2021

Framework of Big Data Analytics in Real Time for Healthcare Enterprise Performance Measurements

Ahmed Mohamed
University of Central Florida

FRAMEWORK OF BIG DATA ANALYTICS IN REAL TIME FOR HEALTHCARE ENTERPRISE PERFORMANCE MEASUREMENTS

by

AHMED ABD-EL-HAFEZ MOHAMED
B.S. Electrical Engineering, Alexandria University, Egypt, 1990
M.S. Electrical Engineering, City College of New York – CCNY, New York, 1994
M.S. Industrial Engineering, University of Central Florida, Orlando, Florida, 2018

A dissertation submitted in partial fulfilments of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the Universality of Central Florida
Orlando, Florida

Fall Term
2021

Major Professor: Luis C. Rabelo, Ahmad Elshennawy

ABSTRACT

Healthcare organizations (HCOs) currently have many information records about their patients. Yet, they cannot make proper, faster, and more thoughtful conclusions in many cases with their information. Much of the information is structured data such as medical records, historical data, and non-clinical information. This data is stored in a central repository called the Data Warehouse (DW). DW provides querying and reporting to different groups within the healthcare organization to support their future strategic initiatives.

The generated reports create metrics to measure the organization's performance for post-action plans, not for real-time decisions. Additionally, healthcare organizations seek to benefit from the semi-structured and unstructured data by adopting emerging technology such as big data to aggregate all collected data from different sources obtained from Electronic Medical Record (EMR), scheduling, registration, billing systems, and wearable devices into one volume for better data analytic. For data completeness, big data is an essential element to improve healthcare systems. It is expected to revamp the outlook of the healthcare industry by reducing costs and improving quality. In this research, a framework is developed to utilize big data that interconnects all aspects of healthcare for real-time analytics and performance measurements. It is a comprehensive framework that integrates 41 integrated components in 6 layers: Organization, People, Process, Data, Technology, and Outcomes to ensure successful implementation.

Each component in the framework and its linkage with other components are explained to show the coherency. Moreover, the research highlights how data completeness leads to better healthcare quality outcomes, and it is essential for healthcare organization survival. Additionally,

the framework offers guidelines for selecting the appropriate technology with the flexibility of implementing the solution on a small or large scale, considering the benefits vs. investment.

A case study has been used to validate the framework, and interviews with Subject Matter Experts (SMEs) have been conducted to provide another valuable perspective for a complete picture.

The findings revealed that focusing only on big data technology could cause failing implementation without accomplishing the desired value of the data analytics outcomes. It is only applied for one-dimensional, not at the enterprise level. In addition, the framework proposes another 40 components that need to be considered for a successful implementation.

Healthcare organizations can design the future of healthcare utilizing big data and analytics toward the fourth revolution in healthcare known as Healthcare 4.0 (H 4.0). This research is a contribution to this effort and a response to the needs.

This dissertation is dedicated to my family

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Luis Rabelo, for his support and patience. This research could not have been accomplished without his guidance.

I would also like to thank my committee co-chair Dr. Ahmad Elshennawy for his encouragement and motivation.

Thank you to Dr. Gene Lee and Dr. Hulya Julie Yazici for their help and being members of the committee.

Special thanks to Dr. Heather Keathley for her advice in conducting interviews.

I am very grateful to everyone in the healthcare organization who helped me during my Ph.D. journey.

TABLE OF CONTENTS

LIST OF FIGURES xii

LIST OF TABLES xv

LIST OF ACRONYMS/ABBREVIATIONS xvi

CHAPTER ONE: INTRODUCTION 1

Background 1

Problem Definition 2

Problem Statement 3

Research Question 4

Research Objectives 4

Contribution 4

Scope and Limitation 5

Document Structure 6

CHAPTER TWO: LITERATURE REVIEW 7

Introduction 7

Healthcare Information Technology (HIT) 9

Big Data 10

Definition and Property 15

Lifecycle 16

Technologies 18

Architecture....19
Analytics Framework....20
Enterprise Data Warehouse (EDW)....24
Extract, Transformation, and Load (ETL)....25
Data Lake....25
Data Analytics....26
Business Intelligence (BI)....27
Healthcare Key Performance Indicators (KPIs)....27
Cloud Computing....29
Research Gap Analysis....32
CHAPTER THREE: RESEARCH METHODOLOGY....36
Research Methodology....36
Research Idea....38
Survey of Literature....38
Gap Analysis....39
Justifying Smart Relationship/Potential Solutions....41
Preliminary Architecture for Big Data Analytics....42
Case Study....44
Validation and Lessons Learned....49
Future Research....49

Conclusion 50

CHAPTER FOUR: RESEARCH FRAMEWORK 51

Introduction 51

Healthcare Systems and Data Mapping 52

Healthcare Business 53

Stakeholders 57

SIPOC-R 62

Data in Healthcare 68

Source of Data in Healthcare 72

Big Data and Quality Improvement Relationship in Healthcare 73

Causal Loop Diagram (CLD) of Big Data and Quality 75

Big Data Tools 77

Proposed Framework 80

Solution Design 85

Value Proposition 89

Cost Analysis 92

Expected Outcomes 97

Implementation of Framework 98

Summary and Conclusion 109

CHAPTER FIVE: RESULTS, VERIFICATIONS/VALIDATIONS AND ANALYSIS 111

Introduction 111

System Architecture of the Framework 111

Deployment 116

Exercise the System 117

Validations 123

Case Study 123

Interviews 124

Analysis 128

CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS 138

Summary of Research 138

Contribution to the Body of Knowledge 139

Conclusions 140

Limitation of Research 142

Future Work and Recommended Research 143

APPENDIX A Transcript of Interviews 146

Interview Instructions 147

Interview 1 150

Interview 2 159

Interview 3.....167

Interview 4.....178

Interview 5.....186

Interview 6.....192

Interview 7.....199

Interview 8.....207

Interview 9.....216

APPENDIX B Permission to Use Figures.....225

APPENDIX C IRB APPROVAL.....228

APPENDIX D IRB APPROVAL ADDENDUM.....230

LIST OF REFERENCES.....232

LIST OF FIGURES

Figure 1: Venn diagram of searched words 8

Figure 2: Big Data and Healthcare Goggle Search Trend 9

Figure 3: Wireless Body Area Network data collection and cloud 12

Figure 4: Model of Big Data Analytics 17

Figure 5: HDFS Architecture 19

Figure 6: Big Data Architecture 20

Figure 7: Big Data Analytics 21

Figure 8: Data Analytics Framework 22

Figure 9: iCARE Framework 23

Figure 10: Cloud-based Data Collection Model 30

Figure 11: Cloud-Based Fuzzy Model 31

Figure 12: The Literature Review Gap 34

Figure 13: Research Methodology Steps 37

Figure 14: Smart Relationship Diagram 42

Figure 15: Big Data Analytics Architecture 43

Figure 16: CMS Transition Timeline from PQRS to MIPS 45

Figure 17: Manual Process Search 48

Figure 18: Automated Process Search 48

Figure 19: Healthcare Mapping for Big data Analytics 52

Figure 20: Process Map of Data Creation in Healthcare 61

Figure 21: SIPOC-R Diagram 62

Figure 22: SIPOC-R of Data Created for Outpatient 64

Figure 23: SIPOC-R of Data Created for ED Patient 65
Figure 24: SIPOC-R of Data Created for Inpatient 66
Figure 25: SIPOC-R of Data Created and Consumed during Revenue Management Cycle 67
Figure 26: HIE Diagram of Sharing Data Electronically 69
Figure 27: Florida HIE- Participants 70
Figure 28: Oneflorida i2b2 participants 71
Figure 29: OneFlorida i2b2 Data and Access 72
Figure 30: Source of Data in Healthcare 73
Figure 31: CDL Data, Quality & Profit 76
Figure 32: Foundation of the Framework 80
Figure 33: Proposed Big Data and Analytics Framework for Healthcare Enterprise 83
Figure 34: Big Data Analytics Solution Design Framework 86
Figure 35: Cloud Computing Models 88
Figure 36: Customized Data Analytics Configuration 89
Figure 37: TCO of Cloud Computing Life Cycle 93
Figure 38: TCO Components of Big Data Analytics 94
Figure 39: List of Expected Outcomes of Adapting the Framework 98
Figure 40: Hosted Data Warehouse Components Diagram 103
Figure 41: Old Organization Structure for Data Related Applications 104
Figure 42: New Data and Analytics Organization Structure 105
Figure 43: The six Pillars of Data Governance 107

Figure 44: Modernized Data and Analytics Platform Transformation – A Composable Architecture 108

Figure 45: Six-View Representing Systems Architecture 112

Figure 46: Big Data System Architecture 115

Figure 47: Chart comparisons from 2016 to 2019 119

Figure 48: Analytics Tools Chart 120

Figure 49: Interview Steps 126

Figure 50: Proposed Framework 141

Figure 51: Big Data and relationship with AI and ML 144

LIST OF TABLES

Table 1: The Literature Research Gap Analysis 33

Table 2: Research Gap Analysis 40

Table 3: Mapping Organization, Process and Data Project 55

Table 4: Criteria of Selection for Big Data and Analytics.....78

Table 5: Estimated Cost of Big Data Analytics in Public Cloud 96

Table 6: List of System Applications 101

Table 7: List of Analytics Tools:.....102

Table 8: Organization for Healthcare Ranking and Awards.....118

Table 9: Trends from 2016 to 2019.....119

Table 10: Cost of Analytics Tools.....121

Table 11: Summary of SME Responses 130

LIST OF ACRONYMS/ABBREVIATIONS

ACO	Accountable Care Organization
APM	Alternative Payment Model
CMI	Case-mix index
CMS	Centers for Medicare & Medicaid Services
DRG	Diagnosis-Related Group
ELT	Extract, Load, and Transform
ETL	Extract, Transform, and Load
HC	Healthcare
HCOs	Healthcare Organizations
HDFS	Hadoop Distributed File System
HIE	Health Information Exchange
IT	Information Technology
LOS	Length of Stay
NLP	Natural Language Processing

CHAPTER ONE: INTRODUCTION

Background

Healthcare generates many electronic health records (EHRs) and electronic medical records (EMRs) in modern countries. These records are warehoused by healthcare providers such as physicians, specialists, hospitals, clinics, labs, radiology departments, pharmacists, and other health professionals (Srinivasan & Arunasalam, 2013). Some studies acclaimed that big data will add about $300 billion annually into US healthcare if used in an innovative and effective method (Belle et al., 2015). The government, healthcare organizations, and the public community consider innovation as critical factors to reduce the cost of healthcare by improving medical decisions, enhancing patient monitoring, and bettering public health observation (H. Yang et al., 2014). Minger (2014) projected that big data is one of three potential areas of innovation in the medical field that will help provide medical care to the public and transform how we live. Google trends indicate that the “Big data” keyword started to become a hot topic in 2011, and it reached its heights in 2015 (Wong et al., 2015). Personalized healthcare will benefit from big data technologies that allow patient-centered care instead of a disease-centered approach (Chawla & Davis, 2013).

The concept of Industrial Engineering (IE), as described in the IE body of knowledge (IEBoK) (Berry, 2018), can contribute to design, architect and develop a solution that allows consuming the massive amount of the available medical records and data to improve the quality outcomes and reduce the spending of healthcare. Furthermore, industrial engineering offers solutions on areas that jeopardize nations' economics, such as healthcare in the United States of America (USA). It deals with a holistic view of the problem with different perspectives,

combining interdisciplinary fields to avoid silos, creating growth opportunities, and understanding the real issue to provide solutions (Darwish & van Dyk, 2016).

Finding a structured approach to utilizing big data that benefits the healthcare industry is addressed in this decade.

Problem Definition

The recent annual healthcare spending in the US is about $3 trillion, and it is accounted for 17 percent of the nation's Gross Domestic Product (GDP), compared to 5 percent of GDP in 1960. This level of spending ranks the USA internationally number 36^{th} in life expectancy (Saria, 2014). Therefore, the high cost of healthcare is the main issue that needs to be addressed to allow better outcomes in health and quality of life.

In 2010, the total amount of digital data exceeded one zettabyte (10^{21} bytes), and it was doubled in the year 2011. Similarly, EHRs have doubled between 2009 and 2011 (Bottles, Begoli, & Worley, 2014). In 2011, the U.S. healthcare organizations generated 150 exabytes (10^{18} bytes) of digital data (Lecklider, 2015). The main challenge is to collect data in a single place and in one format to apply learning algorithms and analytics (Kaur & Rani, 2015).

The challenge of unstructured and semi-structured data is that it wastes data without converting it to usable knowledge. For example, in healthcare, it is hard to separate valuable data from ineffectual data from large volumes of data promptly needed to make fast but accurate medical decisions (Chawla & Davis, 2013). Also, it is hard to perform a research task in a large volume of unstructured data without a proper tool that divides the datasets into smaller chunks for faster processing (Cunha, Silva, & Antunes, 2015).

Healthcare organizations have several Key Performance Indicators (KPIs) and metrics in different areas such as finance, quality, research, customer satisfaction, clinical operations, and

business operations. Daily, weekly, and monthly reports based on historical data are generated, but, unfortunately, many of these reports do not correlate. It requires access to run in different systems with various data types that are not stored in a centralized system location. In addition, many of these reports need information from data that includes structured, semi-structured, and unstructured formats such as database, freeform texting, clinical notes, and images (Kimball & Ross, 2013). The sources of data come from Excel, Extensible Markup Language (XML) file and American National Standards Institute (ANSI) X.12, and scanned documentation in Graphics Interchange Format (GIF), Joint Photographic Experts Group (JPEG), or Portable Document Format (PDF) format.

80% of medical records are in an unstructured format. Clinical notes written by physicians are an example of data that is not currently used. As a result, healthcare providers can better understand the population they serve regarding their conditions and treatments. In addition, they need to know about the illness and its severity to their patients (Shah, 2016).

In addition, the healthcare of privacy act of 1996, "Health Insurance Portability and Accountability (HIPAA)," regulates sharing and exchanging personal health information inhibits sharing data from being utilized effectively (Terry, 2015).

Indeed, both healthcare and big data have challenges that allow researchers to conduct more studies on the topic.

Problem Statement

Healthcare organizations currently use a manual process that requires healthcare workers to search for the needed information from semi-structured and unstructured data types that introduce time constraints and lack of accuracy issues.

Research Question

The lack of a method that employs appropriate technologies to help healthcare organizations efficiently utilize the unused semi-structured and unstructured data types generates the need to create a framework. This framework in which semi-structured and unstructured data can provide valuable outcomes, such as avoiding the timely manual process.

Therefore, the question will be, "how do we introduce the right processes in a framework that consists of the appropriate components of big data and analytics technologies to create value from the voluminous data and accurate outcomes in real-time for healthcare organizations?"

Research Objectives

The objective of this research is to create a framework in a structured approach that (i) will define the appropriate methods to consolidate all available data with various types of formats and location of storage, and (ii) will present the big data tools and techniques to develop Key Performance Indicators (KPIs) for healthcare organizations. This framework will be built around big data and analytics as a foundation that allows healthcare organizations to reduce the cost of operations, improve healthcare quality, and meet the requirements of US government health agencies.

Contribution

This research intends to contribute to the healthcare industry by providing a framework that makes healthcare organizations aware of big data technologies and other required components that could be utilized to enhance healthcare quality and improve the process for better efficiency, cost, and effectiveness. Some models have utilized big data technology to address a challenge in a specific area in the healthcare industry, especially in the clinical space.

However, there is no comprehensive solution that could be adopted to customize and address the challenges in quality outcomes and the stress of increasing the financial spending in the economy.

The distinction of this research comes from emphasizing the industrial engineering approach to architect a comprehensive solution to deliver value on data completeness. In addition, this development can introduce new capabilities such as process automation, which are currently performed manually to submit quality metrics and reports to government agencies for reimbursement.

Therefore, the essential contribution here is to illustrate the value propositions of leveraging big data technology by measuring the performance of the healthcare industry in both quality and financial aspects. Furthermore, the case study will show the components to build a solution model in the cloud to take advantage of the agility and the economic benefits that cloud computing provides by having no upfront purchase, compared to spending for purchasing software and hardware.

Scope and Limitation

Healthcare is a very sophisticated business with many stakeholders, including hospitals, physicians, pharmacies, radiology, labs, and other ancillary services.

In addition, unlike other businesses where both the customers and sellers are the only stakeholders who know the cost of the services in advance and agree on terms and agreements (T&A) before the deal occurs, healthcare transactions are unknown to all. In healthcare, sometimes customers do not know the price upfront before the healthcare services they receive, and they do not know who they will pay, i.e., hospital, surgeon, or other specialists. Hospitals do

not know who the payer is, the customer, insurance company, or both. They also do not know how much they will be paid or will be paid at all.

Because healthcare as a business is enormous, this research to develop a big data analytics framework will not standardize all aspects of various healthcare areas. However, it will be a core that could be easily customized for usage in different areas in healthcare.

Document Structure

The arrangement of this document starts with this chapter, chapter one, that gives an overview of the healthcare industry and its challenges. Then, it outlines the purpose of the research and states the need to develop a framework to utilize big data in the healthcare enterprise.

Chapter two covers the literature review of the previous works of the areas covered in this research to determine the literature gap. The need to discover the works performed in the healthcare industry and its performance measurements, big data, and cloud computing is essential to developing this suitable framework.

Chapter three deliberates the research methodology to develop the framework and the detailed steps to conduct this research.

Chapter four illustrates the proposed framework and the value proposition with cost analysis.

Chapter five covers the validation of the framework using a case study and interviewing SMEs for different outlooks.

Chapter six summarizes the research, lists the contributions, concludes the results, and recommends research areas for future work.

CHAPTER TWO: LITERATURE REVIEW

Introduction

This literature review covers the area of big data and analytics in healthcare and the current articles of both areas, and it is an integrative review to find the answer to these research questions:

- How do healthcare organizations handle the massive amount of data in their Electronic Health Records (EHRs)?
- Is big data used in healthcare, and in which areas?
- How will big data help resolve the challenges in the healthcare industry?
- How to optimize the utilization of big data in a dynamic environment such as healthcare?
- What are the performance measurements that government healthcare agencies require?
- Will cloud computing be an appropriate way to implement big data analytics in healthcare?

The method used to search for the “big data analytics” AND “healthcare” keywords to filter the results with 950 journal articles. Secondly, a search for the “healthcare” AND “performance measurements” keywords is conducted. Thirdly, “big data analytics” AND “performance measurements” are also searched. The Venn diagram shown in figure (1) represents the general areas of search interest, big data, analytics, healthcare, and performance measurements.

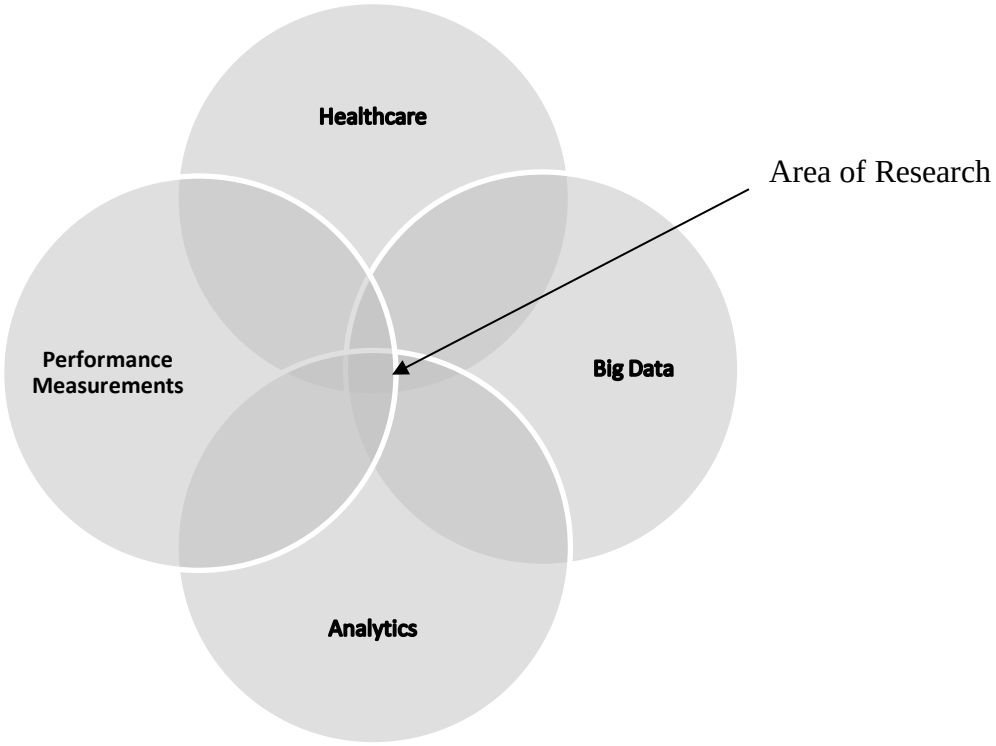


Figure 1: Venn diagram of searched words

Furthermore, 793 doctoral dissertations worldwide were found in the ProQuest database for the period between the year 2012 and the year 2018 for filtered words “health” AND “big data.” After reviewing all dissertations, surprisingly, only “An Epidemiology of Big Data” (Young, 2014) partially met the search criteria. However, “The Effect Of Cardionet Home Tele-monitoring For Congestive Heart Failure Patients: An Observational Research Study” (Patrick, 2015) recommended the use of big data in mHealth for Congestive Heart Failure (CHF) Patients. Still, it was not covered in the dissertation.

Other words “Health Information Technology (HIT),” “Electronic Health Records (EHR),” “Data Warehouse,” “Data Mart,” “Data Lake,” and “Cloud Computing” are searched individually to interrelate with the main topic of the area of research.

All the above searches meet the goal of finding future research opportunities to create a framework of big data analytics in healthcare in a structured approach.

The Google trend in figure (2) shows that there was not much activity searching for big data and healthcare before 2010. However, in the last five years, the trend fluctuates with a maximum of 100 searches on one day in the year 2016.

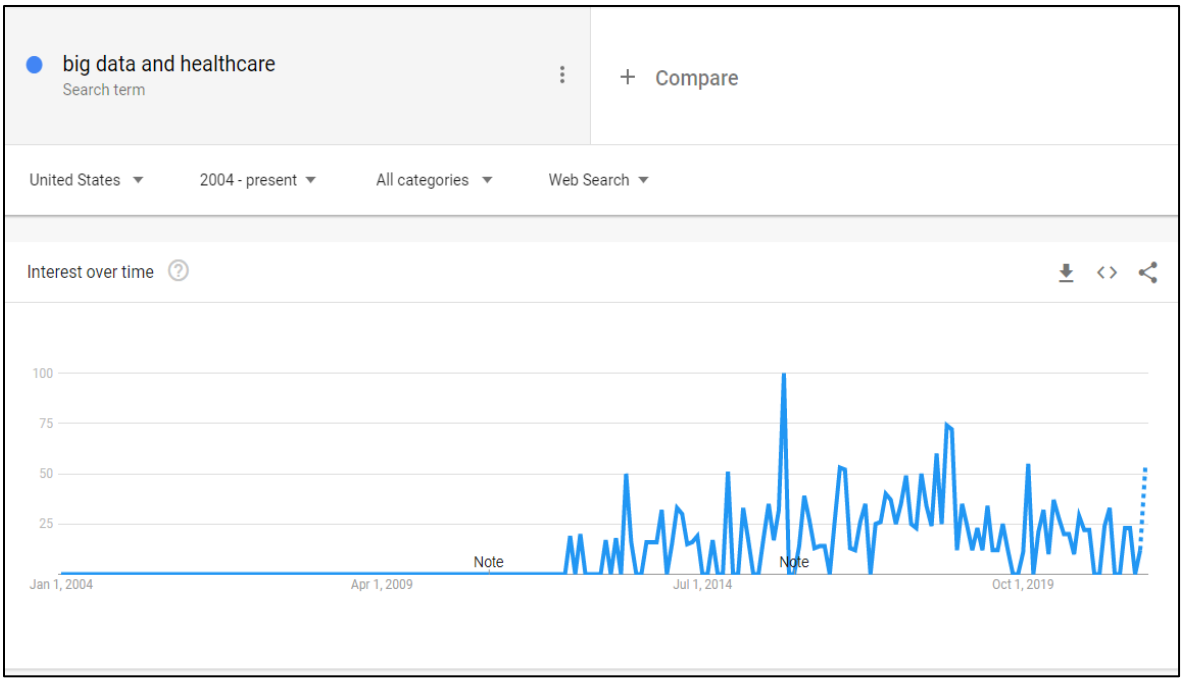


Figure 2: Big Data and Healthcare Goggle Search Trend

Healthcare Information Technology (HIT)

Healthcare organizations generate a considerable amount of data that includes patient medical records in electronic format, patient tracking and patient monitoring, the record of text communication of discharge summaries, and imaging.

It is necessary to define data, information, and knowledge. Datum is defined as something given or a fact that is used to calculate or analyze. Information is defined as forming

meaning from the outcomes of the analyzing process of data. Finally, knowledge is defined as the understanding of analyzing the information (Kuiler, 2014).

Electronic Health Records (EHRs)

The primary function of EHRs is to prepare for patient care delivery and document patient health information during treatment. The health records in EHRs have various roles while treating patients; it supports patient care decision-making and help leaders in their decision-making in healthcare strategy. EHRs contain unstructured and structured data (Häyrinen, Saranto, & Nykänen, 2008). EHRs incorporate electronic medical records (EMRs) and patient-created personal health records. It enables the best way to manage a person's health by sharing patient health information between healthcare providers (Ambinder, 2005).

Electronic Medical Records (EMRs)

The EMRs contain legal history generated in hospitals and ambulatory care about patient's health, and it is the supplier of records for the EHR. It is an application system that is a repository of information containing computerized provider order entry (CPOE), clinical data, order entry, clinical decision support, laboratory, pharmacy, and clinical documentation. In addition, EMRs have the patient's medical record during the visit to the care delivery organization (Garets & Davis, 2006).

Big Data

Big Data is a paradigm shift in thinking about data stored in Data Warehouse (DW) and using the current Business Intelligence (BI) tools to analyze. It is no longer just data sets of numbers and texts that we can organize in the desired format. It is about raw data in structured,

semi-structured, and unstructured formats (Kimball & Ross, 2013). All data are collected in one or more systems that form the concept of big data that is hard to manage to utilize traditional software and hardware tools (Wyllie & Davies, 2015). In healthcare systems, the dataset of medical records does exist, and the big data approach will provide more accurate information in a cost-effective way (Wong et al., 2015). Also, the Federal government is considered a data steward by managing Medicare which contributes to big data. It puts several policies and regulations to accelerate the process of sharing and exchanging health-related information (Madison, 2014). Between years 2012 and 2014, federal government agencies proposed five reports to address significant data challenges. They are "Consumer Data Privacy in a Networked World," "Protecting Consumer Privacy in an Era of Rapid Change," "Big Data: Seizing Opportunities and Preserving Values," "Data Brokers," and "Technological Perspective" report. All five reports recommended the development of data brokers with various suggestions of involvement of legalizations and regulations (Terry, 2015).

In addition to the healthcare organizations' stored data and the federal government's data steward, connected devices with healthcare applications will be another big data provider to healthcare. By 2020, it is expected that 50 billion devices that run applications will be connected to the Internet. These connected devices with healthcare applications will produce vast amounts of data in small portions and semi-unstructured (Cortés, Bonnaire, Marin, & Sens, 2015). These devices are taking advantage of the Internet of Things (IoT), a growing technology expected to help enhance the medical fields by interconnecting monitoring devices to the Internet (Abinaya, Kumar, & Swathika, 2015). In addition, wearable computing devices will change how the healthcare industry does business, and it is expected to reach 169.5 million devices in 2017 (Mezghani, Exposito, Drira, Da Silveira, & Pruski, 2015). Wireless Body Area Network

(WBAN) is a concept that connects body-attached sensors via wireless network or cellular data communication in real-time to computer systems. WBANs can store the collected data about body temperature, the sugar level in the blood, blood pressure, and heart rate, as shown in figure (3) (Quwaider & Jararweh, 2015).

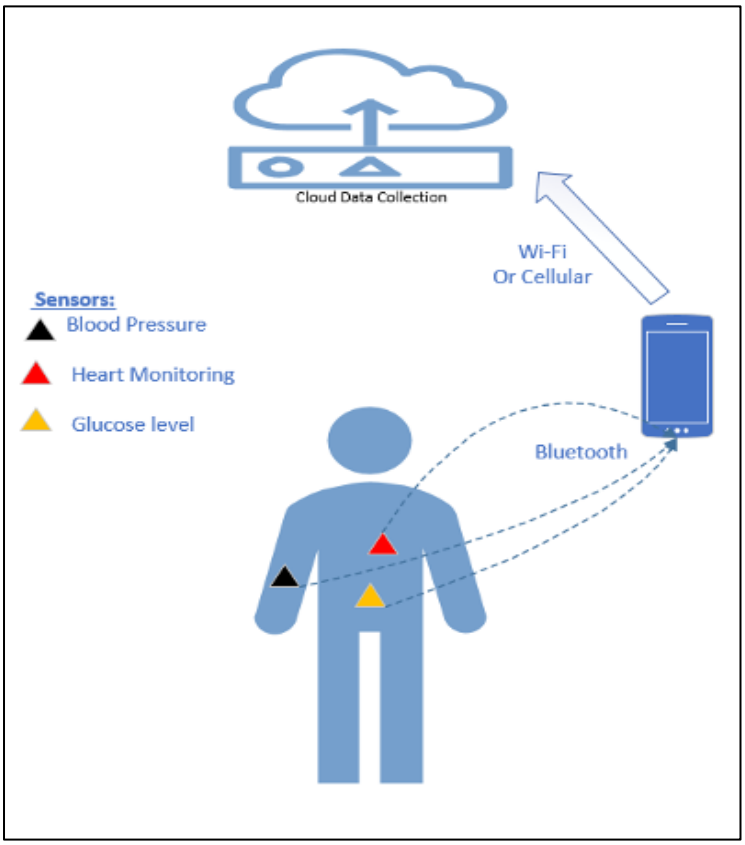


Figure 3: Wireless Body Area Network data collection and cloud
Recreated and Modified from "A cloud supported model for efficient community health awareness," Quwaider, M., & Jararweh, Y. (2015). Pervasive and Mobile Computing. doi:10.1016/j.pmcj.2015.07.012

There are 3 types of data: structured data that is stored in a specific field, semi-structured data that is stored in a specific field but contains information about other data or graphics, and unstructured data that is not stored in a specific field (Kangin, Angelov, Iglesias, & Sanchis, 2015).

Analyzing big data helps identify the data patterns that could predict crises before it happens (Archenaa & Anita, 2015). In March 2012, the USA government allocated 200 million dollars to enhance the tools and techniques of arranging the big data to analyze a massive volume of datasets (Merelli, Pérez-Sánchez, Gesing, & D'Agostino, 2014). Current technologies allow analyzing large amounts of collected data from patients, finding correlations, and building predictive tools of statistical model techniques (Viceconti, Hunter, & Hose, 2015). Utilizing big data in healthcare aims to help physicians identify the risk factors of future diseases by analyzing patients' medical information in real-time (Chawla & Davis, 2013). However, in 2011, the market report of Health Data Analytics stated that data analytic was used in only 10 % of hospitals in the USA (Bottles et al., 2014).

In the significant data era, the amount of information captured is more than what ordinary database tools could handle. Some many challenges and questions need to be addressed to utilize big data. For example, who owns the data? Which data is worth keeping? Is the data analytically meaningful? Big data needs a framework similar to the library system that lets librarians collect, cross-reference, and help find information of the vast amount of data (Frederiksen, 2012). Currently, the focus is on tools to perform data cleaning and data wrangling. It also needs to deal with the analytic-focusing issue with information in traditional databases and graphs databases. For example, a patient's health information records include diagnoses, lab results, and images in healthcare. Still, other information such as events, age, and gender are part of the information (Shneiderman & Plaisant, 2015).

The article “the big deal about big data” (Moore, Eyestone, & Coddington, 2013) listed that executives of healthcare organizations recognized that there are four areas in big data that are important to adopt:

- Data governance and data discipline implementation
- Integrating data to the predicted method.
- Collecting the needed amount of data
- The ability to run several scenarios leads to various conclusions.

Another article, “A Survey of Big Data Analytics in Healthcare and Government” (Archenaa & Anita, 2015), emphasizes the necessity of big data analytics for the healthcare industry for the following reasons:

i. Patient-centric: faster relief, early disease diagnostic, decreasing re-admission rate, and providing proper drug doses
ii. Early detection of spreading diseases: this allows preventive measures to be taken by healthcare providers and make the public aware of it
iii. Monitoring quality in hospitals: to monitor and measure the quality of healthcare services that hospitals provide
iv. Improving treatments is to provide personalized patient treatment and assess risk factors.

Definition and Property

There are several definitions of big data and its properties that have been found in several articles. Baro, Degoul, Beuscart, and Chazard (2015) defined big data as the Log (n*p) is equal or greater than 7; where (n) is the number of statistical data and (p) is the number of variables. They also listed the properties of big data as follow:

i. Big Variety
ii. High velocity
iii. Obstacles in veracity
iv. Obstacles in workflows
v. Obstacles in computational techniques
vi. Obstacles in filtering valuable information
vii. Obstacles in sharing information
viii. Obstacles in finding experts in the field of big data.

Moreover, the Arthurs of the articles addressed other challenges in the big data concept such as:

i. Reuse of data
ii. Detecting false information
iii. Challenges with privacy

Chan (2013) stated that storage and processing power are essential components that deal with the big data characterization of 3Vs: high volume, velocity, and variety. In addition to the 3Vs, veracity was described in the article as the accuracy of the data source and the fittingness of data for the target. Also, the value was described as the benefits to the business by utilizing big data. Healthcare's big data has similar issues, having different types and formats of data

(Variety), integrating a large amount of data (volume), and the fast-growing amount of data (velocity). Moreover, checking the correction of data (veracity) and converting data into (value) is essential to contribute to the healthcare community (C. C. Yang & Veltri, 2015).

Moore et al. (2013) defined the attributes of big data as:

i. Volume: fast-growing data than what was early available
ii. Velocity: real-time availability
iii. Variety: several data sources.

Researchers in socio-medical found that the 3Vs are dynamic and relay on time. Therefore they are not sufficient for their area of research (Wong et al., 2015).

Lifecycle

Archenaa and Anita (2015) listed the following phases of the lifecycle of big data:

i. Collecting Data: store data from different sources
ii. Data Cleaning: identify the junk data and remove it
iii. Classify Data: categorize the type of data being structured, semi-structured, or unstructured.
iv. Data Modeling: it is to analyzes the classified data.
v. Data Delivery: it is the phase that provides reports and charts.

Mezghani et al. (2015), in their article "Stream processing of healthcare sensor data: studying user traces to identify challenges from a big data perspective," listed similar steps to be able to use big data as follow:

i. Data acquisition: it is the ability to obtain and filter data from many sources.

ii. Data cleaning

iii. Integration and Representation of Data

iv. Data modeling and analysis

v. Interpretation: find and present the relationship of events in a large-scale data

Another big data model is proposed by Johri, Singh, and Srivastava (2014) that shows the different phases. It identifies the needed tools in each phase to benefit from big data. The model was done to improve children's health, but it can also be adopted for healthcare in general. Figure (4) below is self-explained with the sequence of steps.

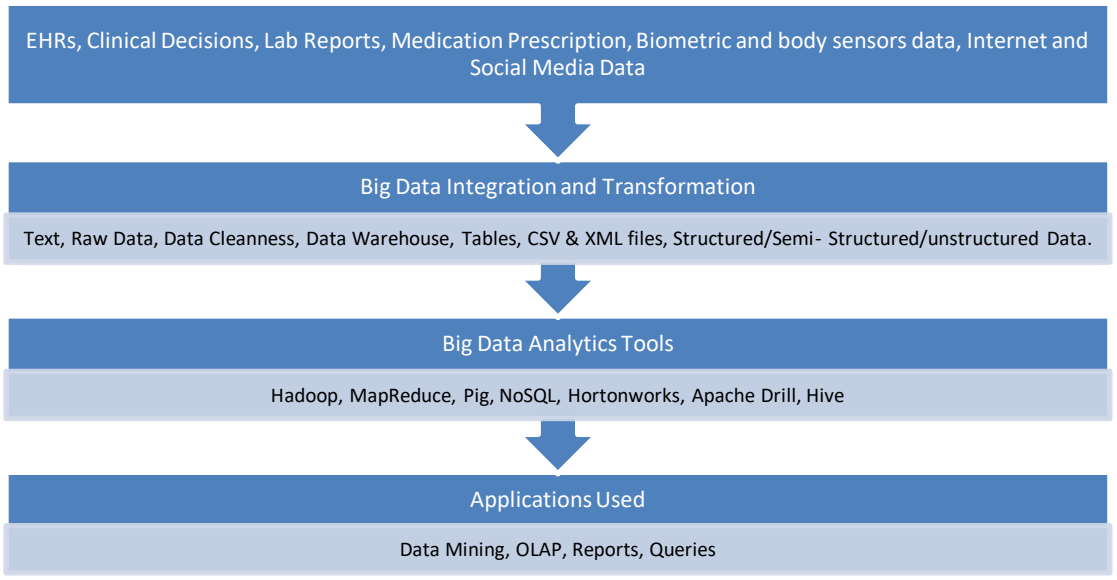


Figure 4: Model of Big Data Analytics
Recreated and Modified from "Predicting Child's Health using Big Data Analytics," Johri, V., Singh, V., & Srivastava, I. (2014). *International Journal of Advanced Research in Computer Science, 5*(7), 193-198.

The tools and technologies that are listed in the model are explained in the next section.

Technologies

Several technologies are involved in big data.

Databases:

i. NoSQL - “Not Only SQL”: it is a non-SQL and non-relational database that has the capabilities of scalability and distribution of big data. The NoSQL database contains key-value-store, key-value-cache, tuple-store, and object database. Apache Hbase open-source and Oracle NoSQL are examples of NoSQL databases that are scalable and reliable (Chan, 2013).

ii. Apache Hadoop: is a framework that contains computers clusters that distribute the process of big data. Hadoop has two main components:

- Hadoop distributed file system (HDFS): is a storage system that distributes files in several computer clusters.
- MapReduce: is a parallel processing system that distributes computing tasks to several servers and accumulates results.

The disadvantage of Hadoop is that it cannot process in real-time because of the high latency (Chan, 2013)

The components of HDFS are:

- Name node: this is the master node that handles requests that are sent from clients.
- Secondary Name node: this is a backup node if the primary Name node fails.
- Job Tracker: this is the MapReduce that assigns tasks to the name node. (Archenaa & Anita, 2015).

Figure (5) shows the flow of tasks between the client and the name node

Figure 5: HDFS Architecture
Recreated and Modified from "A Survey of Big Data Analytics in Healthcare and Government," Archenaa, J., & Anita, E. A. M. (2015). Procedia Computer Science, 50, 408-413.

iii. MongoDB: is a database suitable for text data format, and it conforms to Hadoop and NoSQL. Therefore, it is adequate for the needs of scalability and high-performance computing for big data storage (Cunha et al., 2015).

Architecture

Client-server architecture is presented as a big data model to collect processes and analyze big data. It includes:

- Clients: This is a platform that runs a NoSQL database
- Distributed file systems such as Apache Hadoop
- Server Cluster: parallel computing and High-Performance Computing (HPC)

Figure (6) shows the model of client-server architecture (Chan, 2013)

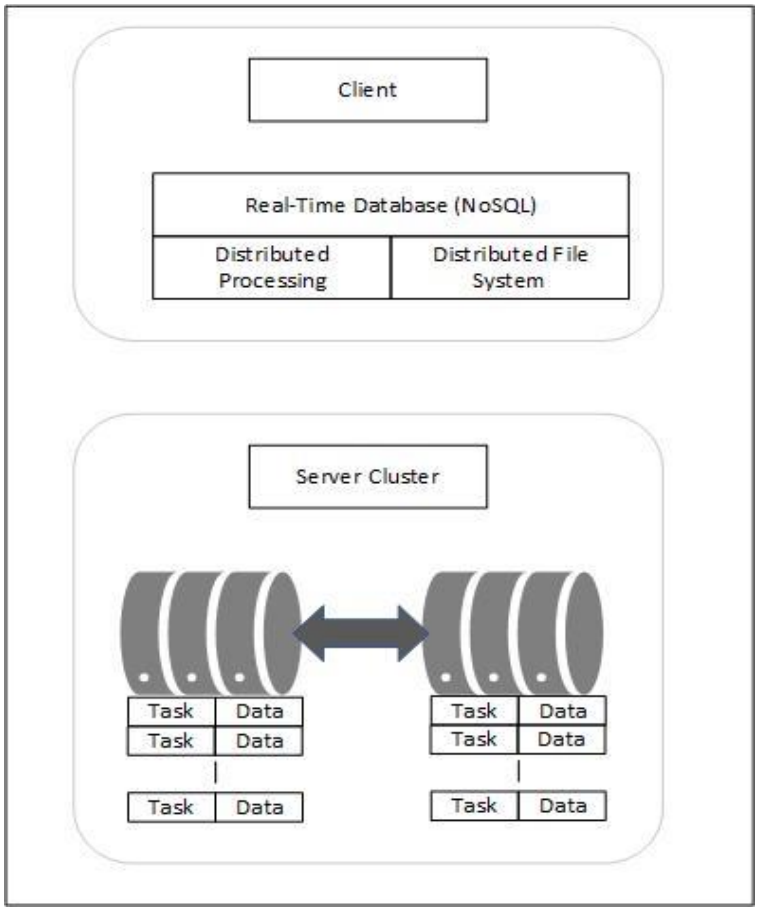


Figure 6: Big Data Architecture
Modified From "An architecture for Big Data analytics," Chan, J. O. (2013). Communications of the IIMA(2), 1.

Analytics Framework

Big data analytic uses parallel and distributed computing process power provided by Hadoop for unstructured data. Map Reduced data will have the ability to be incorporated with the structured data in the data warehouse to get processed by business intelligence (BI), online analytical processing (OLAP), and data mining tools.

Figure (7) below clarifies the components and flow of data analytics (Chan, 2013).

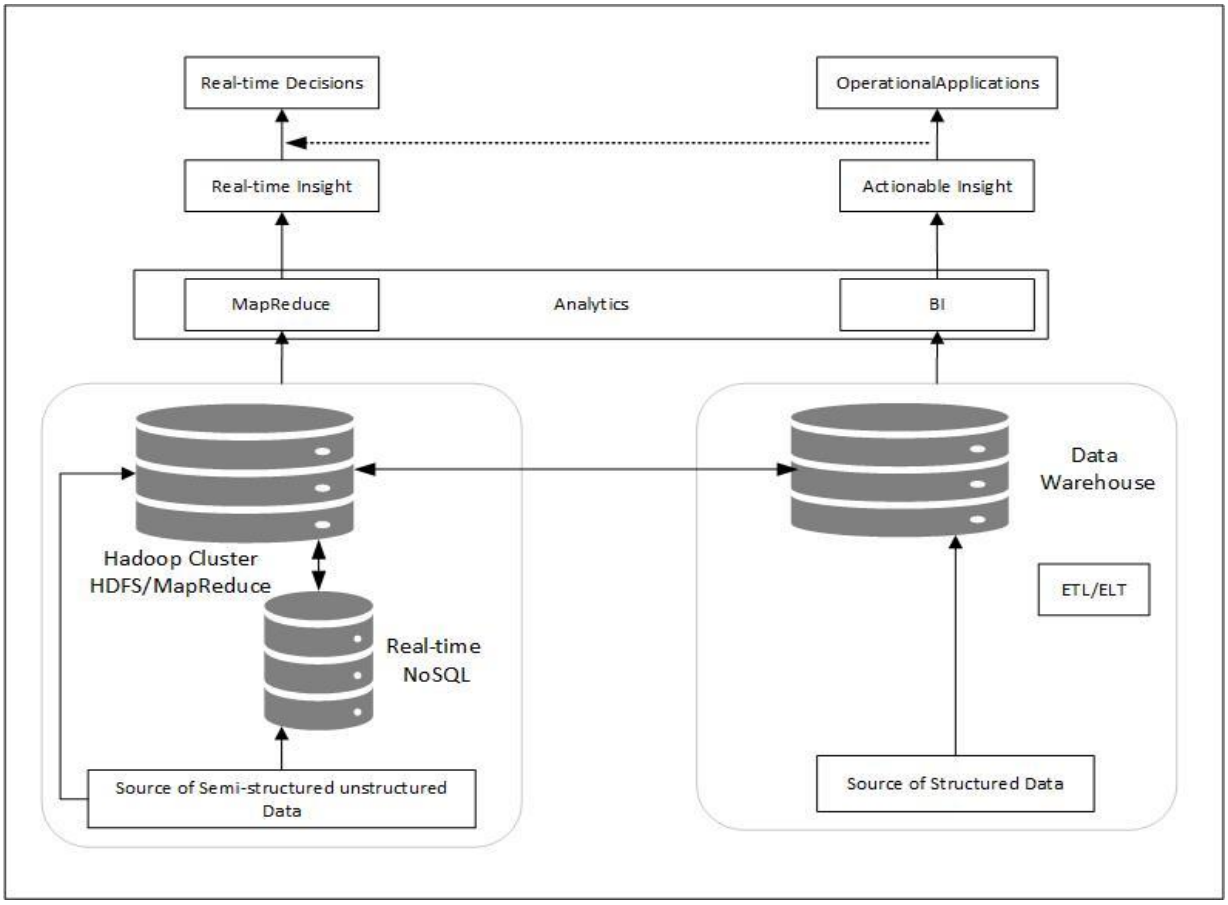


Figure 7: Big Data Analytics
Recreated and Modified From "An architecture for Big Data analytics," Chan, J. O. (2013). Communications of the IIMA(2), 1.

Ontology-based analytics helps physicians utilize the Decision Support System (DSS), especially in emergencies, and the database is accessed via the oncology - Web Ontology Language (OWL) (Abinaya et al., 2015).

Kuiler (2014), in his article "From Big Data to Knowledge: An Ontological Approach to Big Data Analytics," proposed another analytics framework of big data that accommodates both small and big data.

Figure (8) shows the framework's components that are based on ontology and the lexicon.

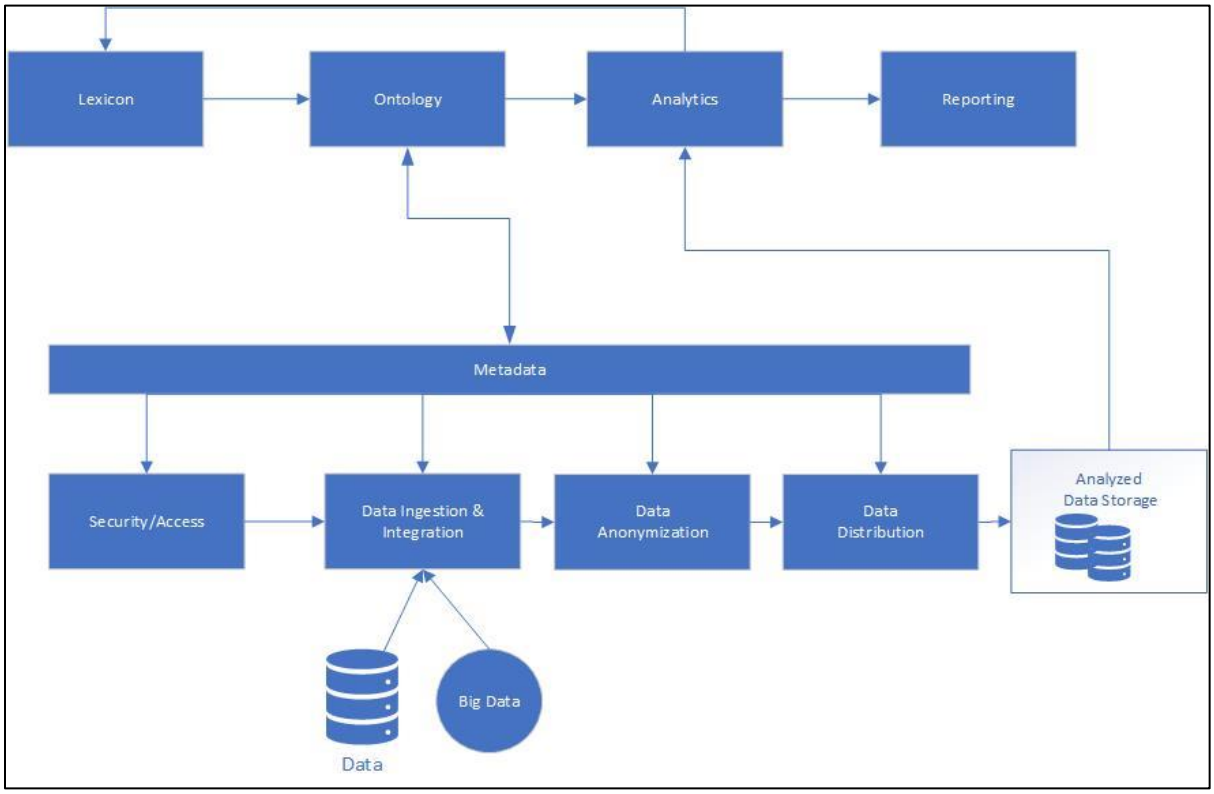


Figure 8: Data Analytics Framework
Recreated and Modified from "From Big Data to Knowledge: An Ontological Approach to Big Data Analytics Kuiler," E. W. (2014). Review of Policy Research, 31(4), 311-318. doi:10.1111/ropr.12077

Both the ontology and the lexicon are the provider of data analytics that could be adjusted. For example, in healthcare, clinical terms are provided by the International Health Terminology Standards Development Organization. Metadata is formed to fulfill the standards, and it also has the capabilities of separating valuable data instead of discarding the noise (Kuiler, 2014).

To provide a "patient-centered" model, Chawla and Davis (2013) introduced a framework called CARE (Collaborative Assessment and Recommendation Engine) that utilizes big data technologies. It creates a personal risk profile for each patient based on comparing with

other patients with similar diseases. It is considered a predictive method to avoid future health threats to patients.

Figure (9) illustrates the components and flow of the CARE framework.

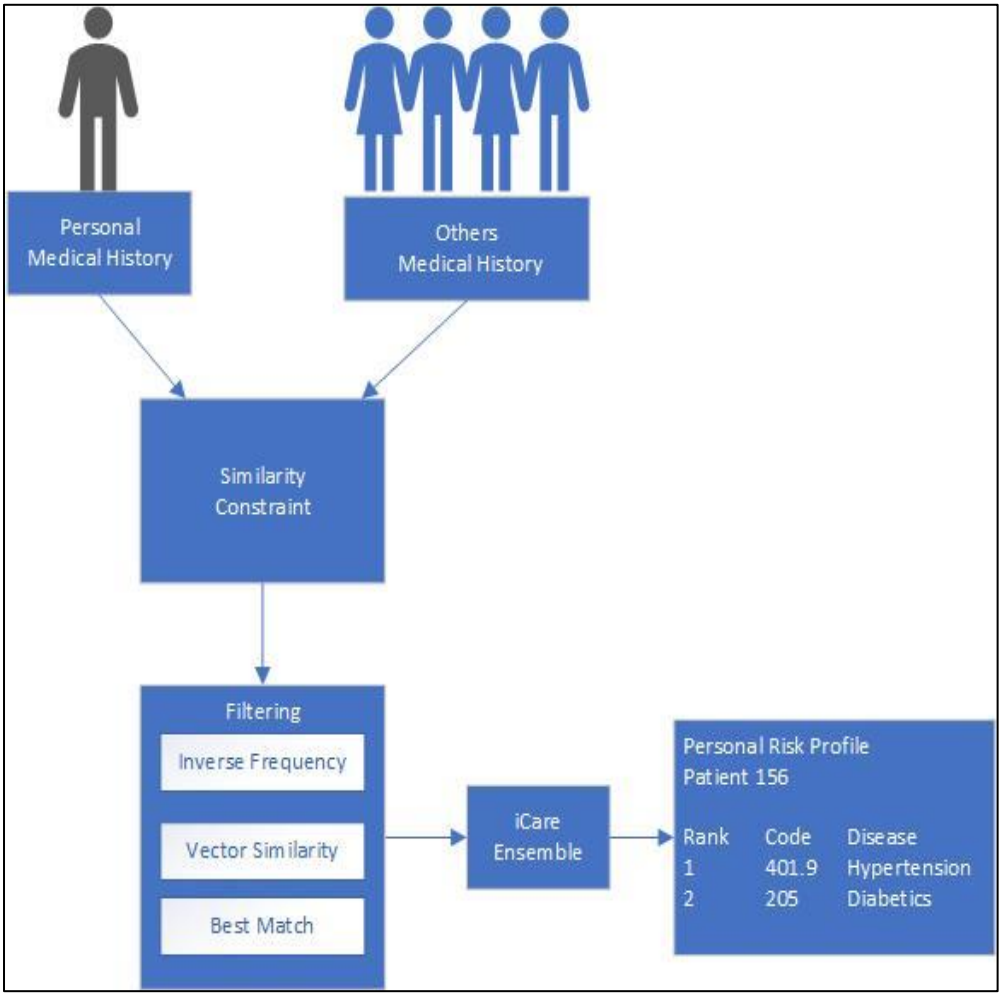


Figure 9: iCARE Framework
Recreated and Modified from "Bringing big data to personalized healthcare: a patient-centered framework," Chawla, N. V., & Davis, D. A. (2013). JGIM: Journal of General Internal Medicine, 28, S660-665 661p.

When a new patient arrives, his /her medical record is compared with all other medical records from other patients. The system detects the shared diseases based on the lab results, images, family histories, etc. The engine of CARE that runs mathematical algorithms will have the capability to predict the new patient. It is an early indicator for likely disease that the new

patient may have. CARE could be an excellent solution to bring big data to the personalized healthcare of individuals.

Enterprise Data Warehouse (EDW)

The concept of data warehousing is described as a process used to share data efficiently within a healthcare organization, a community, or a nationwide network (Wyllie & Davies, 2015). The data warehouse is a central repository of collected data from different sources that needs mapping sources and targets from understandable data of both current and historical information (Johri et al., 2014). Data is one of the most critical assets of any organization when they can transform it into information. For the last 30 years, we gathered a vast amount of data, but it was not easy to utilize for analytics, and we could not get deep into the details of the data. Therefore, businesses and executives could not have easy accessibility to the information to help them decide based on facts. Therefore, the main goals of the DW are to allow easy access to the information, to contain accurate data, to be flexible for modification, to deliver information promptly, to be built in a secure environment, and finally to give a trusted outcome to achieve successful results (Kimball & Ross, 2013).

Pipe (1997) listed the DW in three different categories:

1. Enterprise Data Warehouse (EDW), where several systems store data captured over a period to help in making tactical and strategic decisions. It contains summarized and detailed current and historical records.
2. Operational Data Store (ODS), where new detailed data to provide daily reports.
3. Data Mart (DM) is defined as a subgroup of the data warehouse, and it serves the needs of a specific area to help decide the departmental level (ȚOle, 2015).

Extract, Transformation, and Load (ETL)

ETL is a system that provides everything in DW and Business Intelligence (BI) from the operational source to presentation. It is to get the data from the source, work it, then present it. First is the extraction of data to put into the DW. This step means taking the needed data from the source and placing it into the DW to read and interpret. Secondly is the Transformation step that prepares the data to be of good quality. The data goes through cleansing, duplicate elimination, integrating from multiple sources, and improving data quality in the transformation step. The third and final step of the ETL process is loading, where the data can be presented. Loading allows structuring the data to be handed off to the final delivery phase (Kimball & Ross, 2013).

Data Lake

Integrating DW and big data has become a strategic direction for many organizations to take advantage of unused data for better decision-making. Data Lake provides the technology solution to transform the data architecture to accommodate the big data era. It allows the ETL process of the data stored in DW and the capabilities of supporting unstructured data cost-effectively compared to DW. Furthermore, unlike DW that stores structured data only, Data Lake allows the storage of a considerable amount of structured, semi-structured, and unstructured data in their original formats from many sources (Mullins, 2018).

Fang (2015) listed the capabilities of a data lake as follows:

1) Keeping the unused data available in its native format until it is needed for analytic use

2) It is a scalable, low-cost repository of data storage,

3) The performing of the ETL process when it is needed and the support of all types of data

4) The ability to analyze specific data for certain departments or case studies.

The data lake concept was used as the main component of constructing a data-managed groundwork for a healthcare project in Europe. It employed the available heterogeneous data stored in Data Lake from internal and external sources in an agile way (Kondylakis, Koumakis, Tsiknakis, & Marias, 2018)

The challenges of using Data Lake are introduced because the dumping of all captured data of various formats with no understanding, no managing, and without the exclusion of cleaning the data could lead to inaccessibility of the data lake over time. Rebuilding the schema of the data lake to consider the structure variations stops the data lake from being inaccessible in the long run (Klettke, Awolin, Störl, Müller, & Scherzinger, 2017). Many organizations lean toward cloud computing for utilizing the data lake concept due to the time it takes to implement and the need for individuals with these technical skills.

Data Analytics

In the mid-1950s, business analytics was introduced with tools that collect more information and detect similarities that the human mind could not identify. As a result, it helped profound insight into crucial business trends based on facts to understand more than what is obvious when managers make decisions (T. H. J. H. b. r. Davenport, 2013).

Business Intelligence (BI)

BI is one of the data analytics offering to explore data sets to detect patterns and find trends. Capturing the data is the needed step to keep the operational records in one place, such as DW. The next step is to analyze the stored data to turn it into information for better decision-making. Therefore, ETL in DW supplies BI applications with data that can analyze as a final step. It is essential to understand the business needs and know the big picture from a business perspective to benefit from the data warehouse concept and data analytics (Kimball & Ross, 2013).

Descriptive analytics is considered BI, which analyzes historical data gathered from table format as an input presented in a visualized output (Alghamdi, Baz, Alsubait, & Alhakami, 2021).

Predictive analytics is also categorized as BI that reveals connections and correlations of data on hand to predict future events by learning from previous actions (Eckerson, 2007).

Prescriptive analytics is the more enhanced version of predictive analytics for finding the best possible solution produced from previous accessible solutions (Lopes, Guimarães, & Santos, 2020).

Healthcare Key Performance Indicators (KPIs)

Parmenter (2010), in his book, key performance indicators: developing, implementing, and using winning KPIs, identified three ways to measure the performance of an organization: Key result indicators (KRIs) that state a perspective of how the organization has done in areas such as customer satisfaction and financial measurements, performance indicators (PIs) present the needed actions to find the profitability of a specific product or services, Key performance

indicators (KPIs) that articulate the steps to raise the performance of the organization radically. 10 KRIs/80 PIs/10 KPIs rule is suggested as a guideline for organizations to follow. The author encourages organizations to pay attention to seven crucial characteristics when implementing KPIs for their success. These seven characteristics are

1) Non-financial

2) Ascertain metrics

3) Regularly measured

4) Pursued by the executive leadership team

5) Comprehended by staff

6) Assigned to someone or team for responsibility

7) Have a significant effect.

Regardless of the business type, measuring the organization's performance over time is essential to understanding the volume of products or services that the organization provides and its competing market. Accessing this information with a simplified presentation is a crucial factor of understandability (Kimball & Ross, 2013).

The article titled "Big Data and KPIs: A Valuable Connection" (Fanning, 2016) suggests combining the useful KPIs existing with the outcomes of the big data analytics to provide new insights and improve the organization's success.

There are many KPIs for healthcare, as examples measure the number of incidents and events, and patients' wait for services. The goal is to construct KPIs that learn from old events and send early indication alerts of undesired future events that are expected to happen. (Baldominos, De Rada, & Saez, 2018).

In general, the landscape to measure the performance of healthcare organizations lays under the following list:

1) Patient Harm Incidents
2) Mortality rate
3) Volume and Patient Census (Throughput and Efficiency)
4) Quality Outcomes and readmission rate
5) Revenue and Profitability (Reimbursement, Expense Management, Bad Debt)
6) Market Share

Cloud Computing

Cloud computing is a technology paradigm that will work with big data to provide better healthcare quality. IoT and WBANs will be used on a large scale in the next few years, and both will create a massive amount of data that needs to be stored and analyzed for a better quality of healthcare and effective health awareness. Cloud computing is a scalable system that supports a vast storage capacity and a high-speed computing process needed for big data. Quwaider and Jararweh (2015) suggest in their article a cloud computing model that is based on Local Cloud (LC) and Enterprise Cloud (EP) layers that have the capabilities to collect large-scale data from millions of Monitored Subjects (MS).

Figure (10) describes the components of the proposed model.

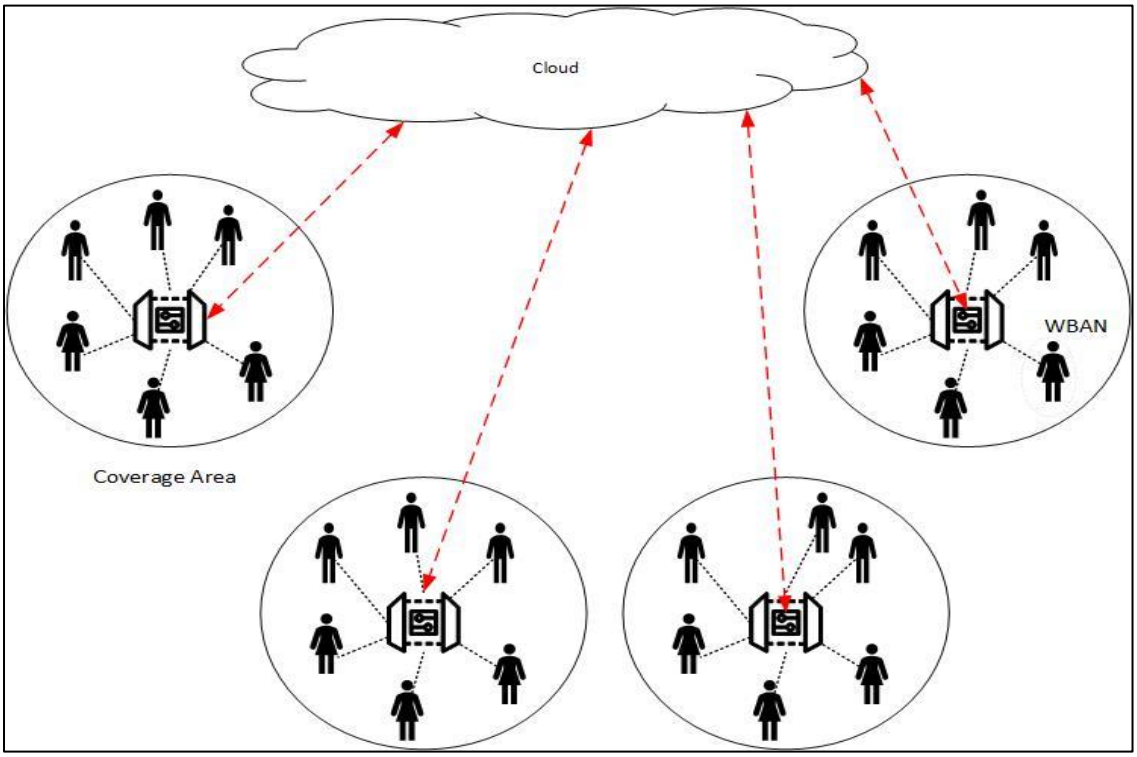


Figure 10: Cloud-based Data Collection Model
Recreated and Modified from "A cloud supported model for efficient community health awareness," Quwaider, M., & Jararweh, Y. (2015).Pervasive and Mobile Computing. doi:10.1016/j.pmcj.2015.07.012

Kimball and Ross (2013) recommended taking advantage of the public cloud for prototyping big data concepts because it is easy to construct and scale. However, they also encouraged organizations to move their data to the private cloud at a certain point to gain control of the stored data. A cloud computing system for healthcare providers in a unified platform of data warehousing (DW) is recommended to collect, store, process, and manage so that medical doctors, hospitals, and insurance companies can benefit from the shared information cost-effectively. The system must follow the requirements to allow security, health information privacy, availability, and integrity (Sundharakumar, Dhivya, Mohanavalli, & Chander, 2015).

Sundharakumar et al. (2015) proposed another cloud computing framework that allows providing low-cost healthcare. The proposed solution considers the 3Vs of big data that allows monitoring patients at their homes. The solution is built based on intelligent techniques and algorithms such as fuzzy logic and neural networks.

Fuzzy logic can set defined values such as on or off, yes or no, and high or low, as shown in figure (11)

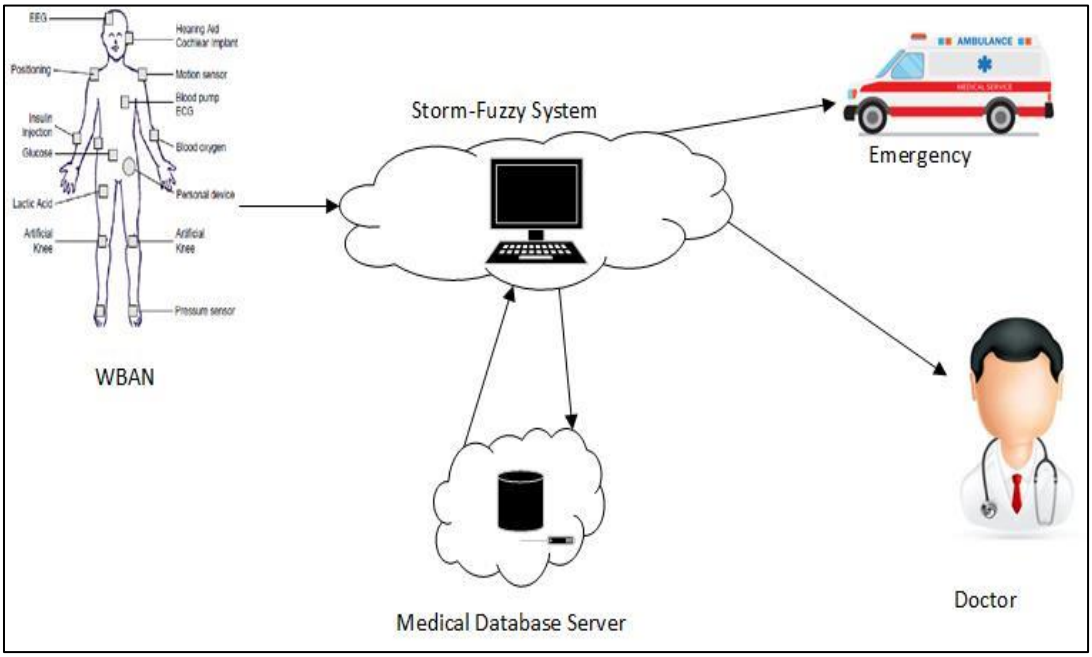


Figure 11: Cloud-Based Fuzzy Model
Recreated and Modified From "Cloud Based Fuzzy Healthcare System," Sundharakumar, K. B., Dhivya, S., Mohanavalli, S., & Chander, R. V. (2015). Procedia Computer Science, 50, 143-148. doi:10.1016/j.procs.2015.04.076

Cloud computing and big data combined will deliver a Knowledge as a Service (KaaS) solution to many healthcare organizations. KaaS is considered an extension layer of the cloud computing model (Mezghani et al., 2015).

Research Gap Analysis

The purpose of conducting a literature review is to identify which research is currently available and work on the gap analysis that pinpoints toward missing components in research that need to be discovered. Aiming for a complete understanding of all pieces involved to build a comprehensive view would allow figuring out the missing components that may lead to a research gap. For example, the case studies that leveraged big data characteristics were built in-house rather than cloud technology.

Utilizing big data in healthcare is a gap that needs to be evaluated for better performance measurement outcomes for both financial and quality aspects. The absence of industrial engineering in existing research is another missing component and is a clear gap. Additionally, there is a need to conduct financial analysis to evaluate the missing saving opportunities for not adopting big data technology compared to the total cost of ownership (TCO) of implementing it. More importance, discovering the impact on business in efficiency and effectiveness is a key to any study that explores the use of any new technology such as big data. Finally, it is vital not to shape a problem to fit technology, as it seems several of the studies did, rather than identifying the problem and finding the appropriate technology that is suitable to solve it.

Table (1) shows the variance of covered topics in the literature review as a gap analysis.

Table 1: The Literature Research Gap Analysis

Authors	Healthcare Specific	Performance Measurements / KPIs	Big Data Model	Data Warehouse/Mart /lake	Hosted in the Cloud
Brennan, Oelschlaeger, Cox, and Tavenner (2014)	X	X	X		
Chawla and Davis (2013)	X	X			
Chan (2013)			X	X	
Kuiler (2014)	X		X		
Mello, Leite, and Martins (2014)		X	X		
Fanning (2016)		X	X		
Sundharakumar et al. (2015)	X				X
Quwaider and Jararweh (2015)	X				X

Research that utilized big data in healthcare was generated and listed in figure (12), but none expand the usage of big data in healthcare performance measurement.

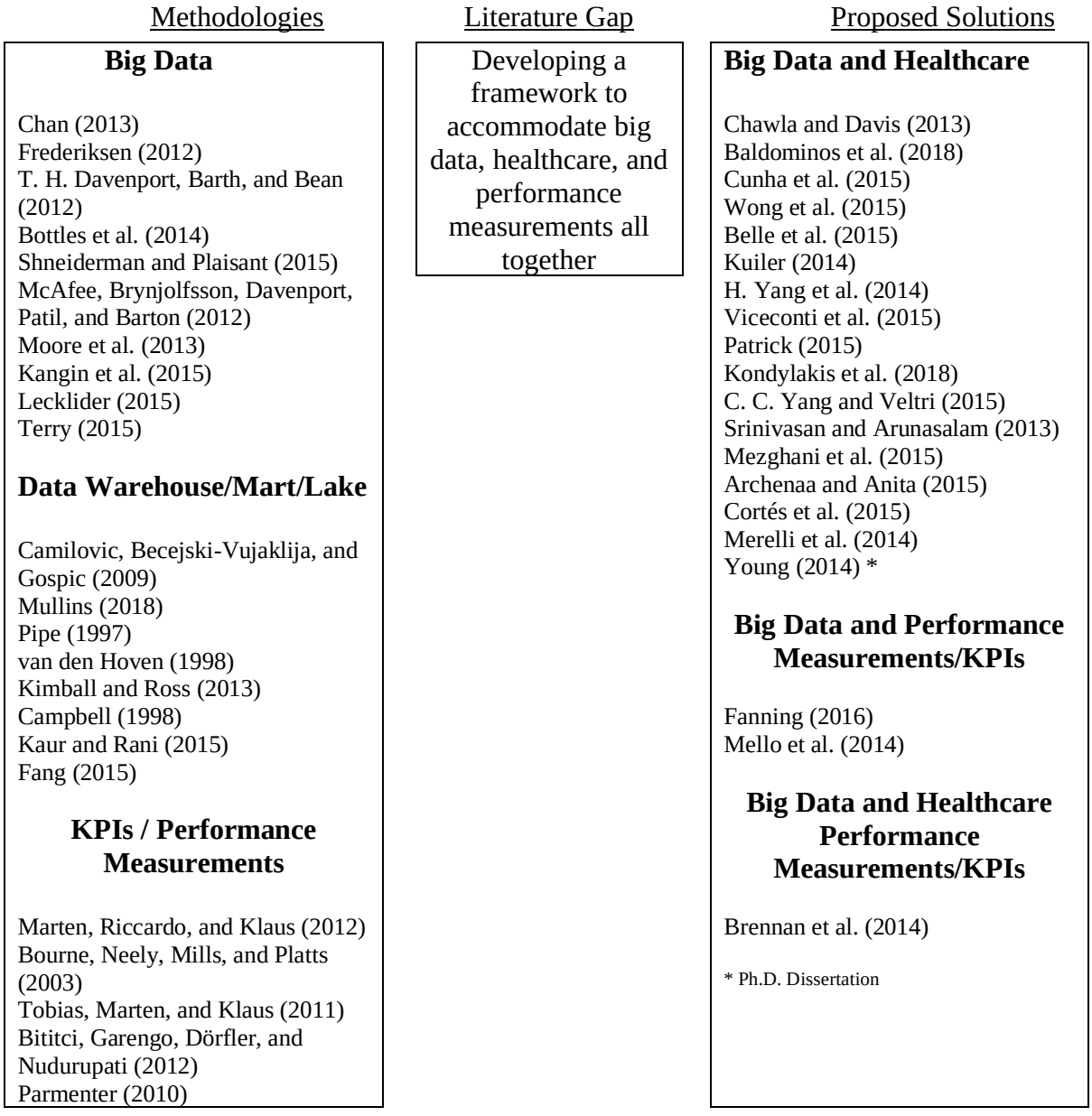


Figure 12: The Literature Review Gap

Indeed, there is missing participation in research that provides an integrated model that combines all three areas of big data, healthcare, and performance measurements. Thus, the conclusion of the gap analysis of this literature review depicted a need for an inclusive framework that integrates all necessary components to build a cost-effective model with the flexibility to modification and customization.

CHAPTER THREE: RESEARCH METHODOLOGY

This chapter covers the "HOW" to achieve the goals of this research. It outlines the research methodology components encompassing the research design and procedures, the data types and sources, the access to collect data, and finally, the financial resource to execute. The end goal is to create a framework that connects the dots of the research idea and fills the gap of utilizing big data in performance measurements in healthcare.

Research Methodology

The research methodology steps in this dissertation begin with the research idea by searching the related topics and technologies in big data and healthcare performance measurements. After deciding on the research point, the second step is to complete a literature survey for an exhaustive exploration. This literature survey aims to discover the current state of academic and theoretical research areas of healthcare as a business and its components, health information technology, Data and cloud computing technologies including traditional and big data concepts, and performance measurement in the healthcare industry. Combining all the literature items collectively to form a holistic outlook to permit them to be used together is identified as a gap that needs analysis for a potential framework or architecture solution. A coherent relationship is documented of all components of the related topics to recognize the purpose of the research's goal, which is covered in the section of Smart Relation/Potential Solution. Next, a Preliminary Architecture of the framework is introduced to construct the model utilizing the tools necessary to apply to big data, followed by the framework development. Then, a case study is implemented to examine the result of the function of the framework and if it meets the expected outcomes as validation 1. Validation 2 and refining the model to enhance the

accuracy of the outcomes is the last step before listing the expecting future research and giving the conclusion. The diagram in figure (13) shows the steps followed in this dissertation.

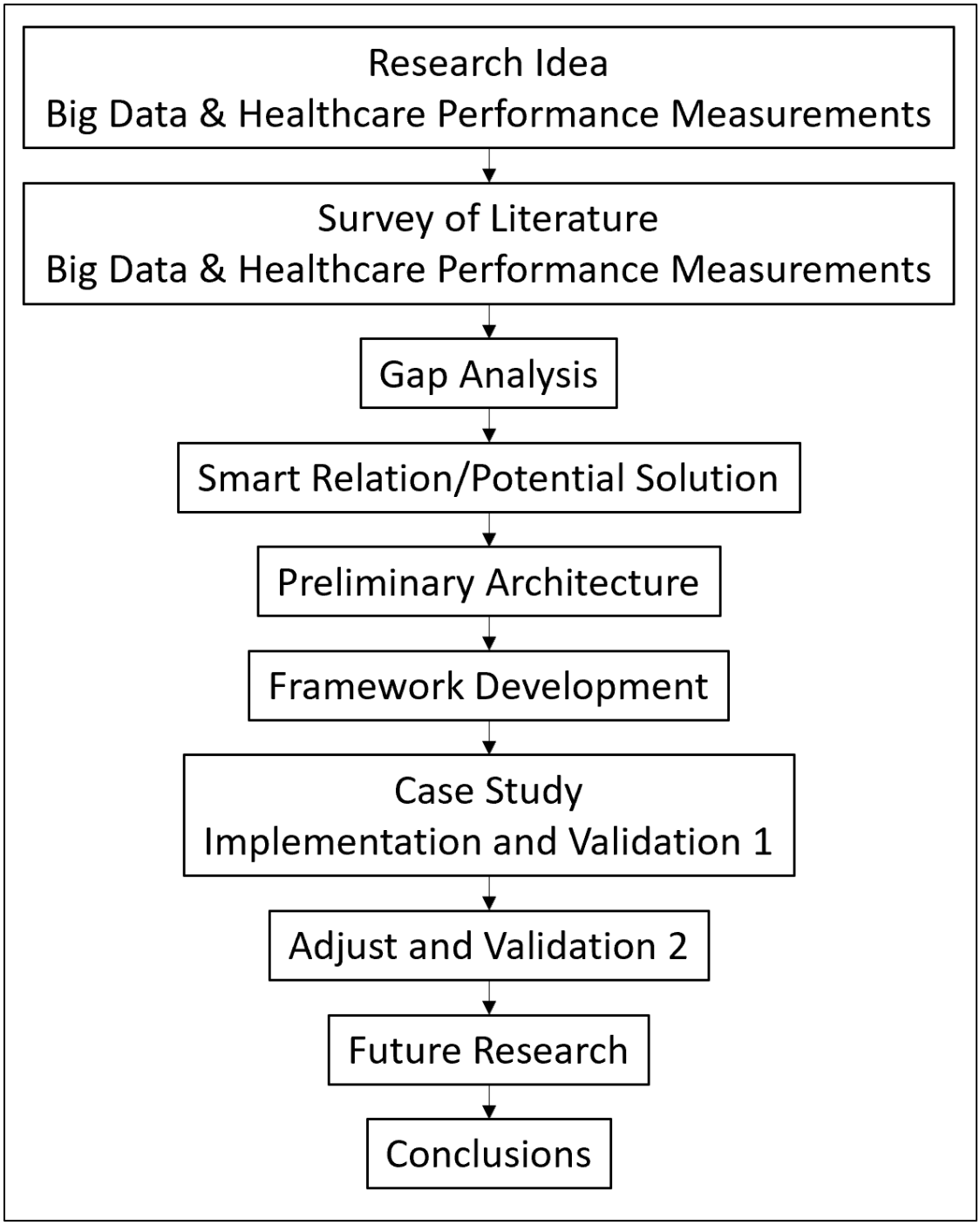


Figure 13: Research Methodology Steps

Research Idea

Several technologies that are recently introduced to the business community, such as Machine Learning, Deep Learning, Artificial intelligence (AI), and Big Data, will change the outlook of the healthcare industry and other related fields. The goal is to explore these technologies that will lead to breakthrough approaches necessary to solve the issues of the high cost and enhance the quality outcomes in healthcare.

It is expected that multiple research foci on the healthcare industry will take advantage of big data features of dealing with images, handwriting, and texts in unstructured data format. The goal is to find innovative solutions to help healthcare leaders be proactive with real-time data in their decision-making. The idea of the research is to architect a solution utilizing big data technology to enhance the performance measurements of the healthcare system.

A literature survey is conducted to find the proposed ideas to improve the performance measurement of the healthcare systems and to present the literature gap that identifies the missing vital points for better optimizations in both quality and cost of the healthcare system.

Survey of Literature

Conducting a literature survey on the research topic and the method followed are essential to answer the “WHY” and “HOW” that other researchers have contributed to this area for a better insight. The goals are to look for academic resources, learn the theoretical approach of previous research efforts, find different points of view or debates, and identify a gap. They were aiming for a 360 degree of view to comprehend the critical problems in the field and the recommended solutions.

The literature survey here focuses on employing big data technology in healthcare, especially performance measurements, from the industrial engineering viewpoint, not from a computer science perspective. There are several research contributions in this field. Still, they are scattered in different areas of healthcare with the emphasis on the medical field to benefit from big data in clinical decision-making. The outcome of the literature survey concludes that there is a missing gap in finding a comprehensive architect or framework that satisfies the needs of using big data in healthcare. The gap analysis in the following subsection determines the missing parts of the current research to fulfill a complete solution.

Gap Analysis

Surveying Literature identified the framework of several case studies that utilized big data in healthcare.

CARE (Collaborative Assessment and Recommendation Engine) framework (Chawla & Davis, 2013), Ontology approach to big data analytics framework (Kuiler, 2014), Architecture of big data analytics (Chan, 2013), and the model of big data (Johri et al., 2014). However, this research's driven efforts concentrated on big data as technology and/or utilizing it in a specific topic in the clinical side of the healthcare industry. Therefore, expanding the search to include performance measurements utilizing big data was crucial to determine the amount of work done to cover this combination. Nevertheless, there was only one research, “Big Data and KPIs: A Valuable Connection” (Fanning, 2016) was found that provided a general concept that was not specific in healthcare. Additionally, “An Epidemiology of Big Data” (Young, 2014) is a doctoral dissertation on Professional Studies in Information Management that covered utilizing big data in healthcare in a specific area of Epidemiology.

Table (2) shows the approaches for the big data in healthcare performance measurements.

Table 2: Research Gap Analysis

Authors	**Problem Addressed**	**Methodology**	**Benefits**	**Drawbacks**	**Extensions / Future work**
Wu (2017)	Predicting Medicare Advantage (MA) members' complaints to CMS	Decision Tree: Large Ensemble with Over-Sampling (LEOS) algorithm	Development of A Big Data Analytics Framework for Forecasting Customer Complaints	The model ignored patients who were not happy but not willing to report complaints to CMS	To address issues that are unrelated to complains
Chawla and Davis (2013)	Personalized Healthcare: A Patient-Centered	Collaborative Assessment and Recommendation Engine (CARE) framework	Provide physicians with a list of patients with high-risk illness	It does not include data other than International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM).	To consider including data from other sources (i.e., labs and disease symptoms)
Kuiler (2014)	Ontological Big Data Analytics	the framework that addresses Little and Big Data Analytics	Give the foundation for transforming data into Information and knowledge.	Addressed large size of structured data but not unstructured or semi-structured	Software development that supports ontology and architectural artifacts framework
Chan (2013)	Big Data Analytics Architecture	Classification of BI & Analytics (BI&A)	Improved the quality of predictions	Limited the tools used to Hadoop and NoSQL	The ethical issue arises because of Big Data
Johri et al. (2014)	discovering child diseases at an early stage as a related factor to the health of child's mother while pregnant	Introduced a big data model to predict a child's health by analyzing the mother health	A good foundation of big data analytics that provided relatable decisions for medical treatment	Used open source with limited capability. Errors in prediction could create a harmful effect	Research in privacy and security.
Young (2014)	Epidemiology of Big Data	They are conducting surveys and collecting data to answer the research question.	Showed the implication about big data and filled the gap between expectation and reality in the field of healthcare	General Concept of big data in healthcare	Phenomenological research to include social media analysis

Hence, the research question “how big data technologies help automate the search of needed performance measurements utilizing semi-structured and unstructured data in real-time and accuracy outcomes in healthcare?” is validated as an identified gap.

Justifying Smart Relationship/Potential Solutions

Healthcare as an industry is a very sophisticated business, and it has several issues in many aspects, including high cost and lack of quality. Adopting digital transformation in healthcare has helped improve the overall experience for both patients and caregivers. As part of Meaningful Use Incentive, physicians and healthcare organizations were required to use EHRs/EMRs to take advantage of the $18B stimulus package for healthcare IT that the US Congress approved in 2009. As a result of the implementation of EHRs/EMRs in the last decade, a vast amount of data became available that allows healthcare organizations to take advantage of data analytics for better decision-making. Even though EHRs/EMRs are used, doctors still prefer to use written-hand notes to update patients' records scanned and stored electronically as an image or a pdf format. That directs to searching the scanned records manually to find information needed about patents, which is time-consuming. This manual operation leads to inaccurate results because of limited human abilities. The need to use big data innovatively appeared to be a worthy effort to develop a search methodology using technology and automation.

Healthcare organizations seek to process data analytics from historical clinical records and notes to propose predictive risk or indicate KPIs in real-time. The accuracy of the information is essential to achieve financial advantage while competing with other healthcare organizations that serve the same patients. In addition, the proposed solution will help the

healthcare organizations' leadership identify the weaknesses in their system that result in losing opportunities and market share, poor patient experience, and costly healthcare. Figure (14) shows the smart relationship diagram.

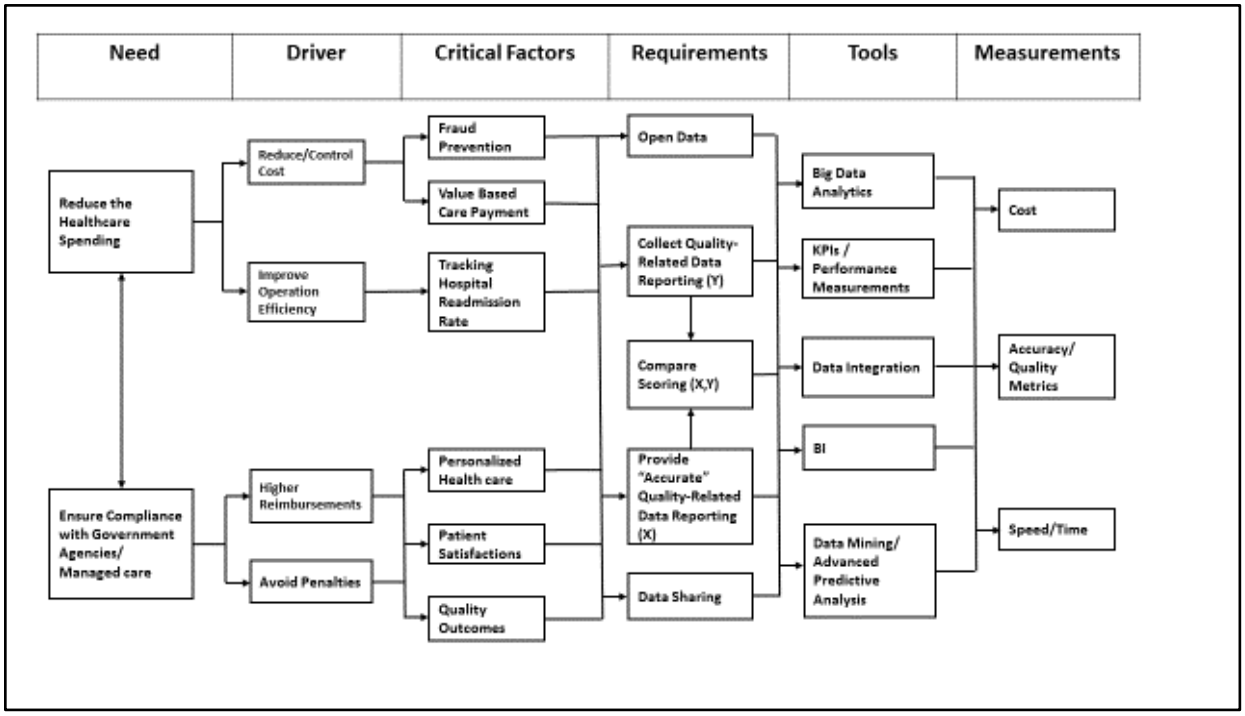


Figure 14: Smart Relationship Diagram

Preliminary Architecture for Big Data Analytics

The performance indicators that the framework will provide as outcomes would focus on the operation services of the healthcare organization inpatient experience as demanded by the Center for Medicare and Medicaid Services (CMS) quality guideline requirements. The framework should optimize processes workflow utilizing big data analytics to automate and customize the outcomes to align the performance indicators with the operation services of healthcare organizations that need to be reported.

The framework shown in figure (15) illustrates the conceptual idea of the model by separating the system into different layers as follow:

1) The Data Source Layer:
2) The Data Storage Layer: A repository data disk space to collect all data with different format types in one storage area.
3) The Data Preparation Layer: responsible for segregating data with its type.
4) The Data Integration Layer: allow the consolidation and transformation between unstructured and semi-structured data types.
5) The Analytics and Reporting Layer: provide operation reports and data analysis.
6) The Presentation Layer: visually shows KPIs and metrics

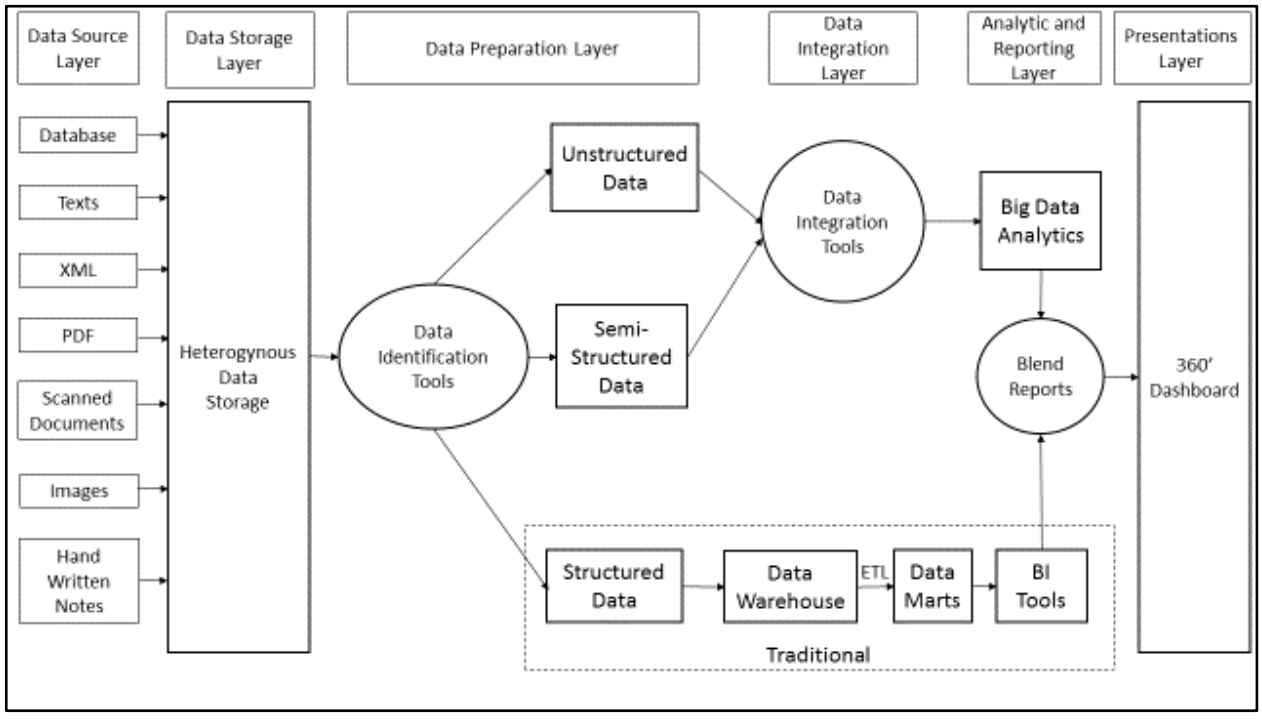


Figure 15: Big Data Analytics Architecture

Case Study

The Center for Medicare and Medicaid Services (CMS) requires healthcare providers to submit an annual quality report for merit reimbursement as instructed in the Physician Quality Reporting System (PQRS) Program.

In the first week of January every year, the healthcare provider downloads a report from the CMS portal that is saved in an excel file format. The report contains randomly selected patients who are treated at this healthcare provider throughout the year. This random sample of patients is picked from the total population assigned to this healthcare provider by CMS. The report needs to be filled out with specific information according to CMS' metrics for this list of patients.

In 2017, CMS merged the PQRS program into the Merit-based Incentive Payment System (MIPS) and announced the change of the data submission method for the 2018 report. The most significant change was that instead of entering a number that represented a specific value that met XML specifications (e.g., 1= no, two = yes, etc.), they were asked to select the appropriate answer from a dropdown box. CMS pushed out an excel template, demonstrating this change, and released the guide explaining the process.

The guidelines for 2019 have recently been released, and the case study will follow the new rules. CMS changes its rules every year, requires a fixable model that is easy to modify and adopt the new rules without revamping the existing model or creating a new system. Figure (16) shows the transition timeline.

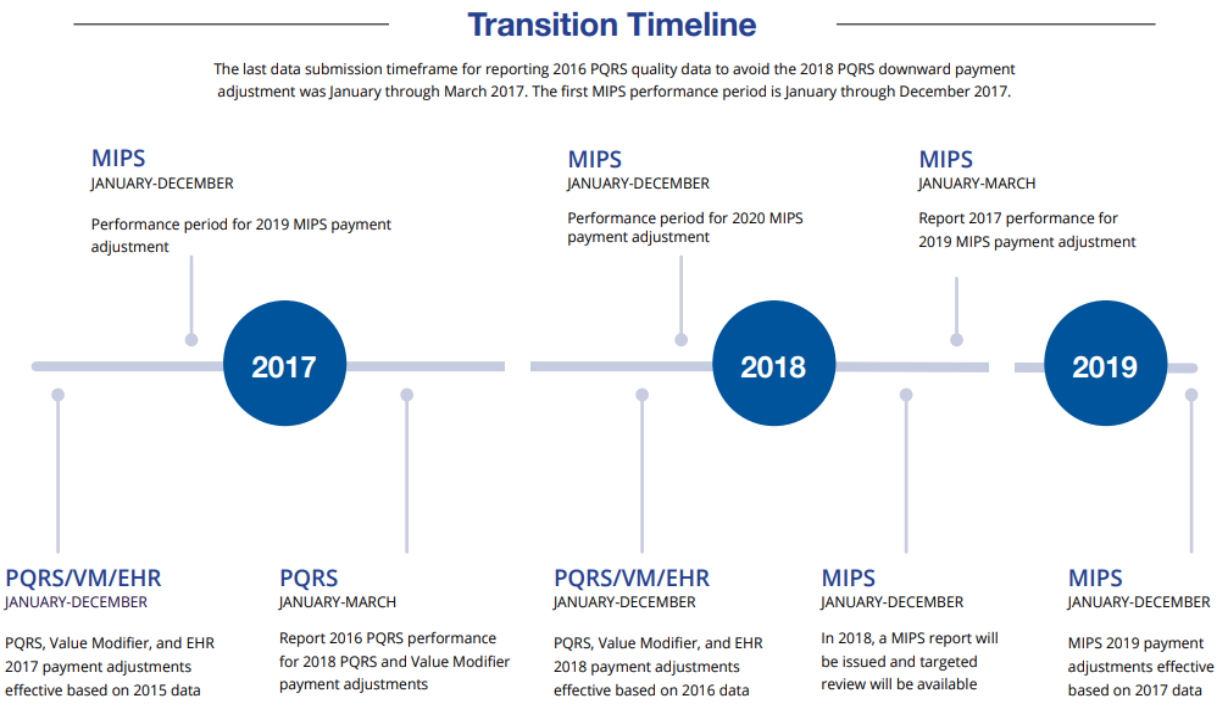


Figure 16: CMS Transition Timeline from PQRS to MIPS

Generally, the metrics focus on measuring the following areas:

1) Diagnoses (Dx):

    a. Coronary Artery Disease (CAD)

    b. Diabetes

    c. Heart Failure (HF)

    d. High Blood Pressure / Hypertension

    e. In Vitro Diagnostic (IVD): Use of Aspirin or Another Antithrombotic

    f. Depression Remission at Twelve Months

2) Ages:
    a. For patients with age between 50 to 75 years old: Colorectal Cancer Screening
    b. For women patients with age between 50 to 74 years old: Breast Cancer Screening
    c. For patients 65 years old and older: Fall Screening and Pneumonia Vaccination
    d. For patients 18 years old and older:
        i. Document Current Meds,
        ii. BMI Screening & Follow-Up Plan,
        iii. Tobacco use: Screening & Cessation Intervention
        iv. Screen for High BP & Follow-Up Plan.
    e. For patients 12 years old and older: Screen for Clinical Depression & Follow-Up Plan
    f. For patients 6 months old and older: Influenza Immunization
3) Age and diagnoses both combined:
    a. Statin Therapy for the Prevention & Treatment of Cardiovascular Disease for ages:
        i. For patients older than 21 years old, previously diagnosed with arteriosclerotic cardiovascular disease (ASCVD)
        ii. For patients older than 21 years old, the previous Fasting/Direct low-density-lipoprotein cholesterol - LDL-C ≥ 190 mg/dl
        iii. For patients between the age of 40 and 75 years old with Dx of Type 1 or 2 Diabetes Fasting/Direct LDL-C level of 70-189 mg/dL

Currently, the information is obtained by manually searching tens of thousands of structured, semi-structured, and unstructured documents in a short period every year between January and March. These documents contain an average of 4 to 5 pages that make hundreds of thousands of pages. This process involves several workers going through the pages via eye scan and manually looking for words/keywords based on CMS's metrics. After filling out the report with Yes/No on diagnosis and treatment, the provider submits the report before the deadline in the second or third week of March of the same year. CMS grades the report on accuracy for Continuum of Care and compares it against other providers on service to benchmark their performance measurements. For example, if the patient were not diagnosed with treatment, CMS would remove back bonuses, merits, or change reimbursement percentage. Therefore, this report is essential for providers in their income revenue.

The health provider gets the information from its EHR care system containing encrypted text notes (unstructured) and links to scanned images stored outside the EHR database. This case study will answer whether this process can be automated and require big data for data completeness?

Figure (17) shows the flow of events taken with the manual process compared to the automated process as shown in figure (18).

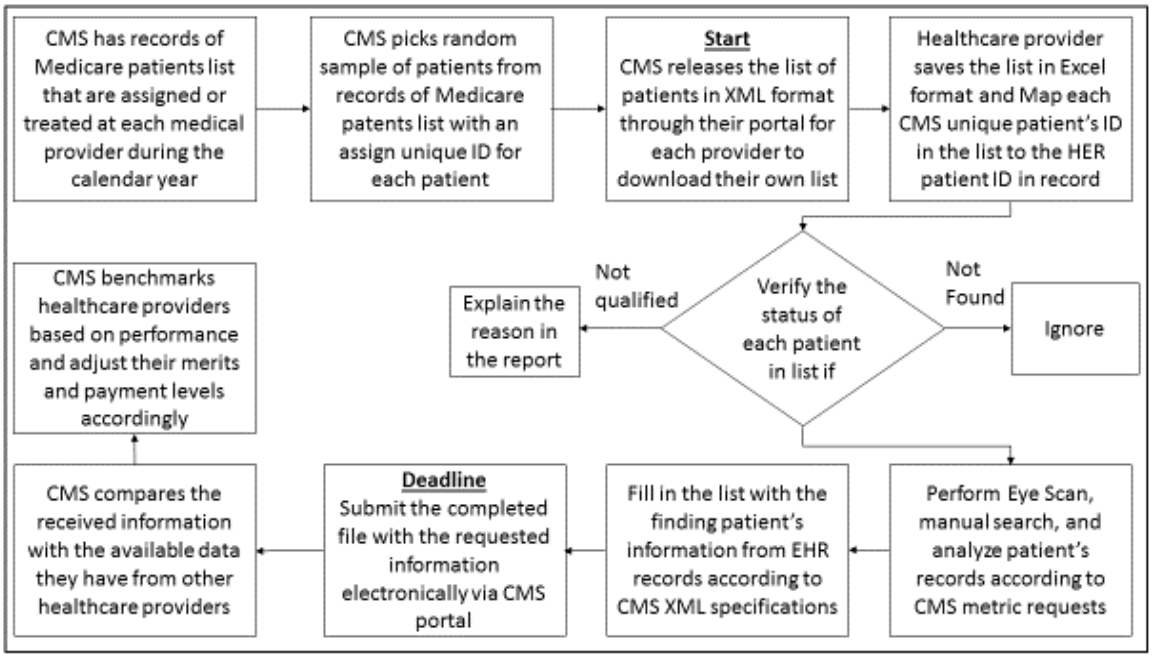


Figure 17: Manual Process Search

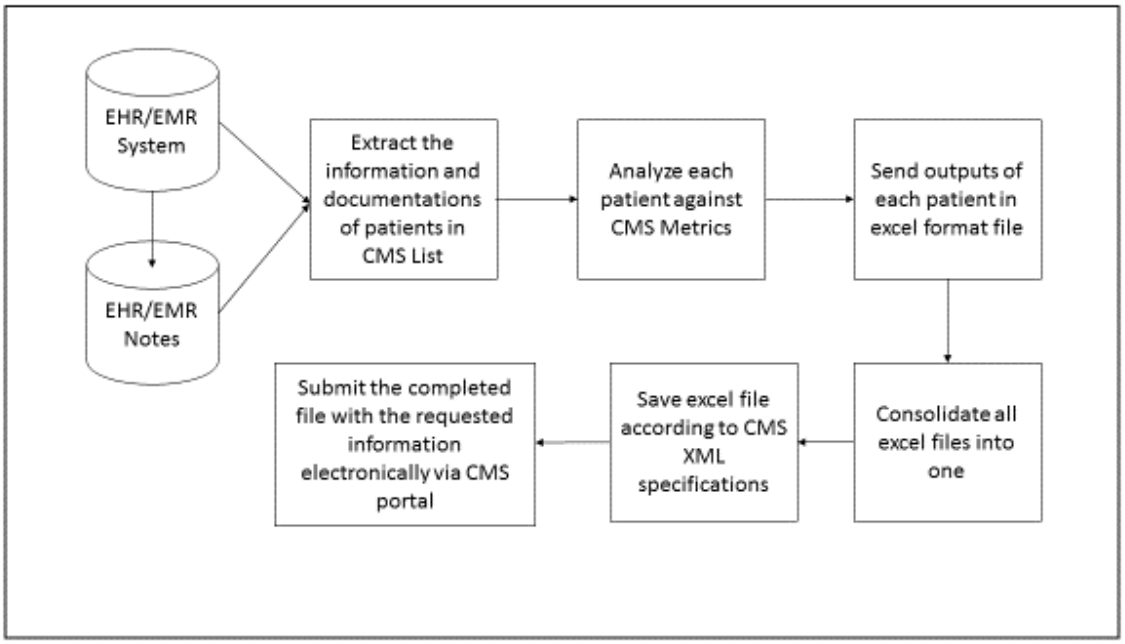


Figure 18: Automated Process Search

Validation and Lessons Learned

The validation of the model will be reached by comparing the outcomes of the model with the results of the manual process. First, however, the model's functionality will be tested operationally as a Proof of Concept (PoC). The accuracy will be verified to ensure that the outputs have the desired quality of information. Additionally, the solution's performance will be measured to reflect the speed and the time consumed to provide the requested information compared to achieving the same task manually.

Future Research

The future research of big data in healthcare will focus on building solutions that integrate all components of big data sources in visualization and cloud computing environments. Other research areas will be conducted in mHealth, wearable devices, and IoT to improve the monitoring capabilities that provide quality healthcare to patients. This research will allow personalized treatment and real-time tracking of physical activities. Moreover, patients' and providers' need to access health records from anywhere via smart devices will lead researchers to introduce a standard protocol to access data universally in a better virtual and infrastructure-less environment. In addition, cloud computing and data analytics will be good research areas that convert clinical informatics into knowledge for better healthcare quality. Finally, innovative security research will be needed to protect privacy when Knowledge as a Service (KaaS) is offered on a large scale.

Conclusion

The ability of the framework to answer many of the questions in utilizing big data in healthcare performance measurements is realized. In addition, it provides a better understanding of big data's impacts in the financial, legal, and technology spheres. There are unlimited opportunities to improve the healthcare system utilizing health information technology and big data tools. It is still in its early stages of implementation but employing information technology in the medical field will stop healthcare from holding up behind the rest of the USA economy. However, it is expected to happen as it did, like other industries that had challenges. Healthcare systems, clinical practices, USA government agencies, and insurance companies will benefit from sharing the big data analytic for better healthcare quality and reducing per capita costs and good services to the community they serve. Big data and predictive analytics will change healthcare organizations' decisions in both business and clinical areas, leading to reduced economic risk.

Big data in healthcare introduces issues in patient information privacy that needs to be addressed by the US government to regulate and organize the sharing of data. It is critical for healthcare organization to look at the big data and cloud computing from the perspective of management and marketing in order to gain value for their long-term commitments.

CHAPTER FOUR: RESEARCH FRAMEWORK

This chapter covers the “What” to achieve the goals of this research. What framework is required for healthcare organizations that want to use available data, including theirs and/or owned by others? The need to strategically adapt the framework to improve clinical delivery and better-quality outcomes in the digital transformation era. Industrial Engineering (I.E.) principles in constructing the framework specify the needed components and steps to integrate people, information, process, and technology into one system. Information Engineering is one of the fourteen areas listed in the IISE Body of Knowledge ("IISE BoK Website," 2021) that is primarily followed in forming the framework, along with other areas, to ensure employing best practices and guidelines for successful executions.

Introduction

The key to leading successful big data analytics for healthcare organizations is establishing an actionable approach with a descriptive structure that complies with their strategic imperatives and digital capabilities. This approach is vital to becoming a data analytics-driven healthcare organization that improves the performance of clinical, financial, and operational outcomes. The framework's objective is to extend the healthcare organization’s data ecosystem in an innovative way that leverages emerging technologies such as big data and integrates technological platforms and advanced analytics.

The study aims to develop and assess the framework by measuring the desired outcomes for healthcare organizations using big data analytics.

Healthcare Systems and Data Mapping

It is essential to understand the healthcare industry to construct the appropriate framework that directs the priorities and ventures of the healthcare organization to reach its goals. There are several departments in healthcare that are required to interact with each other to present the organization as a healthcare system.

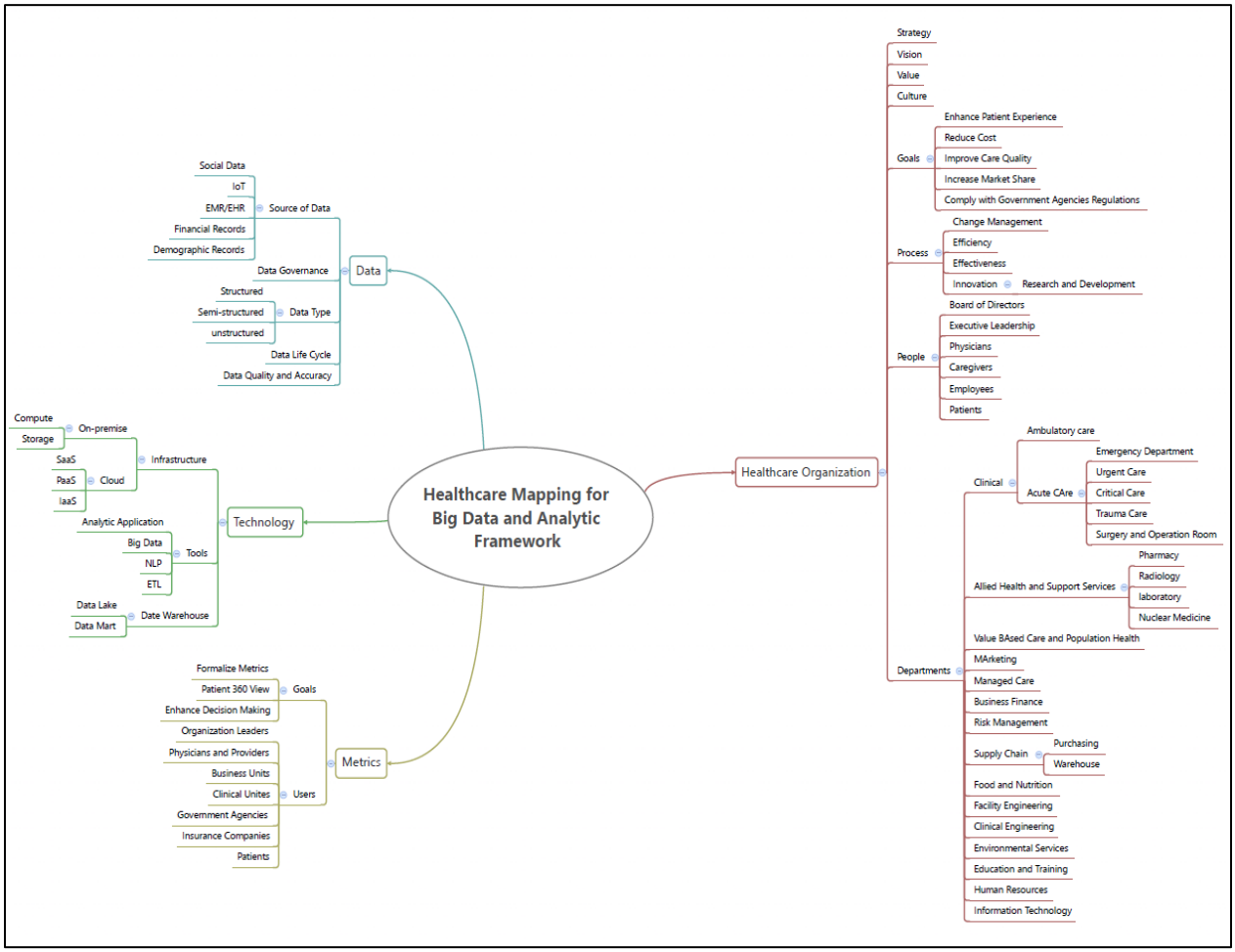


Figure 19: Healthcare Mapping for Big data Analytics

The mapping of the healthcare system in figure (19) illustrates the relationship of the necessary segments of the framework and their involvements in big data and analytics. At the organization level, the vision and the strategy are built around the patients. Not only to provide

the needed care during illness but also to keep them healthy. All departments in the healthcare system exist to serve patients in the clinical setting or on the business side. Many departments such as supply chain and information technology do not interact directly with patients, but they are crucial in daily operations. Data and technology are utilized for creating metrics that are available for internal and external users.

Healthcare Business

The modularity of the healthcare system and the orchestration among departments cannot be achieved without a robust ecosystem that supplies intuition of real-time activities. It starts with the leadership team laying the organization strategy by outlining the goals, the vision, and the value and planning, analyzing the market, and setting performance measurement targets.

As a result of the fourth Industrial revolution (I4.0), healthcare is moving towards a new paradigm with Healthcare 4.0 (H4.0), where healthcare business shifts from hospital-centered to patient-care focus (Tortorella et al., 2020).

Everything patient-related should be at the core of the strategy by employing technologies to enhance capabilities in its ecosystem. The organization's vision must endorse digitalization to benefit from all available analyzed data and be willing to invest intense capital and operational funds to deliver an excellent digital experience to data consumers such as clinicians, business units, and patients.

Therefore, data analytics should be one of the top strategic imperatives the organization's leadership wants to attain. The business objectives of adopting big data analytics must be communicated clearly to stakeholders to ensure they stand behind this objective. The organization’s value will be realized when data innovation is embraced. Neglecting it could cause the organization to lose market share and revenue. Thus, data analytics is a linchpin to

other emerging technologies such as AI/ML. When the data project is approved, listing the project goals requires reference to the organization's strategy to ensure alignment—securing funds and resources, defining the scope, and planning the project's schedule. Finance is an essential requirement for discernibility.

Additionally, designing processes to receive or provide data to and from departments within the healthcare system or other healthcare systems is a critical success. Process design is the center of the overall roadmap to transform the organization into a data-driven organization. It concentrates on the goals based on the organization's purpose, vision, and value defined at the organization level. When the data project starts, it should align with strategic planning to translate it into a tangible task. Measurable means should be identified at the organization, process, and data project levels to evaluate the success of the desired outcomes. The Balanced Scorecard (BSC) is a measurable means of financial and operational goals at the organization level. Key Performance Indicator (KPI) is another measurable method that validates the success of reaching business goals. At the process level, the aim is to measure the Process Capability (C_{PK}/P_{PK}) that determines if the specifications are met or not.

The healthcare business must revamp its processes around the new capability that digital transformation introduced for more effectiveness and efficiency. The revised processes should describe how each department will deal with the data it generates or uses and share it with other departments within the same organization and/or other healthcare organizations.

Table 3: Mapping Organization, Process and Data Project

	Goals	Design	Measurement
Organization	Purpose/Vision/Value	Strategic Plan	BSC / KPI
Process	Process Goals	Process Design	C_{PK}/P_{PK}
Data Project	Project Goals	Project Plan	Project Performance

Table (3) shows the mapping of organization, process, and data projects and how they relate to the goals, design, and measurement.

The Information Technology (IT) department is the leading player who implements the infrastructure and applications and controls and secures data access. IT is expected to act in response to business and regulatory obligations with technical solutions and products rapidly.

The Communications and Marketing departments can use the data to discover how the healthcare organization's community serves and gain their trust. Indeed, social media is one element that needs close attention to capitalize on the positive comments of real-time patient feedback and focus on correcting any negative remarks.

The Regulatory, Public Affairs, and Government Relations departments interface politicians and state governments who set legal rules and laws that regulate healthcare benefits to the state's citizens. Therefore, providing accurate data about the health of the community and patients is essential in establishing trust between the healthcare organization and local & state governments.

The Human Resources department needs to work with other departments to fulfill the needed number of employees. Insight from data analytics allows planning for adjusting the number of employees for each shift by hiring more employees or reassigning them to short-staffed areas.

Having Data is also a liability. The Legal department is expected to give guidance to avoid lawsuit cases against the healthcare organization. They establish the guidelines for the policies and procedures that are required by the law. They must know the impact of digitalization on legal counseling when using or sharing data of patients' health.

The Audit and Compliance department can employ data analytics in audits and fraud Prevention. They should work with IT Security to put Policies and Procedures to track patients' health information access to stop misuse of the data or sharing with other entities that may cause legal actions.

Business finance and budgeting are in charge of funding projects and initiatives. They control the spending of money. Therefore, their involvement is crucial to secure the necessary capital for the data analytics initiatives. They also use data analytics to forecast the financial plan for fiscal years based on growth and market share. The Account Payable department can gain value out of big data as many invoices are scanned as images and need manual labor for verification and auditing.

Supply Chain is another department that can benefit from big data in many forms. For example, in warehousing and inventory, using IoT technology allows locating parts and supplies very effectively. It accurately provides the available inventories so that no shortage or overstocking of medical supplies or medications occurs.

Revenue Management is the account receivable that oversees the process of invoicing patients for their medical billing. Big data that includes unstructured and semi-structured data analytics can improve the financial results and increase revenue. In addition, it can help in analyzing market access, foresee readmission, and benchmarking of efficiency and comparative analysis.

The Care Management department deals with insurance claims for reimbursements. Big data analytics simplifies the process of providing quality reports to Medicare and Medicate government agencies and preventing insurance frauds.

Clinical Engineering and Facility Engineering departments that support the medical equipment and buildings can monitor devices and perform predictive maintenance utilizing IoT technology that reports alarms from equipment malfunctions and issues with the temperature of drugs refrigerators.

Finally, the care delivery area is the primary beneficiary of big data analytics that helps deliver quality care to patients in many aspects. One of the areas is pattern recognition in image processing of MRI, X-ray, ultrasound using deep learning to discover abnormalities. Additionally, many healthcare organizations participate in clinical research in genetics, genomics, and precision medicine for drug development and innovations.

Stakeholders

Identifying stakeholders aims to assess and comprehend their significance, interactions, impacts, authority, and interests toward a successful initiative (Brugha, Varvasovszky, & planning, 2000).

There are several stakeholders involved in the process of data analytics. Those individuals or entities who provide the data, analyze the data, and receive the outcomes of the data analytics are the primary stakeholders. The healthcare organization leaders are consumers of the data analytics outcomes that allow them to determine their organization's performance and understand the market shares to engender business value.

Chief Medical Officer (CMO), Chief Quality Officer – Medicine (CQO-M), and Chief Nursing Officer (CNO) are the clinical leaders. They should be consulted on getting the data's desired outcomes and involved in the strategic data imperative.

Chief Financial Officer (CFO), Chief Operating Officer (COO), Chief Revenue Officer (CRO), and Chief Compliance & Ethics Officer are the leaders from the business side. They also need to be included in the discussion of the data strategy to give inputs of the data they can provide and what reports they desire.

Healthcare Regularity Agencies are external stakeholders such as the Centers for Medicare & Medicaid Services (CMS). For example, the Agency for Health Care Administration (AHCA) demands specific reports and inspections to certify healthcare organizations' processes, quality, and facilities to practice as healthcare providers. In addition, managed care and insurance companies request quality reports of their customers to provide health insurance.

The professional data team consists of several groups. It is supervised by the Chief Data Officer (CDO), whose main objective is to lead the healthcare organization to use data and emerging technology in analytics to produce value by enhancing clinical delivery and improving business performance. In addition, CDO guides the healthcare organization in its data strategy to incorporate digitalization and innovation to widen its ability to leverage data analytics to expand its ecosystem. The role of a CDO is to bridge data and business together to guarantee the alignment of goals and identify opportunities. Furthermore, CDO establishes communication channels with leadership and physicians to transform the organization's culture to value data by educating and sharing information.

Additionally, CDO drives the relationship with technology vendors and partners to select the appropriate technology platforms and software tools for data analytics and endorse following

best practices when implemented. Moreover, CDO conducts cross-functional data governance to attain data control and quality for accurate outcomes. Identifying the needed roles of data professionals is one of the important jobs of the CDO and helping them to develop their skillsets in the data era. The professional data teams include the data engineers whose main tasks are coding, developing, and verifying data quality end-to-end. The data scientists' task is to build algorithm models and use statistics to boost outcomes by extracting useful information from raw data. The data Business Analysis team's task is to analyze data reported from customers, businesses, and cross-department. They work with the business relationship management team to collect and analyze data according to business demands. Additionally, the data integration team that its primary function is to allow sharing data among systems.

Technology vendors and partners are identified as external stakeholders, and their support to accomplish the data strategy goals is critical. They are at the core of the framework in both the financial and technical aspects.

Clinicians, providers, physicians, and caregivers are the major suppliers. They enter patients' health information data into the EMR/EHR systems and use the analyzed data results to make medical decisions and/or conduct medical research.

Finally, patients are the target of the healthcare organization to satisfy to deliver quality healthcare and easy access of their health information and making appointments and paying their bills.

Process Mapping

Process mapping is a way to embody the procedures of the business in a visualizing technique to assist in comprehending the steps of the process. The purpose of process mapping is

for evaluation and clarity to improve the process and the ability to measure each action in the process (Klotz, Horman, Bi, Bechtel, & Management, 2008).

Process mapping is an important stage in creating the framework as it does layout the steps of the patient journey from start to finish. It shows the patient's involvement from the time treatment begins until the service of care is fully paid and how the sequence of activities of each department participated during the treatment (Trebble, Hansi, Hydes, Smith, & Baker, 2010).

Comprehensive Health Record (CHR) is the evolution of EHR/EMR that the leaders of the patient electronic records software development companies prefer to use as a new term instead of HER/EMR. In addition to the medical and health records, CHR includes essential collected data beyond the healthcare provider settings, such as social determination and behavioral care (Sullivan, 2017). Thus, the CHR term is used in the process mapping.

Several workstreams in the process mapping for outpatients, Emergency Department (ED) patients, and inpatients have similarities in their steps. During the scheduling and check-in/registration step, the types of collected data entail the patient accounting, demographics, and insurance information. During the treatment steps, patient medical records, diagnoses, and/or procedure codes are entered for each patient encounter and medical orders entry and results.

Moreover, revenue cycle management and external data also contribute to the system's data gathering or use. For example, data about payments, billing, bad debt, and collections are identified in the system as part of the revenue cycle management. In addition, external data from social media or produced by IoT and monitoring devices are also reflected in the process mapping as part of data sources.

The healthcare process mapping in figure (20) illustrates the relationship of all entities in the healthcare system and their involvements in generating big data that is being analyzed.

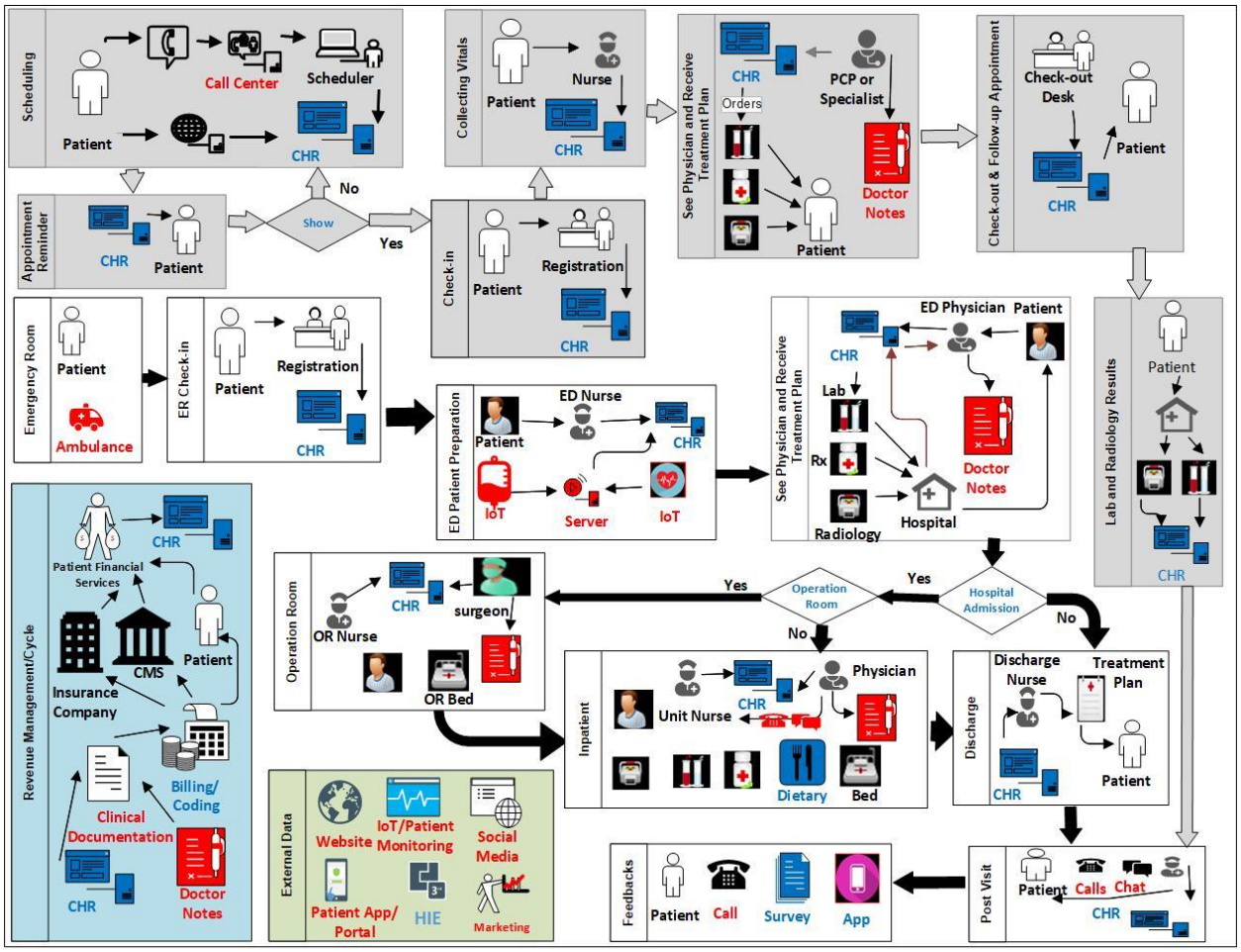


Figure 20: Process Map of Data Creation in Healthcare

SIPOC-R

Suppliers, Inputs, Processes, Outputs, Customers, and Requirements (SIPOC-R) is a quality tool that aids the understanding of the elements of a process and its requirements (Sarkar, Mukhopadhyay, & Ghosh, 2011). The process is the flow of actions that list the "HOWs," the inputs and outputs describe the WHATs, and the Suppliers and customers identify the WHOs (Parkash & Kaushik, 2011). Figure (21) shows the SIPOC-R diagram.

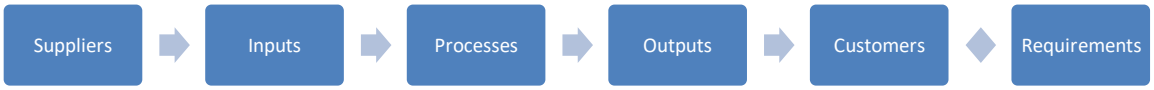


Figure 21: SIPOC-R Diagram

Understanding the demands of the suppliers and customers is essential for knowing the way the data is gathered and stored. Thus, SIPOC-R makes the digital journey easy to understand by identifying the suppliers and the customers of data involved in each step. Moreover, the inputs and the outputs of each step of the process determine the submitted information by the suppliers and data delivered to the customers. Requirements determine the needs of each step for successful execution.

SIPOC-R of data created for outpatient begins with scheduling the appointment to see a physician or a medical service such as lab work or radiology. Recording the call is considered unstructured data that can be analyzed to evaluate the customer satisfaction of overall service. Next, the patient or the guarantor provides the needed information, primarily structured data, whether he/she is a new or an existing patient to the scheduler or via an online application. Before the visit, the system reaches out to the patient for an appointment reminder and asks to

confirm, cancel, or reschedule the appointment. On the day of the visit, the patient information is verified and updated during the check-in process or flagged as no show, if the patient did not show. The show/no-show status is another data that can help the business to adjust the overbooking ratio. Additionally, knowing the cause of no-shows can allow the HC system to help patients who need transportation. Measuring the waiting time before the patient gets into the examination room is another data that can be easily collected to enhance the patient experience.

The system can measure each step's time in the evaluation and enhancement process. In the examination room, the nurse collects the vitals and enters them into the CHR system. The most critical data that most likely was not entered in the system is the conversation between the physician and the patient which, is usually documented in handwritten notes. This unstructured data is the main challenge for many HCOs as it is missed in the submitted claims to payers or CMS for payments. In this situation, SIPOC-R tools help to recognize this data and identify its location in the system. Physicians also enter the diagnoses code in the CHR system as structured data. When patients check out, they get the discharge paperwork with subscribed work orders for lab or radiology. They also can schedule a follow-up for a future appointment if needed. When performed outside the HCOs, lab results and radiology reports cannot be included in the CHR and be missed in the submitted claims. Therefore, a post-visit follow-up call with the patient is needed to ensure the directions from the physician are pursued. It is also an excellent opportunity to hear the patient's feedback on their experience during the visit. The follow-up call can be recorded and analyzed as part of big data analytics.

Figure (22) shows the SIPOC-R of data generated for outpatient.

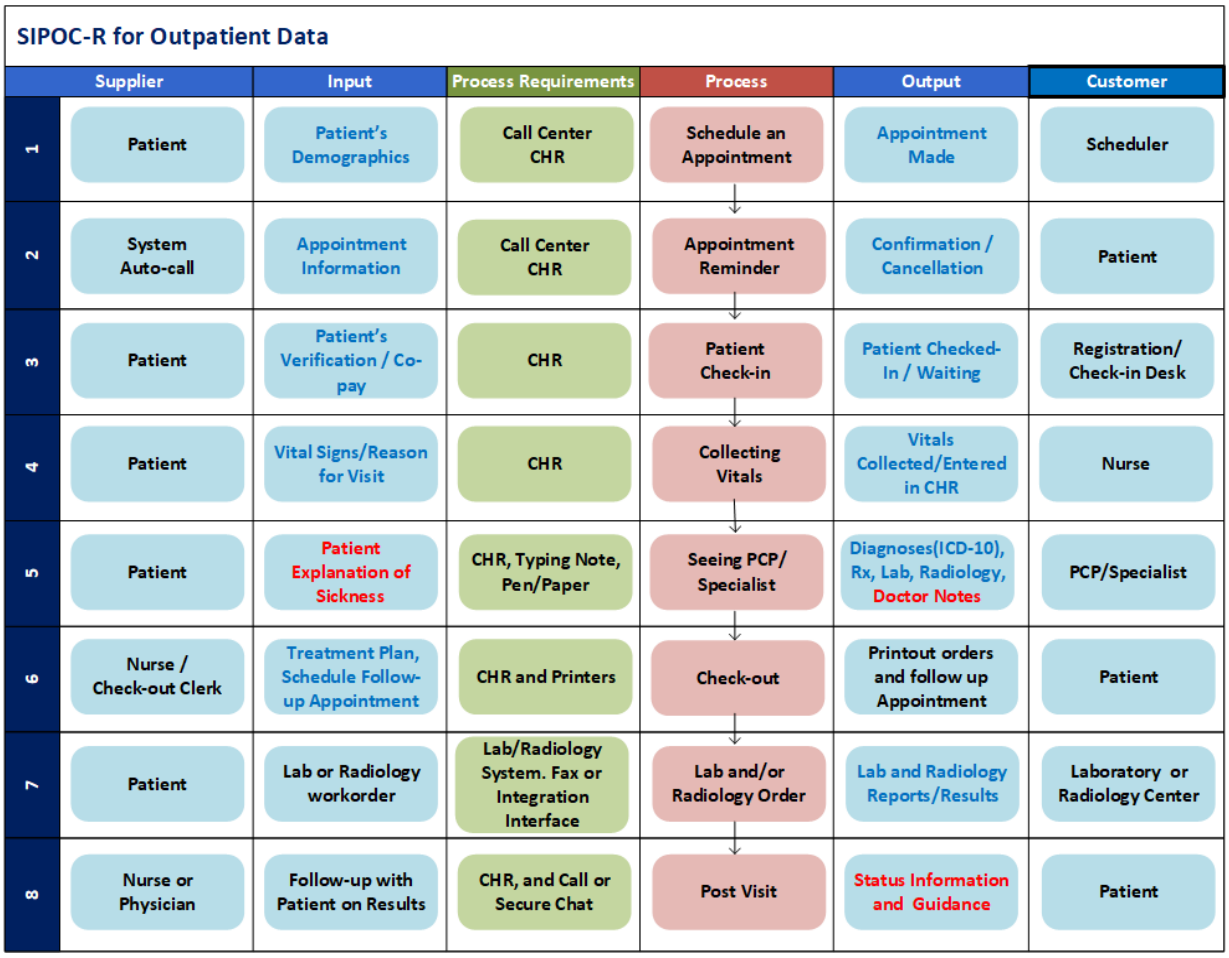


Figure 22: SIPOC-R of Data Created for Outpatient

Similarly, the SIPOC-R of data for ED has a minor modification. It does not need a prior scheduled appointment and can lead to admission as an inpatient or for surgery. Additionally, the IoT and the monitoring devices hooked to the patient feed the system with semi and unstructured data. Also, all the lab works, and radiology orders are performed internally, and external data using Health Information Exchange (HIE) from external sources can be reviewed for patient health history. Finally, the discharge process is more sophisticated than the outpatient process because HCOs could be penalized if the patient is discharged prematurely and was readmitted for the same illness within a short period from his/her first admission.

Figure (23) shows the SIPOC-R of creating data for ED patients.

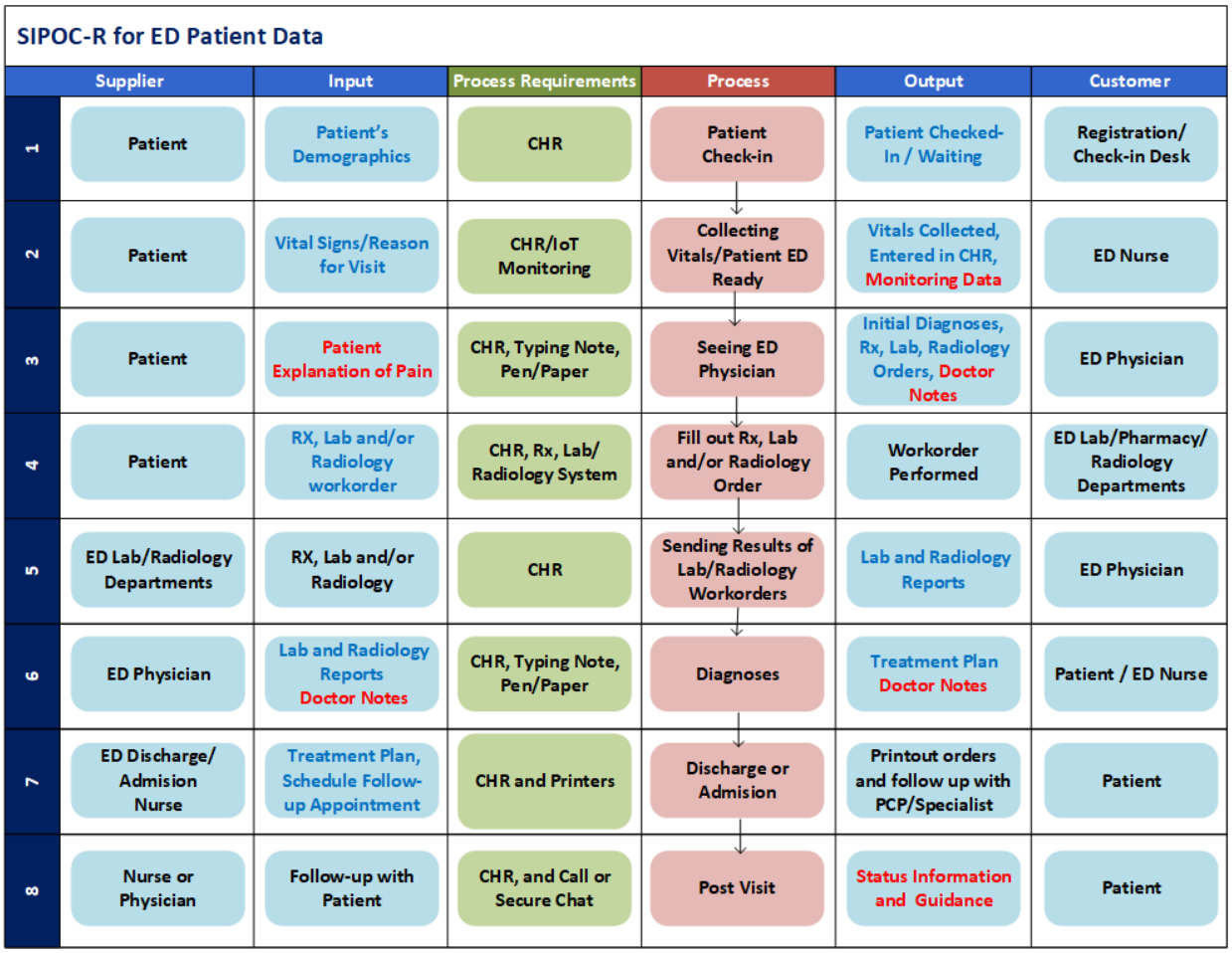


Figure 23: SIPOC-R of Data Created for ED Patient

For an inpatient, the SIPOC-R is a combination of outpatient and inpatient SIPOC-Rs, except that surgery is needed. Patients are admitted to the hospital with a physician order, scheduled surgery, or from the ED.

Figure (24) shows the SIPOC-R of Data for Inpatient.

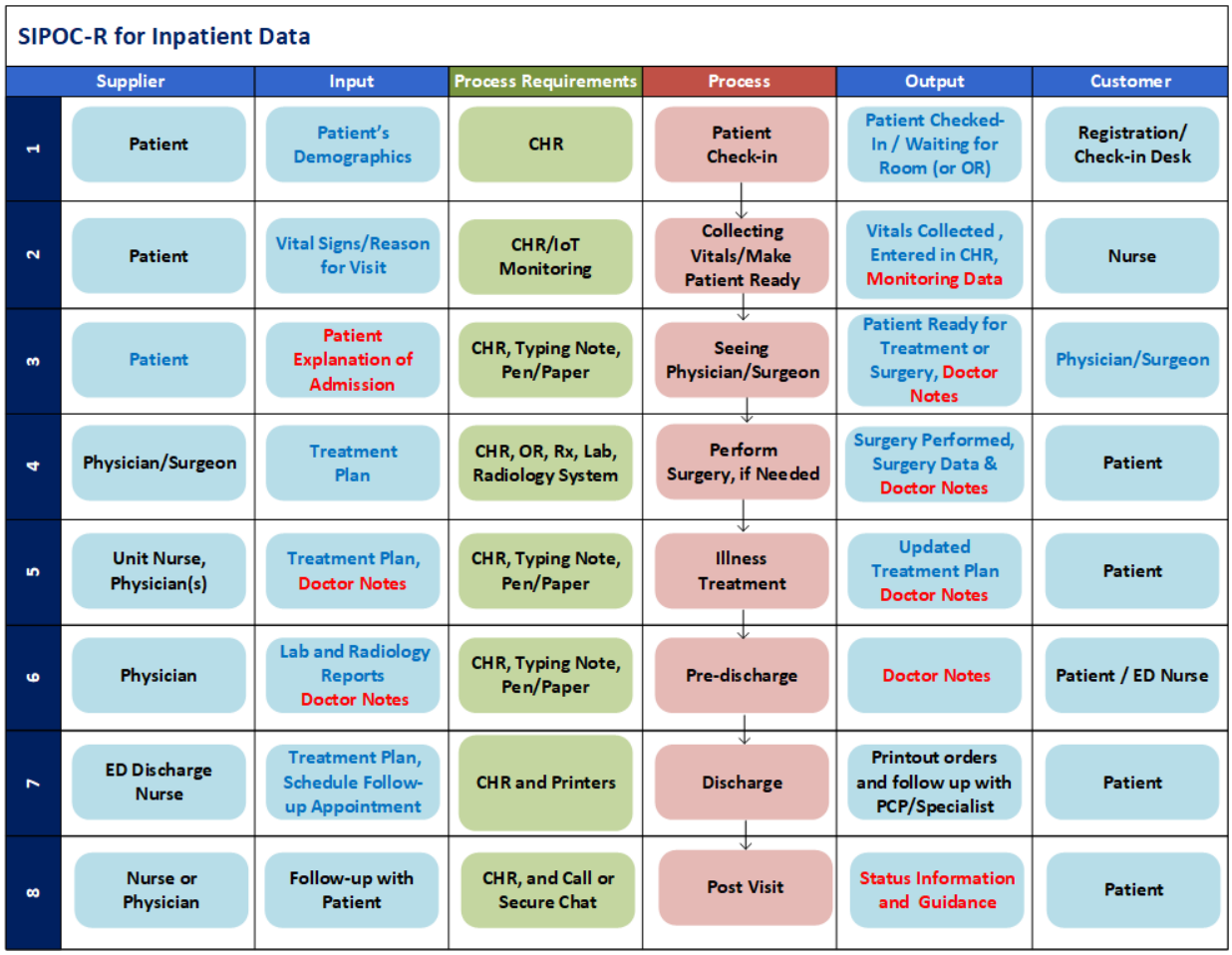


Figure 24: SIPOC-R of Data Created for Inpatient

Revenue cycle management is the most critical department in the HCOs. It is considered the leading consumer of the generated data in clinical documentation that brings income to the organization via submitting claims to CMS and health insurance payers. In addition to the unstructured data of handwritten notes and the structured data entry to the CHR system, specialists and caregivers use dictation using voice recognition software to describe the diagnoses of the case and the protocol used for treatment. Then a medical transcriptionist listens to the recorded voice and updates the written report accordingly. This process can be automated using big data analytics and voice recordings, and the written reports are considered semi-

structured and unstructured data. However, the clinical report is part of the clinical documentation needed for claim submissions, and it must be accurate to avoid denying the claim. It also needs to be complete with the correct treatment codes for proper payments. Moreover, the information included in the clinical documentation contributes to the quality reports submitted to payers and government agencies, as illustrated in the CLD of data and quality section. Finally, the patient financial service prepares invoices to collect payments and enter data into the CHR system as appropriate.

Figure (25) shows the SIPOC-R of each step of creating or consuming data for the revenue cycle management.

SIPOC-R for Revenue Cycle/Management Data

	Supplier	Input	Process Requirements	Process	Output	Customer
1	Physicians	Voice Dictations	Voice Recognition Software	Voice Recording to list Dignoses/ Treatments	Voice Recording	Voice Recognition Software
2	Voice Recognition Software	Voice Recording	CHR/Voice Recognition Software	Converting Physicians Dictations to Capture Items for Billing	Medical Written Report	Medical Transcriptionist
3	CHR / Computer Systems / External Health Records	CHR, HIE, Doctor Notes, Medical Reports	CHR, HIE, Written Note, Medical Reports, Clinical Documentation System	Organizing, Reviewing Medical Records with Correct Billing Code	Clinical Documentation with Captured Coded Items for Billing	Clinical Documentation Specialist
4	Clinical Documentation Specialist	Clinical Documentation with Captured Coded Items for Billing	CHR, Clinical Documentation System	Billing Preparation	Surgery Performed, Surgery Data & Doctor Notes	Patient Financial Services
5	Patient Financial Services	Treatment Plan, Doctor Notes	CHR, Mail, Electronic Data Transfer	Submitting Claim to Payers	Claim Sent	Insurance Company/ CMS
6	Patient Financial Services	Billing Statement	CHR, Mail, Calls	Collecting Payments	Billing Statement Sent	Patient
7	Patient /Insurance Company/ CMS	Send Payment	CHR	Make Payments	Payments Received and Update Account	Patient Financial Services

Figure 25: SIPOC-R of Data Created and Consumed during Revenue Management Cycle

Data in Healthcare

Digital transformation in healthcare started when the American Recovery and Reinvestment Act (ARRA) demanded that all healthcare organizations and providers fulfill the meaningful use requirements by implementing electronic medical records (EMR) in 2014 to sustain their current eligibility for Medicaid and Medicare patients. It followed the Health Information Technology for Economic and Clinical Health Act (HITECH Act) to encourage healthcare organizations to convert medical records from paper charts into digital forms by giving incentives when healthcare organizations procure certified EMR/EHR systems and follow the guidelines of privacy regulations and accuracy. Before that time, data was scattered in different systems with one capability for a specific area and labor-intensive process. EMR/EHR helped to integrate systems and eliminate the silos among them into one system. However, many physicians still use handwritten notes for their patients that cannot be entered in the EMT/EHR system as computerized data. Still, it is scanned and saved as an image, or they type a note in a plain text format that is saved within the EMR/EHR system. Nevertheless, both scanned and plain text is not in a structured format that can be analyzed. Therefore, emerging technologies were needed to take advantage of the new data, such as Big Data technology.

Adaptation of these new emerging technologies allowed linking events when triggered in one area to take actions in other areas, which helped eliminate the disconnect among departments. The efforts of digitalization in healthcare took another milestone step when HCOs started to exchange patient information for better and faster care through Health Information Exchange (HIE) within Region, State, or United States. HIE enables HCOs to share and exchange patient records electronically by participating in the Interorganizational Information System (IOIS) (Adjerid, Adler-Milstein, & Angst, 2018).

The goals of creating HIE are to prevent readmissions, stop prescription drug errors, enhance diagnoses, and reduce unnecessary replicate testing ("HealthIT.gov Website," 2020).

HCOs send patient records of Continuity of Care Documents (CCDs) to the HIE platform. The HIE platform provides the needed patient records to other HCOs via Cross-Enterprise Document Sharing (XDS). The diagram in figure (26) illustrates the HIE flows among HC entities.

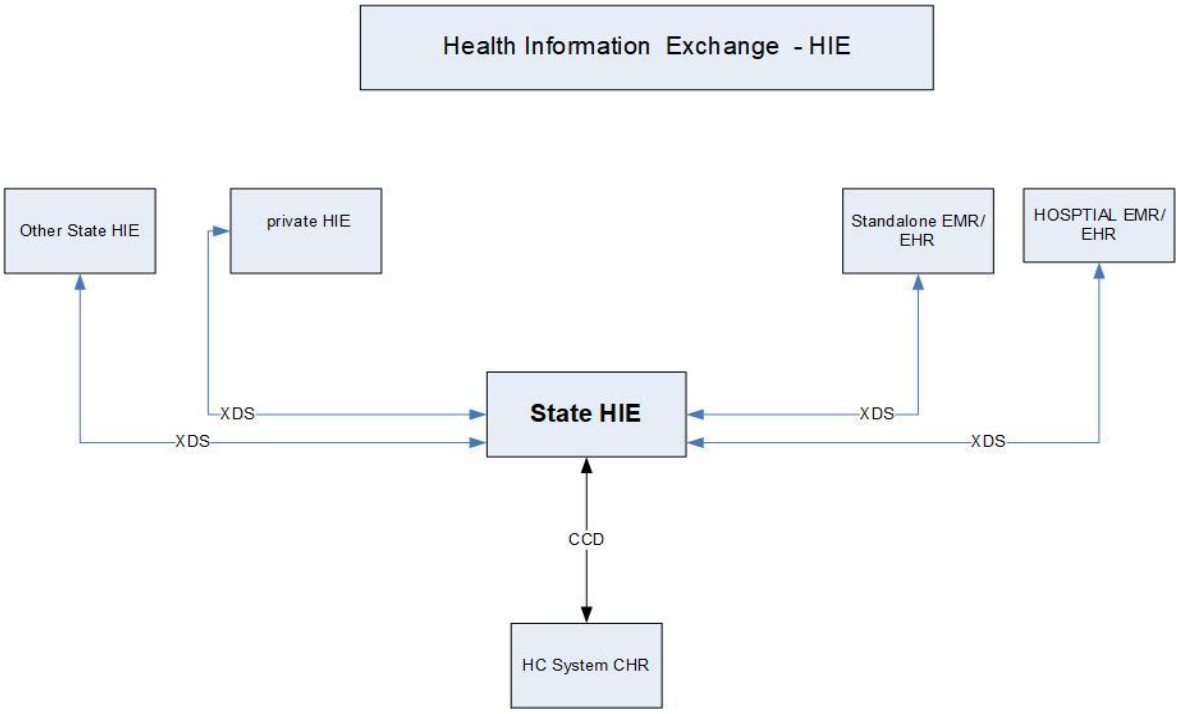


Figure 26: HIE Diagram of Sharing Data Electronically

Florida HIE offers four types of services:

1) The Encounter Notification Service (ENS): Send participants notification about the encounters of their covered members. More than 400 participating data suppliers in the ENS program currently cover 18 million in the state of Florida HIE, as shown in figure (27) ("Florida HIE Website," 2021).

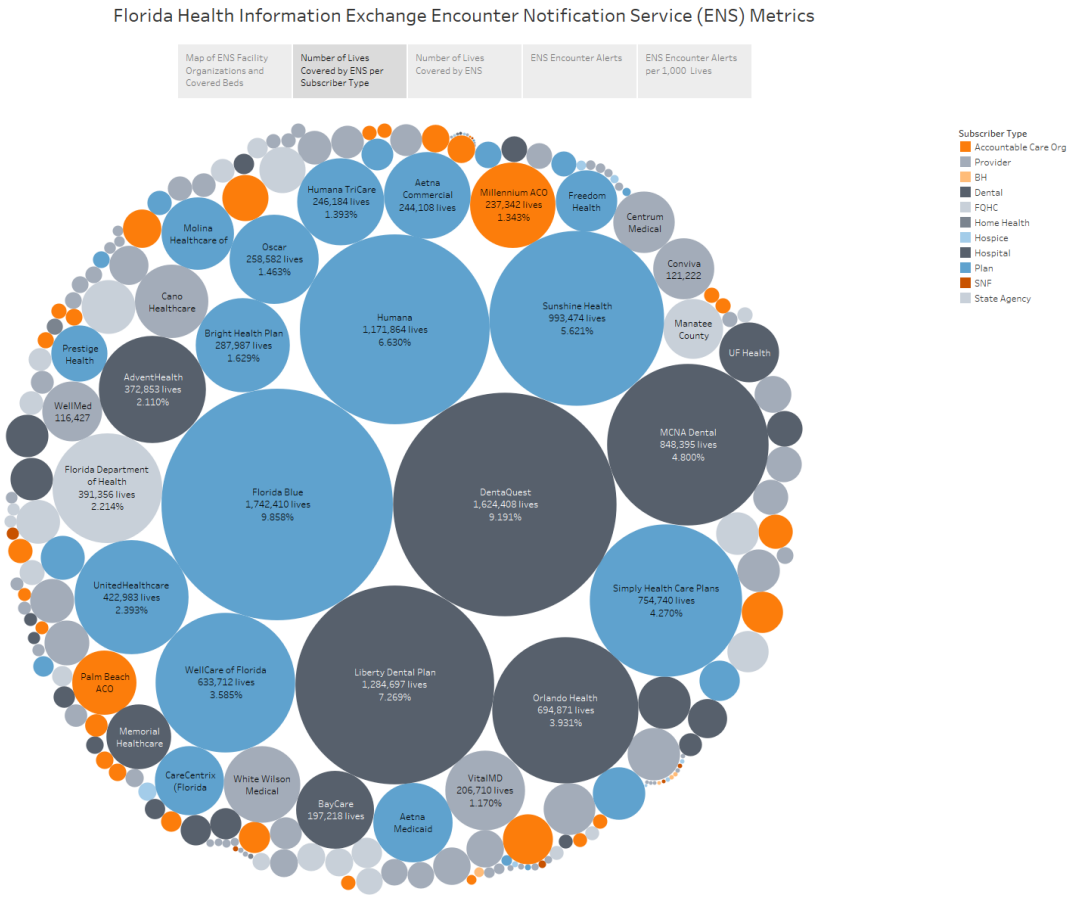


Figure 27: Florida HIE- Participants
Obtained from ("Florida HIE Website," 2021)

2) Proactive Management of Patient Transitions (PROMPT): a web-based interface with a dashboard that provides real-time notification.

3) Direct Messaging: A secure messaging system integrated with HER using American National Standard Institute (ANSI) to address interoperability among systems.

4) Query Solutions: it is a searchable system to find patients' health records across other data sources locally, in the state, or nationally using CCD.

Data from HIE sources can be structured, semi-structured, and unstructured formats.

Informatics for Integrating Biology & the bedside: i2b2 (https://www.i2b2.org/) for structured dataset and National NLP Clinical Challenges: n2c2 that works on the challenges of the unstructured dataset is an example of other organizations.

OneFlorida ib2b formed of 9 HC systems with 4,100 providers, 22 hospitals, and 1240 clinics in 67 counties that serve about 15 million Floridians, which is 74% of Florida population ("OneFlorida i2b2 Data Guide," 2018). Figure (28) shows the HCOs that participate in OneFlorida.

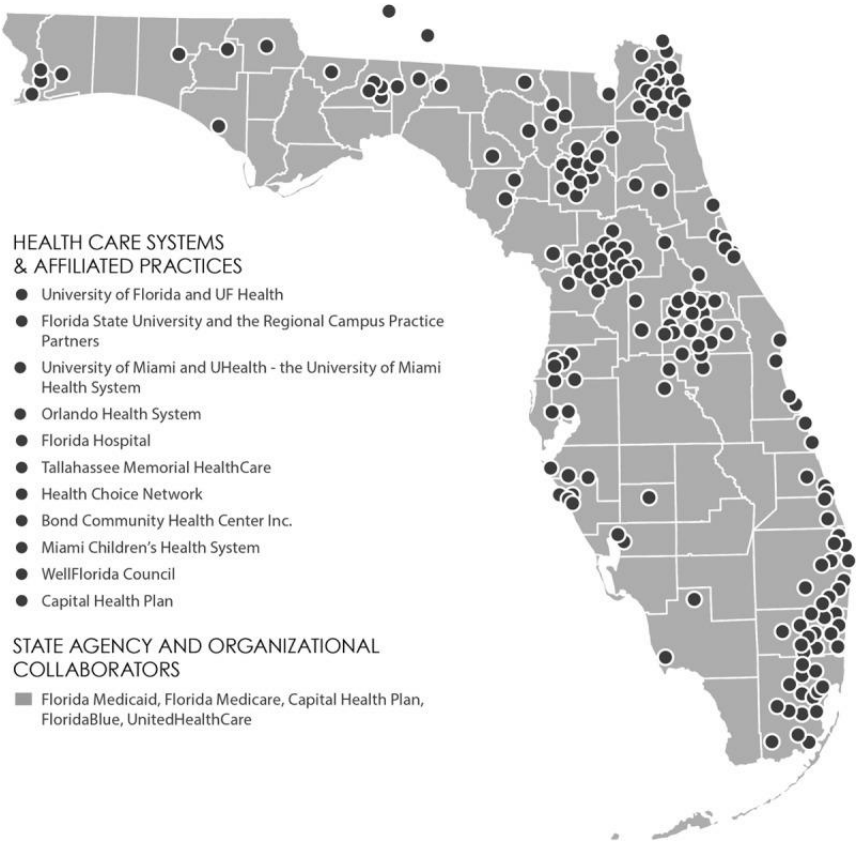


Figure 28: Oneflorida i2b2 participants
Adapted from OneFlorida Clinical Research Consortium: Linking a Clinical and Translational Science Institute With a Community-Based Distributive Medical Education Model (Shenkman et al., 2018)

Figure (29) shows the data and access of OneFlorida.

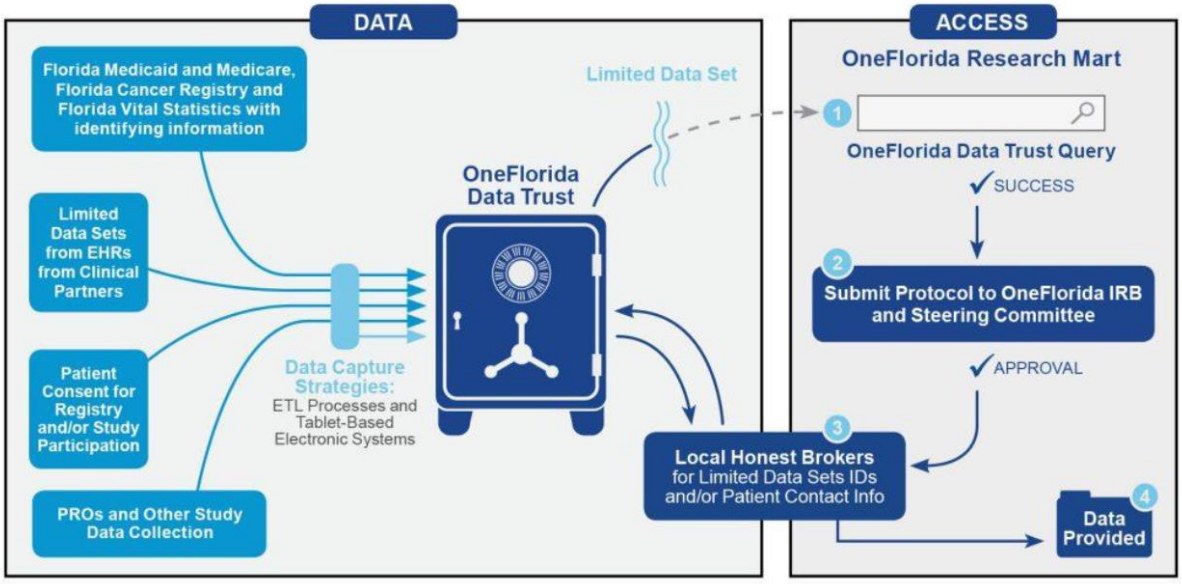


Figure 29: OneFlorida i2b2 Data and Access
Adapted from OneFlorida i2b2 Data Guide.

Source of Data in Healthcare

Data completeness in healthcare comes from several sources, and each system has its data mart that makes the data ready for analytics. However, the primary data source is the EMR/EHR system that contains demographic information about the patient, such as a home address, age, gender, race, etc.…

It also contains blood pressure, height, and weight. In addition, diagnoses of the diseases based on ICD-9 and ICD-10 classification codes, medications, and lab results are entered in the EMR/EHR system.

Another data source is Enterprise Resource Management (ERP). This system consolidates business applications in one software suite that manages the functionality of the business in finance, accounting, human resources (HR), operations, and administrative tasks.

Figure (30) shows the systems that supply data to the EDW.

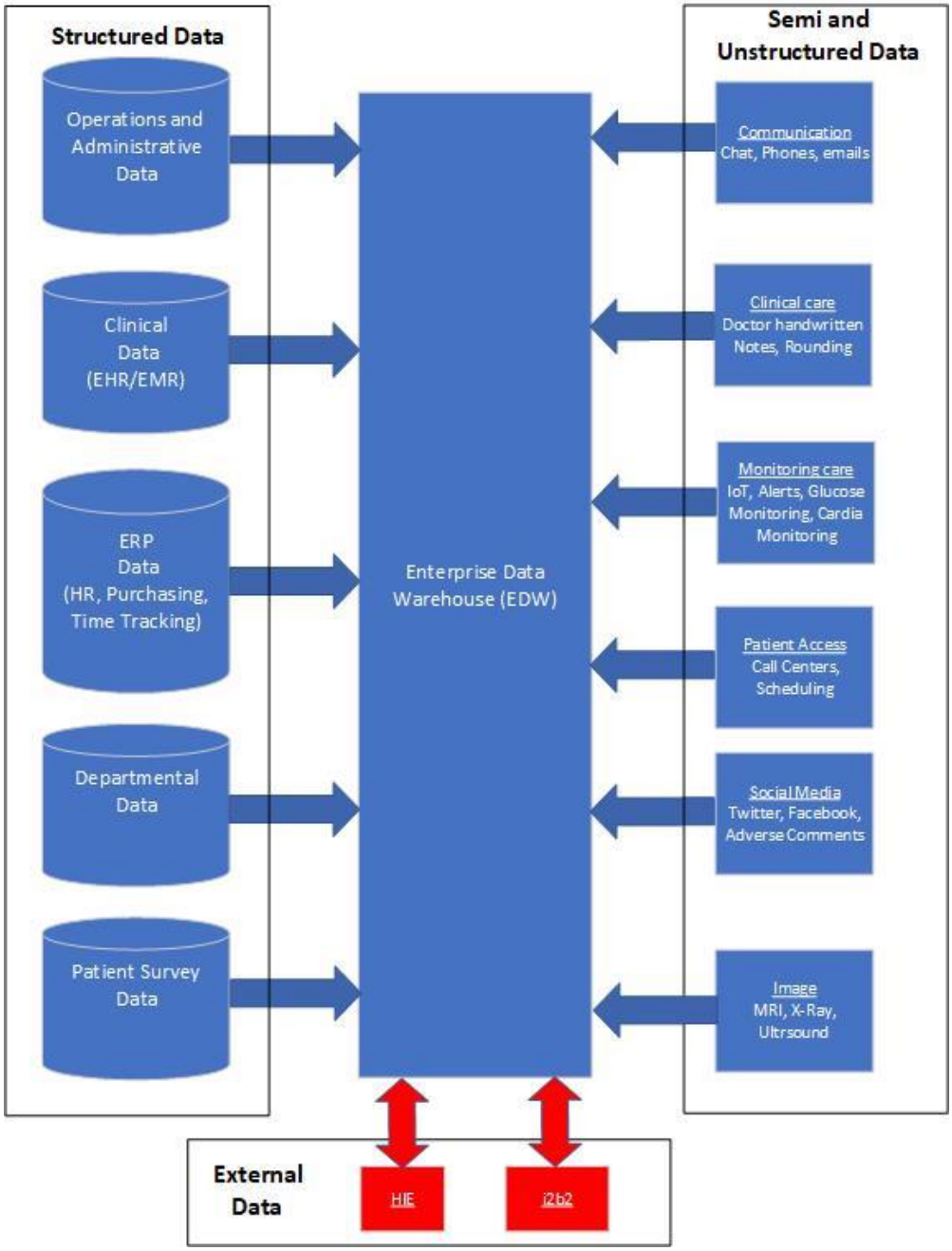


Figure 30: Source of Data in Healthcare

Big Data and Quality Improvement Relationship in Healthcare

Quality metrics need data, and the more data incorporated the more insight. Quality 4.0 is the digitalization of traditional quality management, and big data is an essential element to

improve the quality of design, service quality, and conformance quality (Sony, Antony, & Douglas, 2020). Therefore, many healthcare organizations have started to find a way to pursue big data analytics for comprehensive quality measurement.

Most healthcare providers and organizations face a significant challenge lacking technology and insufficient data to create performance measurement reports for government agencies for accreditations and payments linked to performance. Additionally, The performance reports are publicly published for transparency, and poor outcomes can cause a bad reputation that leads to dropping market share and lessening business (Nelson et al., 2005). Big data technology can successfully address the challenge of inadequate data that allows for good quality reports. For example, the quality reports' performance measures are the mortality, readmission, patient experience, and efficiency and effectiveness of care. Therefore, big data is a crucial driver to deliver performance reports with high accuracy and complete information.

At the hospital level, in 2016, CMS revealed the star rating program. This is an easy-to-understand system based on the level of quality care, with a range of one star as the lowest rating to five stars as the highest rating based on the hospital’s performance of specific quality metrics (Hu & Nerenz, 2020).

At the provider practices level, CMS implemented the Medicare Shared Savings Program (MSSP) in 2012 as part of the Accountable Care Organization (ACO) model to promote collaboration among providers and specialists to share bonus reimbursement based on overall saving because of the superior quality care (Ouayogodé, Colla, & Lewis, 2017).

McCurdy (2001) enumerated HC Accreditation Organizations to compare their accreditation programs and the target HC entities. The study indicated that the number of HCOs requesting accreditation is continually rising to satisfy the requirements of the state and

government agencies to maintain their care certification. It also illustrated the accreditation conditions of quality performance for accepting HCOs under their accreditation program umbrella. The National Committee on Quality Assurance (NCQA), Joint Commission on Accreditation of Healthcare Organizations (JCAHO), and the American Accreditation Healthcare Commission (URAC) are the three major accreditation institutes. All operate under not-for-profit organization status.

Other institutes offer high-performance awards to extraordinary HCOs in their high-quality achievements by benchmarking and adopting innovations using emerging data technologies during their digitalization transformations journey. HCOs receive awards by being ranked high from institutes such IBM Watson Top 100 hospitals, Leapfrog grade, US News & World Report rankings, and HIMSS Most Wired as a prestige recognition that positions them as exceptional leaders in the HC industry.

Emergency Care Research Institute (ECRI) is another independent, not-for-profit organization that serves as an evaluator of patient safety, evidence-based medicine, and technology decision support that lead to more safety, better efficiency, and higher effectiveness HC ("ECRI Website," 2021).

Causal Loop Diagram (CLD) of Big Data and Quality

The dense and vigorous condition of the processes within the healthcare ecosystem combined with the data life cycle can lead the efforts to attain high-quality outcomes with the anticipations of increasing the revenue and market share.

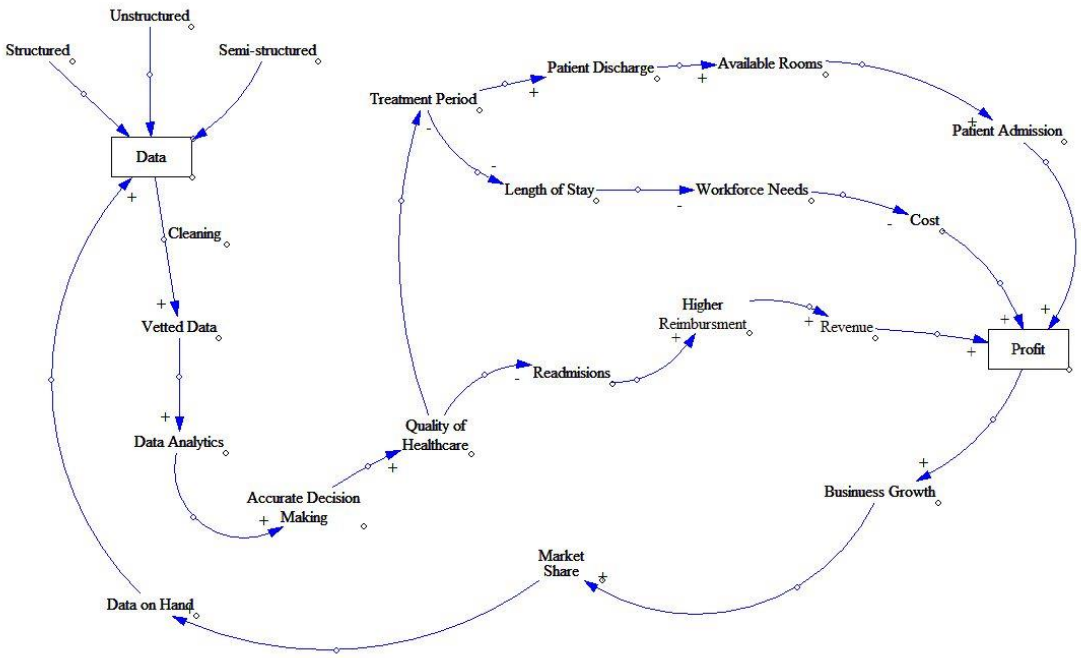


Figure 31: CDL Data, Quality & Profit

The CLD, shown in Figure (31), exemplifies the relationships among data, quality, and profit in healthcare. More data provides extra perceptiveness that helps both clinical and business sectors make better decisions. On the business side, it is a tangible method to control the operation expenditures by avoiding overstaffing and projecting the appropriate number of workers based on needs. On the clinical side, incomplete patient data can cause delays in diagnosing illness and providing the right treatments promptly that can likely alleviate the length of stay.

Data incompleteness is a crucial challenge for healthcare organizations when they submit quality reports to government agencies and insurance companies about patients, they treat by failing to get reimbursement. Therefore, technology such as big data is valuable in process automation, reducing the manual procedure and the person-hours needed in creating quality reports.

Big Data Tools

Big data platform analytics is one of the foundational digital capabilities of systems or tools for improving outcomes.

It is expected that many healthcare organizations have many numbers of both embedded analytics and standalone analytics applications. Many are duplicates as they were purchased in ad-hoc fashion based on departmental needs. This approach would not scale, and it would cause a negative financial impact on the business without creating any business value. The interoperability of data analytics tools must be considered when picking partners and big data technologies to avoid creating a big clutter.

Criteria of Selection

What criteria should be applied to select the appropriate big data and analytics tools to synthesize end-to-end solutions? First, it necessitates various tools that can be integrated. Thereby, multiple tools are a necessity to meet requirements to deliver big data analytics. This depends on several aspects, but it is mainly related to the case-by-case scenario and if the desired analysis is traditional BI analytics such as descriptive and predictive analytics or advanced prescriptive analytics.

The advanced big data analytics architecture must run in a parallel computing environment with many processors to deliver real-time analytics. In addition, it should have agility, elasticity, scalability, availability, and security features to accommodate business needs and innovation.

To decide on the essential functionality of the tools, the MoSCoW model is used to rank the criteria requirements that categorize them into classes:

M: Must have as the required criteria

S: Should have identified the preferred criteria

C: Could have as optional criteria if it does not cause impact

W: Won't have it now but would like to have it later.

This method of selection allows the prioritization and helps understand the need to choose the adequate tools (Alshehri & Benedicenti, 2013)

The criteria for big data and analytics tools are listed in the table (4).

Table 4: Criteria of Selection for Big Data and Analytics

Criteria	Must-Have	Should Have	Could Have	Won't Have
Data Catalog	X			
ETL/ELT	X			
Data Modeling	X			
Meta Data Management	X			
Data Preparation	X			
Data Integration	X			
Connectivity to Source of Data		X		
Data Analytics	X			
Advanced Analytics	X			
Same Platform for Data and Big Data				X
Visualization	X			
Reporting	X			
Automation		X		
Security	X			
Cloud			X	

Other measures should also be considered when selecting the big data and analytics tools to help implement a successful solution. These include speed and efficiency, ease of use, ease of creation, and ease of integration available in cloud technology.

The function of the tools

The function of the tools should synthesize a solution that works together collectively into one enterprise platform that is managed and governed to deliver value.

Security and HIPAA privacy rules are the deal-breaker in selecting the tools that comply with their requirements. The tools should have the ability to log all the activities for auditing purposes. In addition, all data communication from the data source must be encrypted and only allow permission to access for authorized users.

Built-in data integration to enable importing Comma Separated Value (CSV) and XML format files and the ability to perform multithread data loading are essential functions with the flexibility for data preparation to allow renaming, dividing, or creating columns. In addition, it is recommended to include a data masking feature to hide or encrypt confidential data.

Data storage management function capable of dealing with different data types such as documents, text, log files, images, audio, and video is required to provide the information about the source of data in detail.

The tools need to support Application Programming Interface (API) to allow applications to interact together and Software Development Kit (SDK) that includes a set of modules and instructions to develop applications software. This function delivers a user portal to the enterprise in web services and mobile application formats.

The database function manages multi-services for SQL and NoSQL for big data and no-ETL analytics in real-time. In addition to online analytical processing (OLAP) connectivity, it is required for the tools to connect to HDFS and text files via third-party product or Apache Hive/Spark.

Moreover, it must be capable of supporting data extraction for relational queries via Open Database Connectivity (ODBC) and Java Database Connectivity (JDBC) drivers.

One tool function as the data pipeline to handle activities in the data factory while generating ETL and ELT jobs and workflow. In addition, it is used to clean and ingest data sets to make them ready for processing in a group instead of running them separately.

Analytics tools are a crucial component, and they need to be compatible with the storage and data lake used in the platform and solution. The outcomes of the analytics tools are presented using a visualization tool with an easy way published in reports or dashboards. In addition, dashboards present the results in several interactive chart formatting such as histogram, box plot, heat map, and tree map.

Proposed Framework

Developing a framework of big data analytics for healthcare enterprises incorporates the people, process, and technology model as an excellent foundation to add on top of another three elements: the organization, the data, and the outcomes. As shown in figure (32), the relationship of all six frameworks components would lead to automation, standardization, and innovation for successful outcomes.

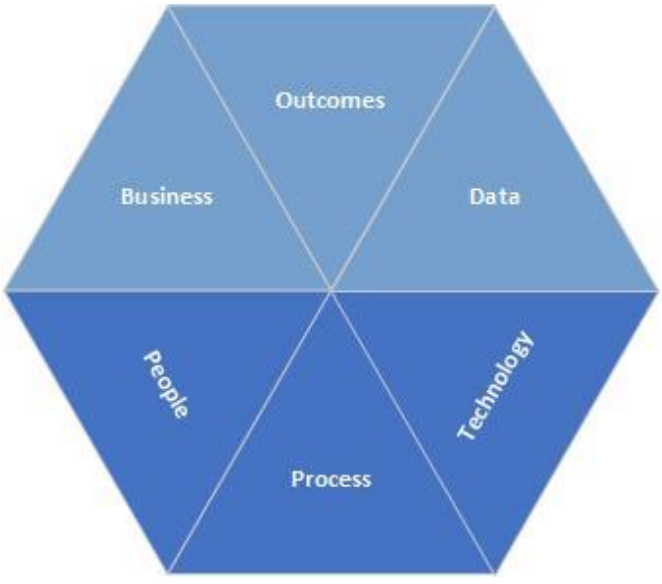


Figure 32: Foundation of the Framework

People are defined as all the stakeholders, such as leadership, employees, and customers, who turn medical information into value that serves clients and meets their needs. Technology entails using computers and applications to collect and save customer data for analysis to improve outcomes to customer needs with accurate information in real-time (Almotairi, 2009).

The proposed framework includes a source of data, data governance, processes, data life cycle, defining metrics/KPIs, and determining the criteria for selecting the proper analytics tools. Furthermore, the framework contains information on the appropriate infrastructure of either hosting in the cloud or on-premises. Identifying skill sets and resources is also part of the proposed framework.

Additionally, Cates, Gill, and Zeituny (2005) showed the amalgamation of people, process, and technology in their model of "The Ladder of Business Intelligence (LOBI)" that presented six levels of maturity: Facts, Data, Information, Knowledge, understanding, and Enabled Intuition. Thus, the framework aims to reach the level of enabled intuition by using the data completeness to provide directions to the future state of healthcare organizations to mature their usage of data and enhance their existing capabilities.

To fully realize the capabilities of big data, implementing a tool to analyze data is not sufficient. Still, a framework that blends people, processes, and technology is necessary to achieve value. Technical people with the right skills, knowledge, and competencies are critical elements in the framework. In addition, the correct process of managing data that addresses how to govern the data and its lifecycle are essential components to control the domains in which healthcare organizations have a wide range of available data. From a technological perspective, it is about tools, platforms, and technology infrastructure. Therefore, the proper framework is not data or technology centric. Still, it should be an excellent combination of these components

where if it is implemented, it allows HCOs to have the ability to innovate sustainably, help them to target specific opportunities, and drive their strategic imperatives.

The framework connects all components and offers a simple landscape that is more powerful and adds value to take advantage of big data capabilities, 3Vs as an example. The most powerful advances of the framework are emphasizing patient care and receiving data in real-time in a centralized location. This emphasis must be done without separating forms of data as these are considered data capture vehicles as well as data delivery vehicles, whether it's off an EMR entry, a wearable patient device, telehealth, or a patient monitoring data that are all considered sources of big data which, we use to do analytics against and extract knowledge in real-time to the clinical and business areas.

Additionally, the framework needs to ensure the inclusiveness of the data from all areas and data quality through assurance mechanisms to give the organization a level of confidence and trust in the coming outputs of metrics and KPI. Therefore, having data governance that defines data ownership and deals with the accuracy of the data and when it is obsolete is an essential element in the framework.

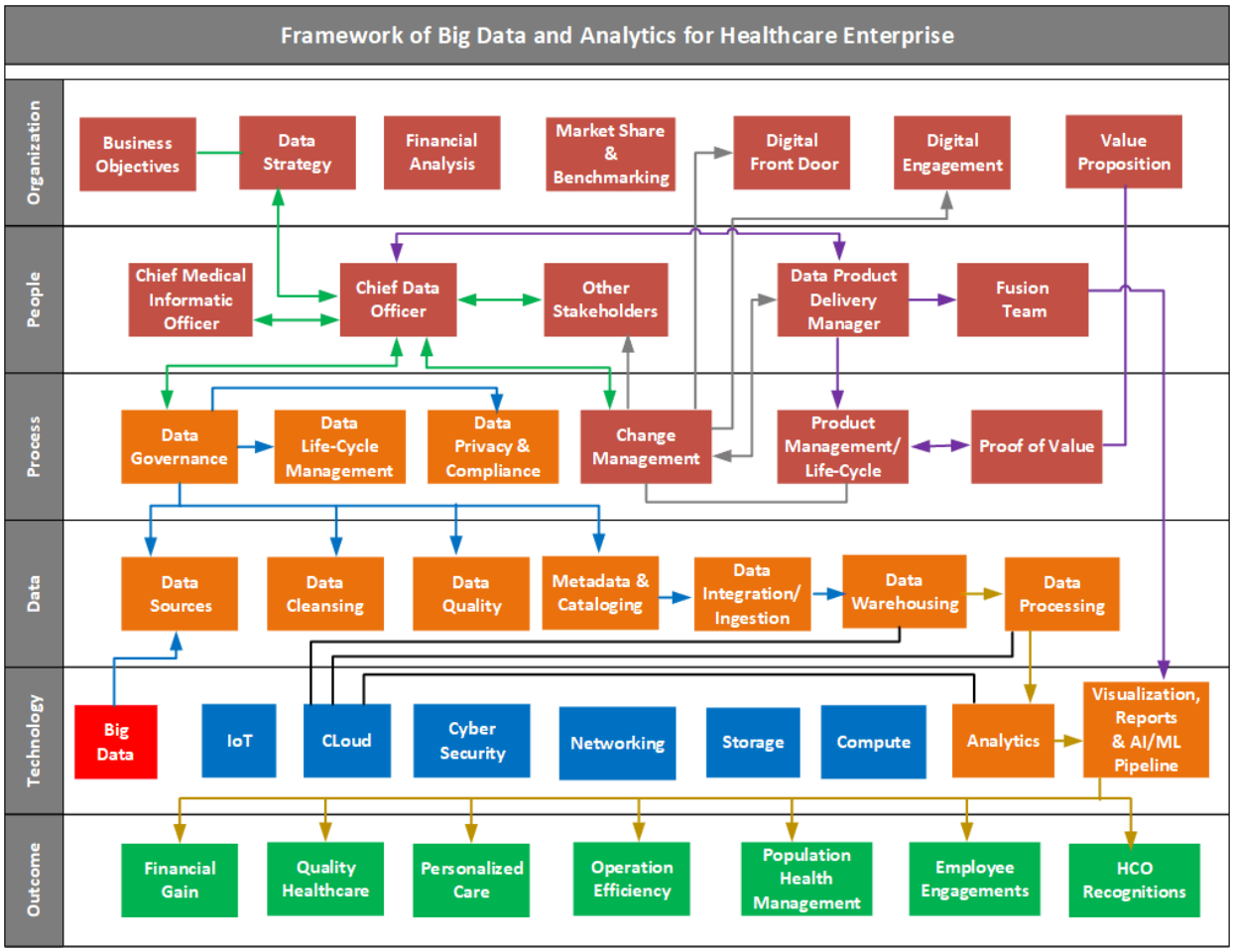


Figure 33: Proposed Big Data and Analytics Framework for Healthcare Enterprise

The proposed framework shown in figure (33) includes 41 essential entities categorized in 6 different areas: Organization, People, Processes, Data, Technology, and Outcomes.

Each entity represents a function or a role, and some entities are linked to one or more entities. However, the standalone entities that are not linked to any other are still crucial to the framework for a complete view. The linked entities are connected with a straight line to show the linkage between them. The one-way or the two-way arrow indicates an interaction between the two entities in one direction. Additionally, each color of the line or the arrow implies a relationship between several entities that interact together in several but similar channels.

The five-colored blocks display the entities in groups to construct a cluster of a distinguished purpose. The significant data entity in the red block is the main component in the framework, and it is under the technology area. The orange blocks are the core of the framework that places all data and analytics entities in one group, and they are distributed among processes, data, and technology areas. The blue blocks show the needed technologies that involve in utilizing big data and analytics. The brown blocks are the supportive entities in the framework that guarantee a successful result. Finally, the green blocks are the outcomes that show the expected value of adopting big data analytics. The proposed framework is very refined, and each entity encompasses a vast amount of work and resources.

At the organization layer, HCOs need to determine the overall business objectives and how the data strategy allows getting the most out of data by leveraging capabilities to drive and deliver the HCO's business goals.

Financial analysis is needed to ascertain that the financial gains and the Total Cost of Ownership (TCO) are straightforward. Also, it helps identify the impact in business using the what-if scenario of implementing big data analytics or not.

Understanding the HCO's position in the healthcare market it serves is essential for better business planning and benchmarking. In addition, it is recommended to use SWOT analysis for big data and analytics.

Digital front door enables HCOs to be engaged with their patients from when they ask for care to when their care journey is completed. In addition, Digital engagement provides patients with transparency on the cost of the needed healthcare services like other industries and helps take advantage of social media.

The value proposition is to trend towards inclusion improvements in certain areas to increase AI and Robotic Process Automation (RPA) to position the HCOs for the future and achieve their business goals. This should minimize complexity and manage fewer vendors in the data domains across the different business functions.

At the people layer, the Chief data officer (CDO) and Chief Medical Informatic officer are the key stockholders working with the HCO leadership to create the data strategy.

Other stakeholders involved in the data domains are covered in the stockholder's section earlier in this chapter that includes patients, government agencies, insurance companies, clinicians, caregivers, IT, and clinical and business units in the HCO.

Process, data, technology, and outcome layers are covered in detail in this chapter and chapter 5.

Solution Design

HCOs should not waste any of their data. While big data is typically not used by the business to generate value, it can bring more information to internal users for better insight or share with external consumers for profit. Thus, big data analytics is a business service that requires planning and following data product management methodology to define the roadmap that involves selecting the technology, providing an end-to-end solution, service supports, and product lifecycle for continuous improvements and adding capabilities.

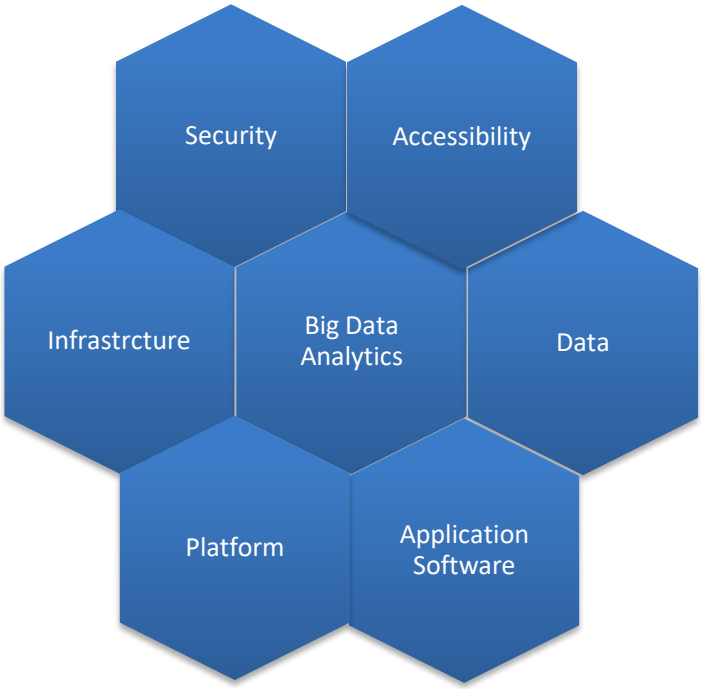


Figure 34: Big Data Analytics Solution Design Framework

The framework of the solution design of big data analytics, as shown in Figure (34), consists of several components that are considered the building blocks of big data analytics. Networking, computing, and storage are the needed infrastructure for big data analytics. Data is stored in the storage and transferred from one host to another via a high-speed network. High-performance computing with a powerful Central Processing Unit (CPU) is essential for big data analytics for real-time results. Database, big data, and analytic tools are the platforms needed to perform big data analytics. Software applications are part of the framework's foundation as they are considered the source of the generated data that we want to analyze. Accessibility in real-time to big data analytics from anywhere and from any device is a critical component of the

solution design. Finally, security and privacy are built around the design to protect the system's data and information. All the previous components introduce enormous challenges to the IT department of the healthcare organization.

Emerging technologies such as cloud computing can overcome many of these barriers. It is highly recommended to use cloud technology to benefit from its agility for the fast delivery of business needs and its elasticity to scale up or down based on business demands dynamically. Therefore, cloud technology is suitable for innovative products or services to allow the business to determine if they are viable with managed costs. Thus, enabling big data analytics is a good candidate for cloud computing. Cloud services have various cloud models such as Infrastructure-as-as-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS), as shown in Figure 34. Each of these cloud services could be built on-premises as a private cloud, hosted off-site in a co-location, or at a partner data center. The real benefit of the cloud technology is obtained when the public cloud is used in a multi-tenant concept that allows several customers to share the same resources in a secure and private model without paying for the entire infrastructure cost. Public cloud is the most economical way to deploy big data analytics, and it enables IT to use infrastructure in a composable approach as well as big data analytics applications in a modular fashion that is easy to customize and directly consumed with self-service, instead of launching a vast farm of servers that can cost a fortune. Additionally, IT can combine private and public clouds into the hybrid cloud to take advantage of both worlds. Thus, the concept of the cloud is the same regardless of it is being private, public, or hybrid.

Figure 35 shows the cloud computing models: SaaS, PaaS, and IaaS.

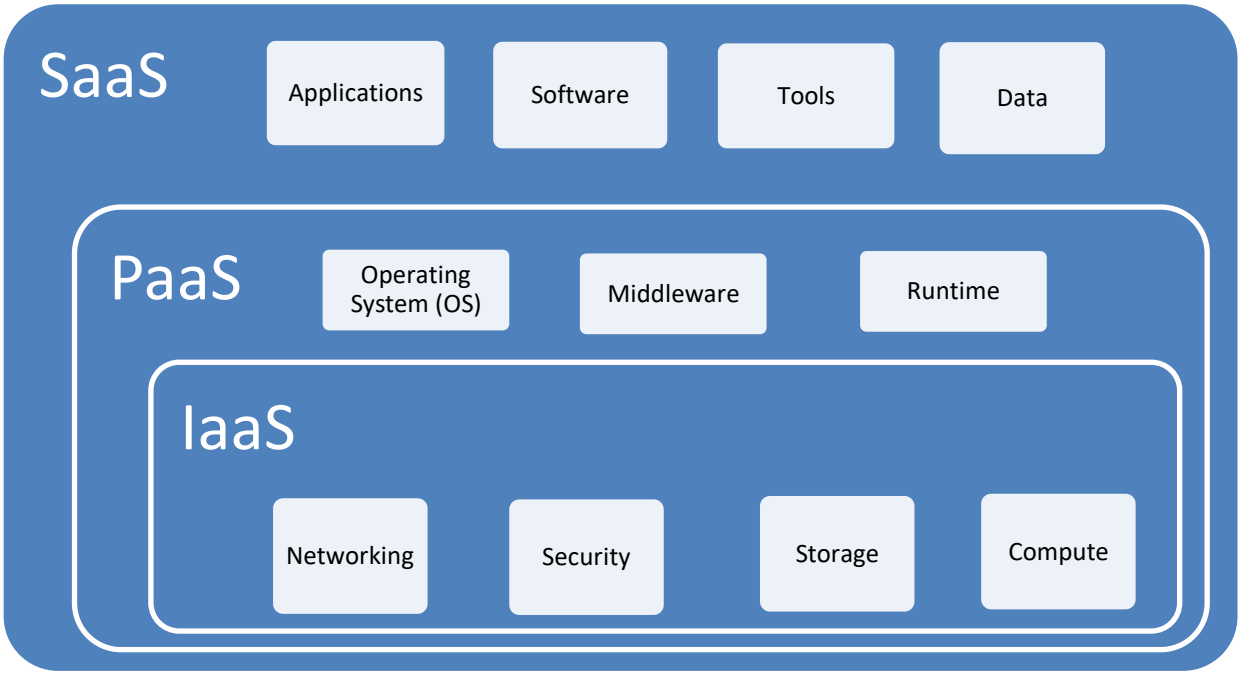


Figure 35: Cloud Computing Models

Consequently, cloud computing plays a vital role in big data analytics solution design. Therefore, this initiative should align with the overall comprehensive cloud strategy rather than ad hoc that the analytics team leads.

Figure 36 shows the big data analytics services that are offered in Microsoft's public cloud.

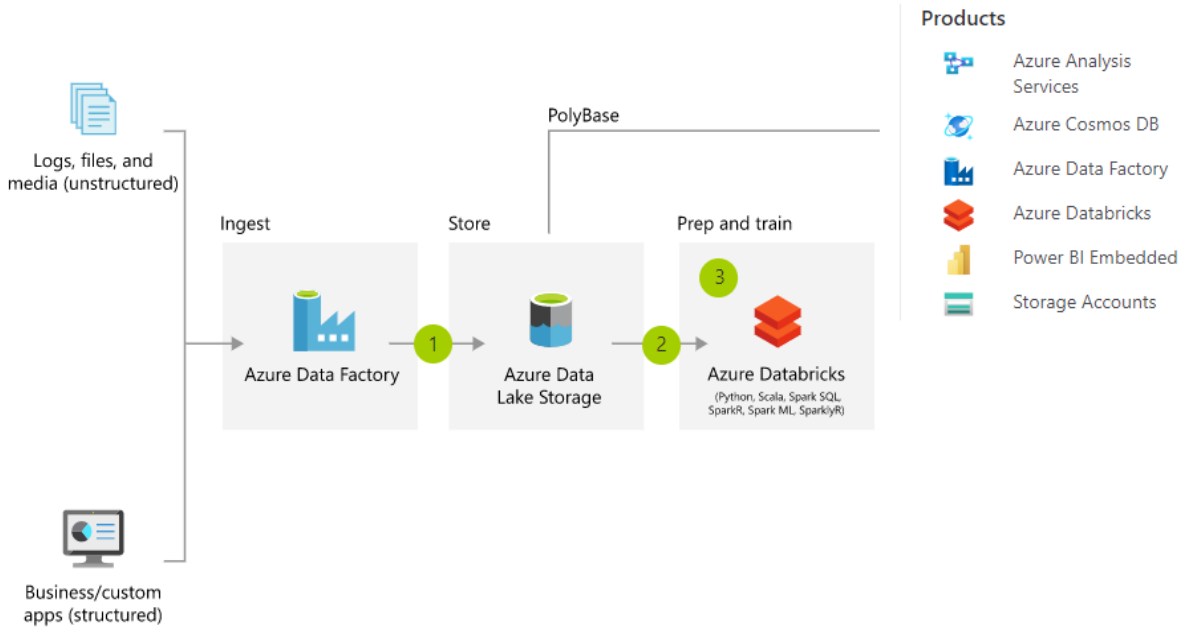


Figure 36: Customized Data Analytics Configuration
Obtained from ("Microsoft Website - Azure Pricing calculator," 2021)

Value Proposition

The fundamental value proposition of adopting the framework is to optimize the benefit of big data technology to allow accessing patients' "complete" data in real-time to enrich efficiencies in care delivery during the patient's treatment journey. Additionally, it supports the end-to-end pathway and workflow for better patient engagement. It also increases patient access with a digital front door experience for the patients. Therefore, it is cutting the cost of healthcare and adding to the revenue through new business opportunities.

Healthcare organizations strive for integrated solutions that focus on driving productivity, efficiency analytics, and new capabilities such as digital front door and visual care to acquire new patients through new fueling innovation technology-centric offerings. These include telehealth, remote patient monitoring, virtual hospital, senior care, and population health powered by data technology. In addition, providers and clinicians will be empowered with self-service capabilities in system tools that streamline the way they treat patients to be more productively in a mobile fashion.

An essential part of the value proposition is the enabler perspective that helps drive quality regulatory to get recognized as a leader organization in healthcare such as Leapfrog or Watson top 100 hospitals, which is incredibly powerful from a marketing perspective for growth and strengthening economics. The key measures of benchmarking against other healthcare organizations are mortality, readmissions, and acquired infections. Therefore, the value is driving quality that cultivates expertise on leveraging new data and advanced analytics to improve the patient journey while being cured. Additionally, it can drive a differentiated patient experience with a flexible scale and architecturally digital platform around data analytics.

The use of data analytic technologies transforms the business model toward digital business to provide new revenue and value, producing opportunities by evolving traditional healthcare organizations to become composable Healthcare. This composable healthcare can compete with retail providers that started to pour billions of dollars into building digital capabilities as they enter healthcare while offering a low cost on the continuum of care services. In addition, many retail providers have expanded their existing pharmacy and optometry care to offer other healthcare services such as virtual care retail clinics, health screenings, primary care, and vaccinations with price transparency.

Therefore, for traditional healthcare organizations to have market growth and market share gains through new products and services, they need to accelerate their digital capabilities and data analytics. This acceleration allows the healthcare organization to reach out to Primary Care Physicians (PCPs) to be part of the community connect for mutual benefit and encourage other healthcare organizations through joint ventures and partnerships acquisitions for a larger market share.

Readmissions are another challenge that many healthcare organizations deal with when they treat patients outside their network. They cannot ensure that those patients do a follow-up with their primary care physician. They end up getting readmitted again after a few days. CMS applies penalties and fines against readmission for not treating outside network patients as their patients. A solution such as wearable devices using IoT for tethering remote patient monitoring to stay in touch is an excellent example of using data analytics throughout the patient treatment cycle.

Patient engagement via texting and emails for advice and counsel before patients visit a specialist helps establish a relationship between the doctor, the nurse, and the patient that allows patient education about their illness. Getting data back is incredibly valuable because it creates a stronger bond with patients as it almost has ongoing communication between caregivers and their patients. Healthcare organizations desire to drive patient experience throughout that treatment journey and improve each touchpoint and interactions for excellence patient experience throughout their lifetime.

Cost Analysis

This section aims to identify the Total Cost of Ownership (TCO) of the big data analytics initiative for the lighthouse project as a proof of concept (PoC). The lighthouse project allows examining concepts of data initiatives on a small scale to determine values. The analysis assesses the gain of achieving the goals that will turn into benefits or business values and determines if it is worth the time and money needed for implementing the solution on a larger scale.

Utilizing the public cloud is an ongoing operating cost. Therefore, HCOs do not need to finance the procurement of hardware infrastructures such as networking, computing, and storage using their capital funds. Additionally, most software companies offer subscription-based licenses for their applications and tools in the cloud that is also considered an operation cost, rather than selling perpetual licenses.

The Total Cost of Ownership (TCO) of the big data analytics product is calculated based on its life cycle and roadmap. A product's life cycle starts with the planning phase and ends when the product retires or is replaced.

Walterbusch, Martens, and Teuteberg (2013) provided their respective TCO of different cloud computing models. The TCO covers the five phases of the product life cycle:

- The cost of the initiation phase is where the time is spent to decide the cloud type(s) and services.
- The cost of the evaluation phase is to select the right cloud service provider based on Service Level Agreement (SLA), technology, security, and price requirements.
- The cost of the transition phase includes the implementation and configuration of the services.

- The cost of the operation phase to run and maintain the service for monitoring and reporting the performance and metered usage charges.

The paper did not consider the cost for the dissolution phase as the service can be stopped at any time without penalty. However, if the data stored in the cloud is critical and needs to be transferred back, there will be a cost associated with data migration.

Figure (37) shows the sequence of phases of the cloud computing life-cycle.

Figure 37: TCO of Cloud Computing Life Cycle
Recreated and Modified from "Evaluating cloud computing services from a total cost of ownership perspective." Walterbusch, M., Martens, B., & Teuteberg, F. J. M. R. R. (2013).

The TCO of the big data analytics is calculated to include the initial cost as capital expenditure (CapEx) and the ongoing cost of the operational expenditure (OpEx).

The CapEx includes the consulting and the professional fees to implement the project as many HCOs do not have the technical expertise necessary for deployment.

The OpEx includes two variable costs and two fixed costs. The monthly charges of the connectivity to the cloud (Con_i) and the monthly charges of the cloud services usage (Cld_i) vary each month based on usage with a range from i=1 (first month) to i=n (last month). The two fixed charges are the monthly charges of the software license subscription (Sft) and the monthly labor cost (Lab).

Therefore, TCO is the sum of CapEx and OpEx

TCO = CapEx + OpEx ……..Equation 1

Where OpEx $= \boldsymbol{Sft + Lab + \sum_{i=1}^{n} Con_i + \sum_{i=1}^{n} Cld_i}$ …….Equation 2

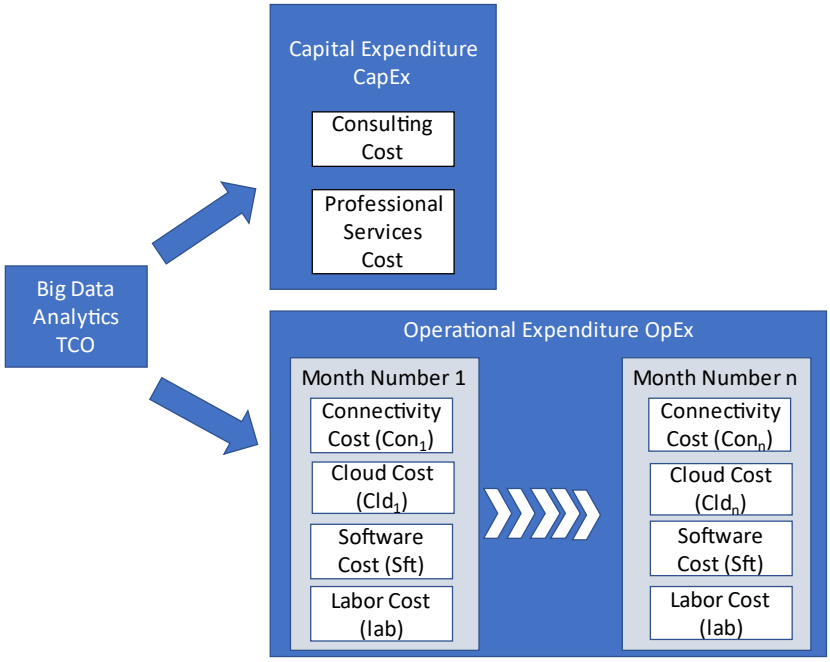


Figure 38: TCO Components of Big Data Analytics

Figure (38) illustrates the TCO components of big data analytics. Even though software vendors and cloud providers offer great help to educate their customers, hiring a consultant to look after the service is highly recommended. The consultant will provide guidance and direction to select the appropriate technology and cloud services provider. The consulting fee for big data analytics is typically a flat rate of around $30,000.00, but it can vary based on the length of engagement. Although, it can go for several hundreds of thousands of dollars if a distinguished consulting firm is hired to construct a wholly digital and cloud strategy. Professional service is needed to implement the solution by Subject Matter Experts (SMEs) in cloud technologies, data

base, big data, analytics, data warehousing, and security with an estimated 120 hours to complete the job at $250.00 per hour for a total of $30,000.00. Therefore, the total estimated CapEx is $60,000.00.

The central portion of the OpEx budget is for the cloud, and it decided based on an estimated calculator that the public cloud providers supplied online ("Microsoft Website - Azure Pricing calculator," 2021).

Table (5) shows an itemized list of the components needed for advanced analytics for the lighthouse project of 1 TB for big data with a monthly cost of $13,285.50. Human labor is also part of OpEx, and about 50 hours a month at $50 per hour is expected for operation support for a total of $2500. Thus, the total monthly OpEx is about $15,785.50.

Table 5: Estimated Cost of Big Data Analytics in Public Cloud

Microsoft Azure Estimate		
Advanced Analytics on Big Data		
Service type	**Description**	**Estimated monthly cost**
Azure Synapse Analytics	Tier: Compute Optimized Gen2, Dedicated SQL Pools: DWU 500 x 730 Hours, 1 TB of storage with Geo-redundant disaster recovery	$5,592.05
Azure Data Factory	Azure Data Factory V2 Type, Data Pipeline Service Type, Azure Integration Runtime: 100 Activity Run(s), 100 Data movement unit(s), 10,000 Pipeline activities, 10,000 Pipeline activities – External, Self-hosted Integration Runtime: 100 Activity Run(s), 1,000 Data movement unit(s), 10,000 Pipeline activities, 10,000 Pipeline activities – External, Data Flow: 1 x 8 Compute Optimized vCores x 730 Hours, 1 x 8 General Purpose vCores x 730 Hours, 1 x 8 Memory Optimized vCores x 730 Hours, Data Factory Operations: 100 x 50,000 Read/Write operation(s), 100 x 50,000 Monitoring operation(s)	$5,045.12
Azure Analysis Services	Developer (Hours), 1 Instance(s), 720 Hours	$95.04
Power BI Embedded	1 node(s) x 720 Hours, Node type: A2, 2 Virtual Core(s), 5GB RAM, 301-600 Peak renders/hour	$1,445.83
Storage Accounts	Block Blob Storage, General Purpose V2, LRS Redundancy, Hot Access Tier, 1,000 GB Capacity - Pay as you go, 100,000 Write operations, 100,000 List and Create Container Operations, 100,000 Read operations, 100,000 Archive High Priority Read, 1 Other operations. 1,000 GB Data Retrieval, 1,000 GB Archive High Priority Retrieval, 1,000 GB Data Write	$19.44
Azure Databricks	All-Purpose Compute Workload, Premium Tier, 1 D3V2 (4 vCPU(s), 14 GB RAM) x 730 Hours, Pay as you go, 0.75 DBU x 730 Hours	$504.80
Azure Cosmos DB	Standard provisioned throughput (manual), Single Region Write (Single-Master) - East US (Write Region); 400 RU/s x 730 Hours; 1,000 GB transactional storage, 2 copies of periodic backup storage; Dedicated Gateway not enabled	$273.36
Virtual Machines	1 D2 v3 (2 vCPUs, 8 GB RAM) x 730 Hours; Windows – (OS Only); Pay as you go; 0 managed disks – S4, 100 transaction units; Inter Region transfer type, 5 GB outbound data transfer from West US to East Asia	$152.62
Virtual Network	1000 GB data transfer from East US region to East US region	$20.00
Cloud Services	1 D2 v3 (2 Core(s), 8 GB RAM) x 730 Hours	$137.24
	Total	**$13,285.50**

Expected Outcomes

The expected outcomes of the framework allow appropriate technologies to provide high accuracy of complete medical records (volume and depth) to enhance clinical documentation necessary to attain system-wide quality in real-time and revenue targets.

The revenue depends on accurate documentation that ensures appropriate reimbursement and collections against denials. As the healthcare reimbursement model changes, causing fines for HCOs with lower quality of care, embracing the framework shall prevent inadequate patient treatment by alerting providers with the severity of illness—complete health records to prevent making wrong conclusions that reduce the risk of mortality of the patients. Moreover, CMS decides the reimbursement rates to hospitals for Medicare and Medicaid patients by applying the Case Mix Index (CMI) that meticulously evokes the mixture, difficulty, and seriousness of diseases of patients treated at an HCO facility.

Additionally, it will eliminate the manual process of reviewing the clinical documentation. For example, many HCOs have a challenge of reviewing 100% of charts that may cause a lost opportunity for prospective payment to maximize reimbursement by classifying patients with an inappropriate Diagnosis-Related Group (DRG). This allows automation to review charts to support case management throughout the patient's stay until the time of discharge with the Length of Stay (LOS) information.

Figure (39) shows the expected outcomes of big data as part of data completeness for analytics. In addition to reimbursement and quality, it adds automation and real-time reporting capabilities as expected outcomes.

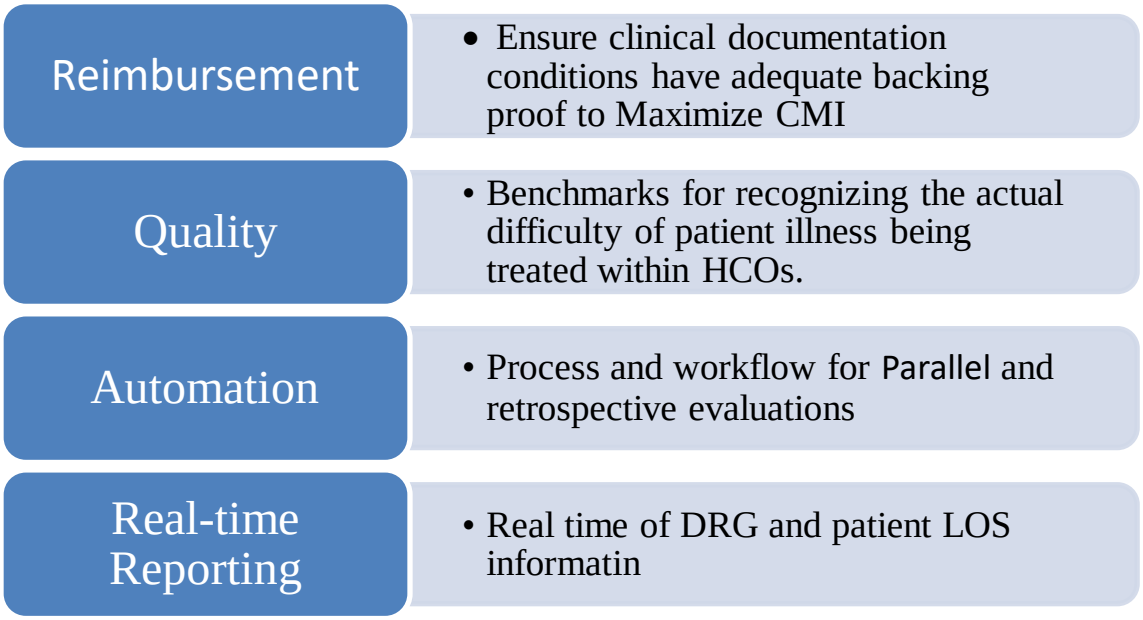


Figure 39: List of Expected Outcomes of Adapting the Framework

Implementation of Framework

The benefit of adopting a framework for healthcare organizations is to avoid rolling out a tool or a system that did not meet the needs to accomplish the expected outcomes. If a well-defined methodology is fundamentally used correctly, it will lead to successful results. Unfortunately, many healthcare organizations do not recognize the process aspect or their technical team's skill and competency aspects and focus only on the tools. As a result, HCOs can spend money on software applications and designate time and resources. Still, they have to be focused on putting all parts of the puzzle collectively to work coherently.

The objective of the case study is to support the benefits of implementing the framework to employ big data to add capabilities such as automation, deep image learning, ML, and AI to HCOs. In addition, there is a need to enhance and refine the use of the available data and analytic tools.

The framework was a multi-year process to implement in one of the HCOs in the state of Florida with the following profile:

- $4.3B annual revenue
- 3,200-bed system
- 15 hospitals and emergency departments
- 4,200 physicians and 80 medical specialties and subspecialties
- 22,000 team members
- 150,000 inpatients and nearly 3.1 million outpatients.

The implementation of the framework started in 2018 by identifying the necessary elements to revamp the entire data strategy that addresses the challenges of the HCO. It required a complete IT transformation and organization restructuring to partner with other departments rather than being a technology service provider. To ensure alignment with this HCO six strategic imperatives were formed:

- Embrace Quality & Safety
- Become Best Place to Work
- Earn Physician Loyalty
- Drive Growth & Innovation
- Enhance Ease of Use
- Strengthen Economics

Like other HCOs in the market, this HCO realized the change of reimbursements from pay-for-service to pay-for-value, aiming to deliver high-quality healthcare at a lower cost with an excellent experience for both patients and providers. The solution to those challenges is to leverage the fully available data, including semi-structured and unstructured data, for better analytics in the clinical and business areas to make data-driven decisions in real-time utilizing emerging technologies.

As background, the initial evaluation revealed the inconsistency of the applications with exacerbated silos because of the lack of standardization. Despite the number of analytic tools and their high TCO, it is evident that the organization's data capabilities had been suffering because of the gap in the technical skills to support the available tools for data analytics. Both data and analytics functions were controlled by IT, which led to other departments and business units establishing what is called “shadow IT.” As a result of lack of coordination, too many data analytic tools and systems of the same function were found that added to the operating expenses with no tangible benefits.

Additionally, this model caused delays in making decisions in clinical and business matters due to spending more time piecing several outputs together.

Furthermore, the analytic data system design and the follow was very complicated and least desirable, which led to low performance and unstable system. Therefore, this HCO needed a proper framework to coordinate efforts and reduce redundancies.

Regarding data sources, the finding showed 1,275+ applications were in use by the HCO that generated 3,339 TB of data, in which 436 applications were identified as the core of the ecosystem that the HCO used as a source of data for reporting and data analytics for clinical and non-clinical areas as shown in table (6).

Table 6: List of System Applications

	Clinical System	Number of Applications
Clinical	Allied Health	92
	In/Outpatient	30
	Inpatient	78
	Outpatient	49
	Population Health	4
	Clinical Total	**253**
Non-Clinical	HIM	28
	Patient Access	24
	Patient Experience	3
	ERP	27
	Others	101
	Non-Clinical Total	**183**
	Total	**436**

Additionally, this HCO is participating in One-Florida and HIE for exchanging data and health information with CMS and other HCOs in the region.

For Analytics Tools, 84 embedded and stand-alone analytics applications and tools were identified with more than 11,000 data elements that provide reporting and BI used across 38 business units. The list of the analytics application includes QlikView, Tableau, Power BI, IBM SPSS, Crystal Reports, SQL Server Reporting Service (SSRS), and HealthCatalyst EDW. In addition, about 20+ other analytics applications were identified as analytics applications but were not included in the analytics application's portfolio. These applications hosted on-premises with outdated technology presents security and support risks to the enterprise without delivering the best value to the business. Furthermore, the cost of these applications is very high, and the

operational value is very low due to the fragmented and poor performance of the overall analytics application's portfolio.

Moreover, users needed a lot of training to use the applications that kept them away from their core healthcare business.

Table 7: List of Analytics Tools:

Business Unit Area	Number of Applications
Administration	2
Audit Compliance	4
Care Assessment & Discharge	1
Care Delivery	40
Care Initiation / Registration	1
Care Management	6
Billing and Reporting	1
Encounter Initiation/Scheduling	1
Finance	20
Human Resources	1
Legal	1
Medical Education and Research	3
Performance Management	1
Regulatory	2
Revenue Cycle	8
Supply Chain	2

Table (7) indicates the analytic tools for reporting and BI using the data sourced from the HCO's ecosystem applications.

In 2018, Health Catalyst Enterprise Data Warehouse (EDW) had approximately 8.8 TB of data from 14 source systems for analytics with a 50% compression ratio, including the database and indices. Total environment footprint includes OS and data on all servers, including

the EDW and audit servers which consume most of the data. Health Catalyst's EDW is not a system of record or archive platform for source systems, databases, or applications. The HCO keeps its existing data sources separate from the EDW. If needed, additional data elements can be extracted to the EDW. QlikView hybrid licensees' model is used for visualizations with inadequate data flow and limited accessibility for users at the HCO side.

Figure (40) shows the layout and the flow of the data before implementing the framework.

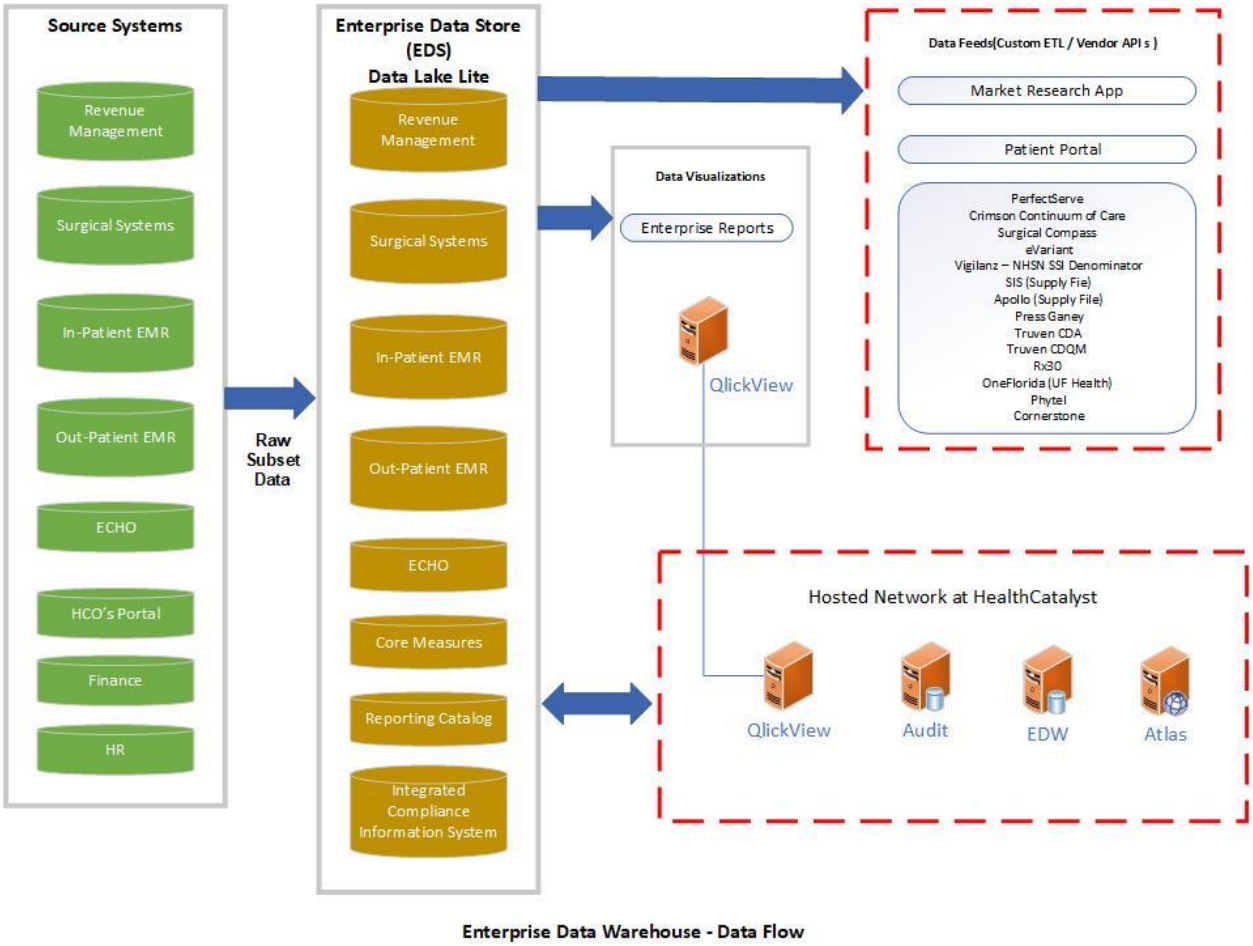


Figure 40: Hosted Data Warehouse Components Diagram

The primary challenge of the HCO is how to stop spending money on data and analytic tools that are silos, redundant, and outdated.

Implementing the framework has allowed the HCO to overcome this issue. It has offered the principles of revamping the entire data strategy with a new fragmented landscape for data investment. It started with restructuring the IT department in a function and/or a role structure to support technology at the HCO. For example, the applications team acted as a function maintaining the data warehouse and responsively dealt with data integration tasks to the clinical and business areas that request reports and KPIs information.

The role of the applications team members depended on their skillsets to execute a set of assignments and their availability to deliver needed reports, as shown in figure (41). Thus, the IT structure for data-related areas limited the HCO from proactively achieving the data strategy.

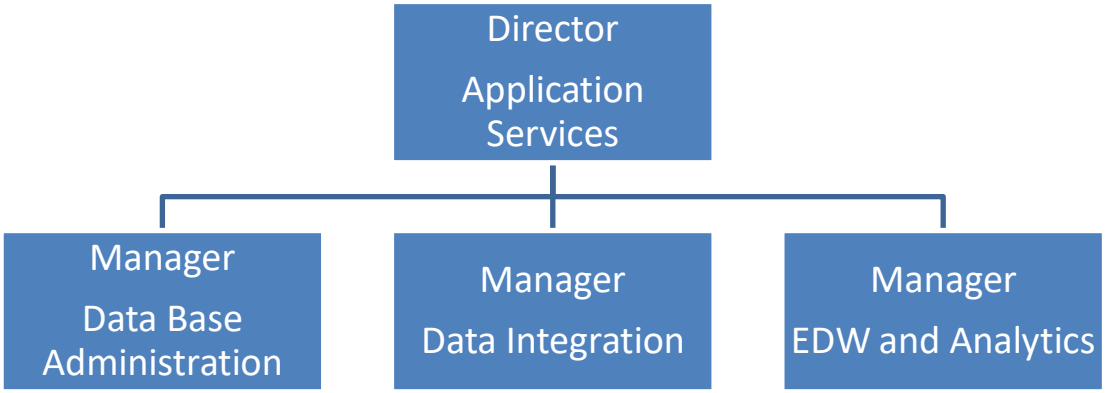


Figure 41: Old Organization Structure for Data Related Applications

The framework allowed addressing the challenges with disparate data and moving away from the organization's central focus because of the high cost and complexity.

At the People layer, the change of the organization structure presented an explicit focus on data as a product that drives toward an enterprise level, rather than having business units apply the Do-It-Yourself (DIY) approach to leveraging data analytics and creating reports.

It started with appointing a CDO who has healthcare acumen, big picture mentality, and strategic agility to allow collaboration between business units and IT. The role of the CDO is to oversee the venture capital in the data and analytics technologies to align opportunities and investments. Additionally, it led the data team responsible for business analysis, enterprise analytics, and data science with defined job descriptions. In addition to the CDO, four leadership positions were needed to support the direction of the HCO's data strategy as functional areas. The new CDO organization structure is shown in figure (42).

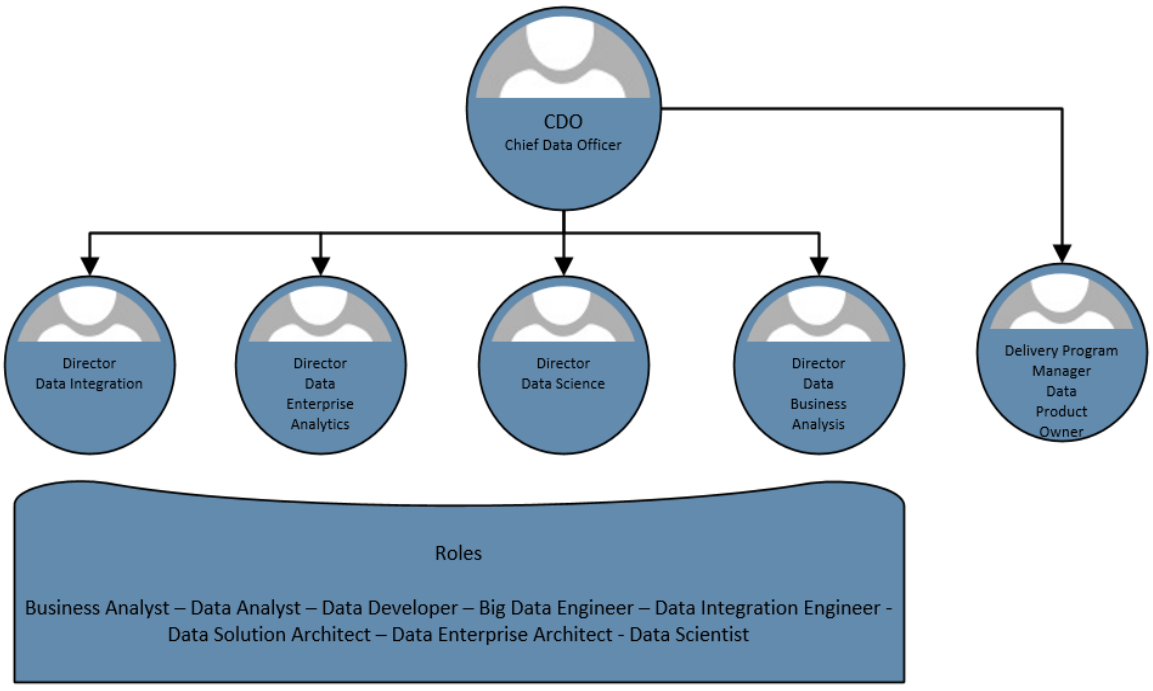


Figure 42: New Data and Analytics Organization Structure

A data product owner/delivery manager is a new role that shepherds possible opportunities using data from the starting idea to execution. This position was created to develop the guidelines for managing the data products for enhancement and operational support to reach a level of product maturity with standardization and product life cycle for new releases. In addition, establishing a product roadmap and setting key performance indicators (KPIs) are key tasks for the data product owner that involves showing the success of the data product and value realization.

At the organization level, the CDO worked with the HCO's executive leadership to align data strategy with the quality and performance of the organization's overall strategy by innovation. This required collaboration with other business units for a new fragmented data and analytics landscape that serves patients and drives growing market shares.

At the process layer, an effort stated to revamp the processes around data owner, data steward, and data custodians for data governance attainment that line up with business processes. The goal is to enable the organization with data transparency where accuracy and trust data are easy to access and data literacy where users can read, communicate, and understand the data, resulting in value. Data governance is essential to ascertain that a translucent choice and responsibilities for data decisions. It classifies what the needed decisions, which makes them are, how they are made.

The data governance introduced six pillars, as shown in figure (43):

- Data privacy controls access and encrypts the data to ensure that the data is secured and only accessible to appropriate users.
- Data Management is to keep the data arranged and structured for usability according to the policy and procedure.
- Data lifecycle is classifying the data active, archive, delete, and purge.
- Data Quality is to be confident in the data available, and it suits the purpose of usage.
- Data Ownership is to identify the department that is accountable for the data assets that the department generates.
- Data Valuation is to deliver the total value of data analytics that is aligned with the strategy.

Figure 43: The six Pillars of Data Governance

At the technology layer, focusing on a subgroup of legacy analytics tools that have outdated technology and need to be checked to remediate risk and lower the costs, the framework optimized the analytics applications portfolio by standardization. This focus eliminated 9% of duplicate analytics tools with about $1.4M of cost-saving in-licenses and support.

Furthermore, modernized applications were added to the data and analytics applications portfolio to migrate toward an enterprise platform that takes advantage of the contemporary software offered in the cloud supporting predictive analytics and utilizing the available structured and unstructured data.

Figure (44) shows the layout of the migrated data and analytics platform shifting towards cloud computing.

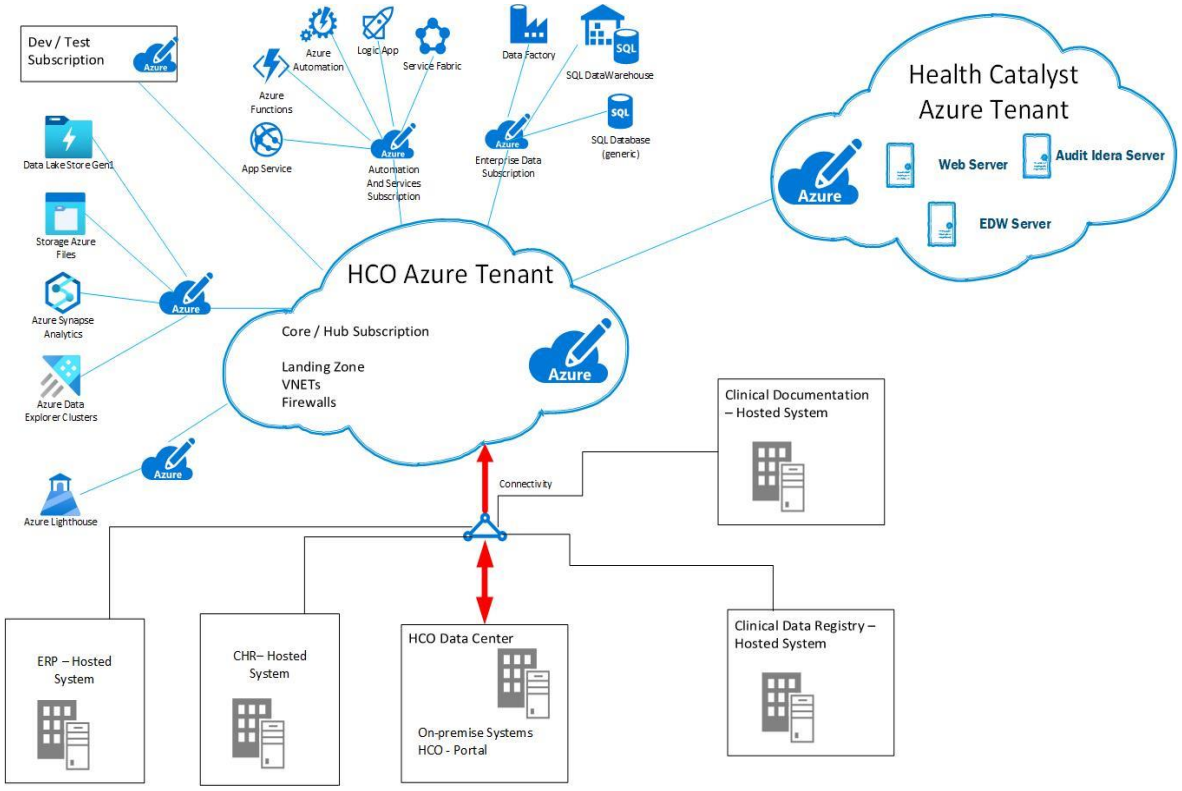


Figure 44: Modernized Data and Analytics Platform Transformation – A Composable Architecture

The solution is designed to be modeled as building blocks in a composable fashion that can be assembled or reassembled and expanded based on selected functions and tool abilities to meet business goals.

The applications that supply data have location independence and are distributed in several places: on-premises, vendor-hosted, and public cloud. Additionally, Health Catalyst has remained a vendor of the EDW hosted in their tenant in Azure cloud that offers innovative data warehousing platforms.

The analytics functions are performed in the cloud as a SaaS, and it only requires checking a box or clicking in the desired functionality.

This solution effectively placed data completeness into the hands of the business with little interaction from users when executing data analytic jobs.

Summary and Conclusion

This chapter described the nature of healthcare and mapped the business units with different departments that form the healthcare system. It showed how data are generated in healthcare and used SIPOC-R to show the supplier and customers of the data in each phase of the treatment from the time patient entered the system to the time of ending the service and providing feedback.

It included the big data technology and the criteria of picking the applicable data and analytics tools in the appropriate platform.

The proposed value was discussed to show the benefits of utilizing big data analytics to the HCOs.

It covered the proposed conceptual framework with 41 components in 6 different layers, Organization, People, Processes, Data, Technologies, and Outcomes.

The organization layer presented the strategy of healthcare organizations from the business aspects has changed because of the new emerging technology that became available in the recent years.

The people layer identified the stakeholders in the different departments of a healthcare system that collaborate to serve patients with the best care possible.

The process layer involves understanding the process map during the patient's journey in the healthcare system while getting treatment for illness.

The data layer illustrates the source of data in healthcare, how it is stored in the system, and how it is shared with other healthcare organizations.

The technology layer listed the involved technologies in implementing big data analytics such as cloud, IoT, and security.

The outcomes layer presented the expected results and values of adopting the framework.

CHAPTER FIVE: RESULTS, VERIFICATIONS/VALIDATIONS AND ANALYSIS

Introduction

A Framework is a representation of directly linked and cohesive processes essential to achieve a sophisticated task. It tells "WHAT" must be done to ensure building a solution that orchestrates the efforts between the components necessary to flourish and adds value to the business. However, it is crucial to examine and validate the framework to demonstrate that it is worth the time and effort to adopt.

System Architecture of the Framework

System architecture describes the physical plan and interrelationship of the elements of the design. The characteristic of the architecture is to standardize and deliver consistency. It is used to lay out the design of the system and the correlation of its components in a comprehensive diagram (Steiner, 1998).

Architecting a system means forming a complicated and multi-disciplinary system in an integrated architecture for full development. It is a systematic approach that includes multiple domains to make certain decisions during implementation. The assortment of the components that share the same attributes of a system is called a view. There are six views of system representations, Purpose/Objectives, Behavior, Managerial, Form, Performance, and Data. Grouping components and connectors make a style that constructs the system. The style also identifies the limitation of the system (Maier, 1996).

Figure (45) shows the six views of system representations.

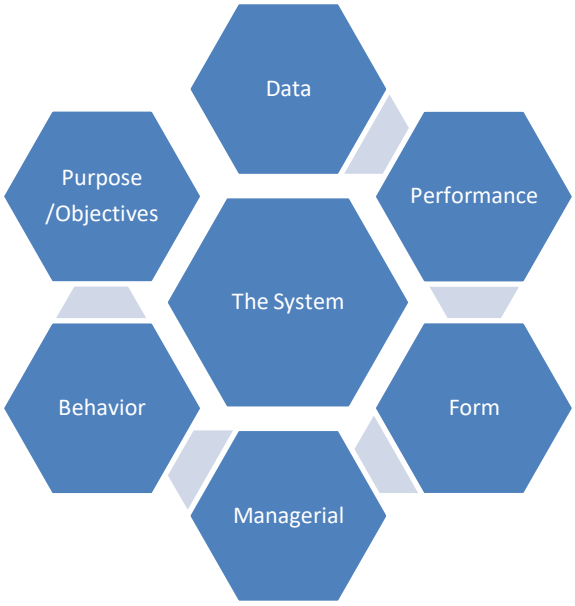


Figure 45: Six-View Representing Systems Architecture
Recreated and Modified from Maier, M. W. (1996). *Developments in system architecting.* Paper presented at the Proceedings of ICECCS'96: 2nd IEEE International Conference on Engineering of Complex Computer Systems (held jointly with 6th CSESAW and 4th IEEE RTAW)

The system architecture of big data and analytics integrates five architectural areas that serve as a blueprint at the enterprise level to allow HCOs to attain their goals of big data strategy. The system architecture abstractly follows the proposed framework. It starts with the business architecture that determines the vision, the strategy, guiding principles, the service capabilities, the operating model, the metrics and performance measurements, and the culture of being an innovation HCO in data and analytics areas. Additionally, the business architecture sets the policy and procedures that need to be followed to meet the rules and regulations of the healthcare government agencies related to data. Furthermore, business architecture identifies the stakeholders and puts the organization structure with the appropriate skillsets to support data initiatives.

The process architecture establishes the activity and workflow procedures that govern the data lifecycle and management processes. It also poses the rules to ensure that the data is secured and privacy is achieved.

The technology architecture depicts the hardware platforms for the core infrastructure and edge devices. The core infrastructure specifies the standards for computing, storage, and network protocols and topology. Cloud network is another core technology that is currently used for data and analytics. Edge and endpoint devices are the end-user, IoT, and biomedical devices that connect to the network. Security covers both core network and edge devices for protection from cyber security attacks.

Application architecture represents three components: the application software, the database, and the application access as user interface/experience (UI/UX).

The significant data architecture is the key element in this system architecture, and it needs to associate with the other four architectures. The data source comes from the applications and edge devices in structure, semi-structure, and unstructured formats and with 3Vs of big data attributes. The data management components are the main instruments for the success of the entire effort, and the malfunction in this area can lead to undesired outcomes. Most of the big data and analytics implementations fail because of inappropriate data management performance. The main activities of data management are:

- Data integration is used when the data source is entirely coming from internal systems, and data ingestion is pulling data from both and internal-external.
- Data cleansing or data cleaning is the process of identifying unreliable and obsolete information from a dataset. This process excludes incomplete and inaccurate data from being processed.

- Data validation is a step after date cleansing that verifies the accuracy and the integrity of the data before using it for processing.
- Data curation is needed to organize and prepare the data for usage. It describes the data using metadata schema which, provides information such as time and date of data creation, the purpose of data, and the creator of the data.
- A data catalog is an assortment of metadata that helps search and analyze the data. It also provides a portfolio of accessible data.
- Data lineages are the activity of knowing each step of the data flow from the source to usability for documentation and visibility.

The data hub is the centralized point for all data at the enterprise level. It connects and manages the data in both data warehouses for structured data and data lake for unstructured data. Data virtualization is a logical place that consolidates all disparity data in different systems in a centralized location that delivers data in real-time.

Data science prepares the data for advanced analytics by allocating big data to be more usable by ML to build an automated model for data analytics and enable the system to learn from data. NLP allows recognizing text, and data mining is used to detect patterns in datasets, and it has 4-type of techniques: for descriptive analytics, clustering and association rule discovery are used. For predictive analytics, Classification and regression are used.

Data processing includes transaction processing that processes one batch per transaction, and it needs user interactions. This batch processing process is a group of transactions all at once without user involvement. Stream processing is a real-time big data processing that performs immediate progression actions when the data is created and received.

Presenting the data processing results can be visually in a graph to show the correlation of the datasets. Interactive data visualization permits users to work with different scenarios to discover more insight to interpret the datasets better. However, the traditional data analytics tools still use scorecards and BI reports.

Big data analytics' business outcome leads to real-time clinical decisions, the ability to get alerted before future issues happen and take the appropriate actions based on data insights.

Figure 46 shows the big data analytics system architecture that consists of 5 sub-architectures with their components.

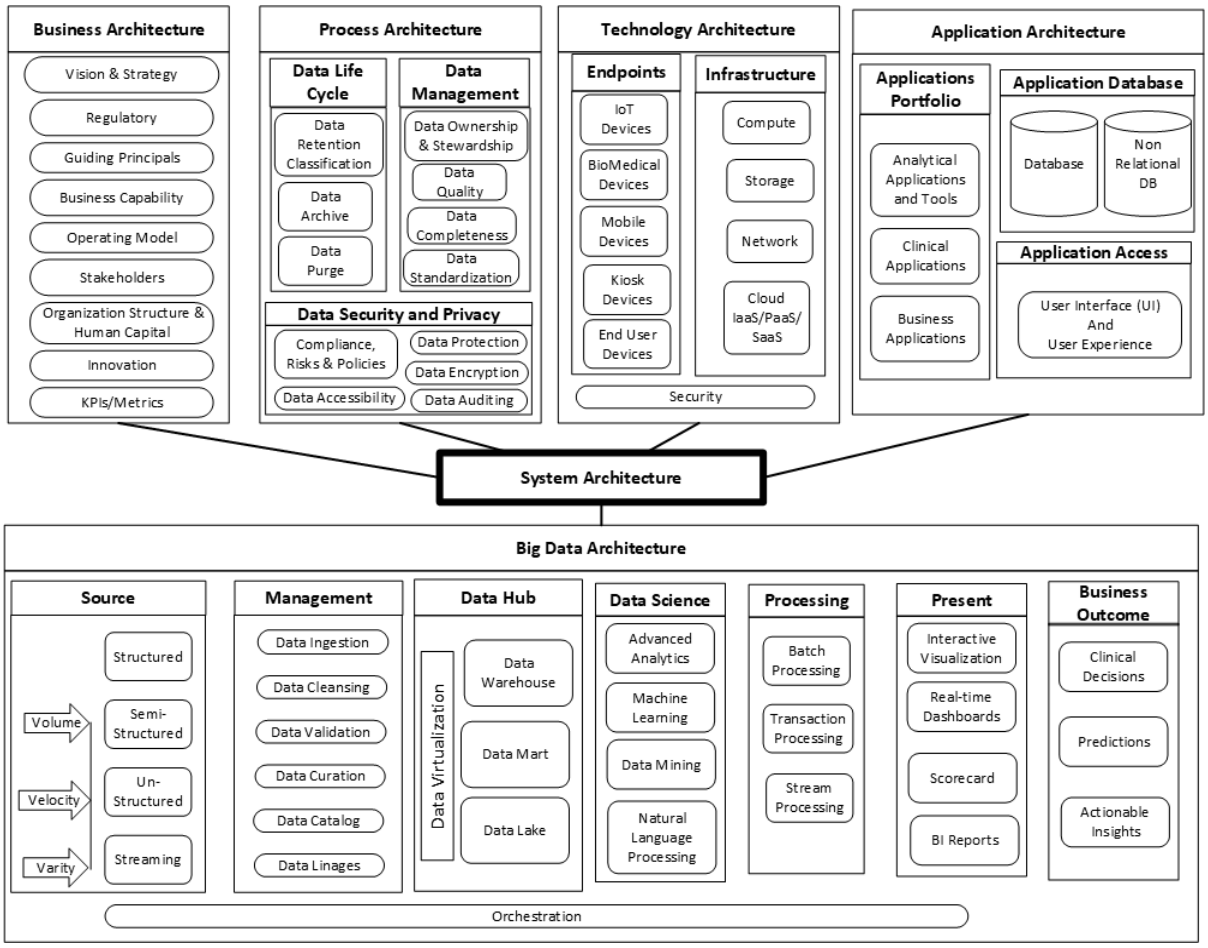


Figure 46: Big Data System Architecture

Deployment

Standing up a lighthouse project as Proof of Concept (PoC) is recommended for implementing new technologies in digital transformation. Performance metrics to measure the success of the lighthouse project should be identified so that it is quickly determined if the project is moving in the right direction or being able to decide to spend more time and money developing a more long-term and a larger scale as a new strategy. This may not commit to a target improvement. Still, suppose the lighthouse project goes well. In that case, this capability can move forward with a minimum success criterion. It could be an exploratory initiative to learn from and shut it down in an early stage. This is about helping business development, operative optimization for a superior organization, and better decision making were spending the money properly for expanding to the next level beyond implementing certain technology.

The deployment of the lighthouse project starts with a squad team. Gartner (2019) research website, it is recommended to create what is called a fusion team. The fusion team is formed with individuals of various disciplines to deliver value rather than service or technology for data projects. It is a collaboration model among key contributors led by a delivery manager who coordinates efforts to ensure a successful implementation with monitored capital expenditure and enterprise architecture that needs to be focused on solving big complex problems practically and working slightly ahead of the organization. Creating a highly productive fusion team should consist of administration and business units, business analysts, clinicians, clinical informatics, IT team, and vendors. These components are dedicated to the project and explain the adjustment of gains and losses during the deployment based on the declaration of the strategic intentions with guiding principles around data analytics.

This approach allows coordination to avoid duplicating other data analytics projects in silos and creating tremendous mess, introducing other systems to improve their areas without

looking at the enterprise level. Therefore, it helps to establish some foundational services and structures to safely work in parallel in different business areas, introducing new digital capabilities for enterprise digital foundation.

Exercise the System

In healthcare, value is represented with several qualitative and quantitative measures for financial and non-financial metrics, including tangible and intangible. The tangible assets such as buildings, equipment, and inventories are 15% of the organization's value in today's economy. The intangible assets are not translated to monetary value, but they are essential to show the framework's benefits. Some intangible measurements are accreditation, reputation, awards/recognitions, patient engagement/loyalty/satisfaction, and physicians' satisfaction. Value in healthcare is classified as access, perceptions, capability, impact, and financial gain, and it should be balanced, efficient, and reliable for tactical and strategic results. Every step of the patient's care journey must be measured, documented, and reported in the healthcare industry, including costs, outputs, quality, time, patient satisfaction, innovation, and HCO reputation. Volume and market share growth are key values needed for measuring the success of the HCO (Phillips, Buzachero, Phillips, & Phillips, 2013).

With the data analytics complete, the HCO has a better-quality outcome and can create and submit quality reports to the organizations in table (8) more efficiently and promptly.

Table 8: Organization for Healthcare Ranking and Awards

Quality Program	Report Frequency	Benchmarking	Risk	Financial Metrics
CMS Value Based Purchasing (VBP)	Annually	X	X	X
CMS Reporting Inpatient/Outpatient	Quarterly	X	X	X
CMS Readmission Reduction	Annually	X	X	
CMS Hospital-Acquired Condition (HAC) Reduction Program	Annually	X	X	
AHCA	Biannually	X (only Florida)	X	
Watson Health 100	Annually	X	X	X
Leapfrog	Quarterly	X		
US News & World Report	Annually			
Becker's	Annually			
Healthgrades	Annually	X		
Comparion	Annually	X	X	X

The improvement in quality care permitted the HCO to win the Watson Health 100 Top Hospitals award in 2021. To earn this award, the HCO should have a high-quality survival, a low number of infections, low 30-day readmission, and low 30-day mortality rates. Additional criteria such as a short wait time in the emergency room, good financial performance, and great patient satisfaction are also considered in the evaluation. Additionally, an Everest award was granted for the top performance and high rate of improvement over 5 years.

Moreover, the projected leapfrog safety grade for the latest evaluation is an A, and the predicted CMS star rating is 4 stars compared to 3 stars in previous years.

Between 2017 and 2018, this HCO had a 34 % market share compared to its competitor in the area with 45% based on the number provided by the Agency for Health Care

Administration (AHCA). However, whereas in market share for Tables (9) and figure (47) show the trends in 2019 compared to the previous 3 years.

Table 9: Trends from 2016 to 2019

	2019	2018	2017	2016
Average Daily Census	1460	1352	1264	1262
Admission	118157	109806	101746	100570
Patient Days	532793	493137	461323	461922
ED Patients	368328	336091	319072	319864
Outpatient Visit	730356	732405	697808	683510

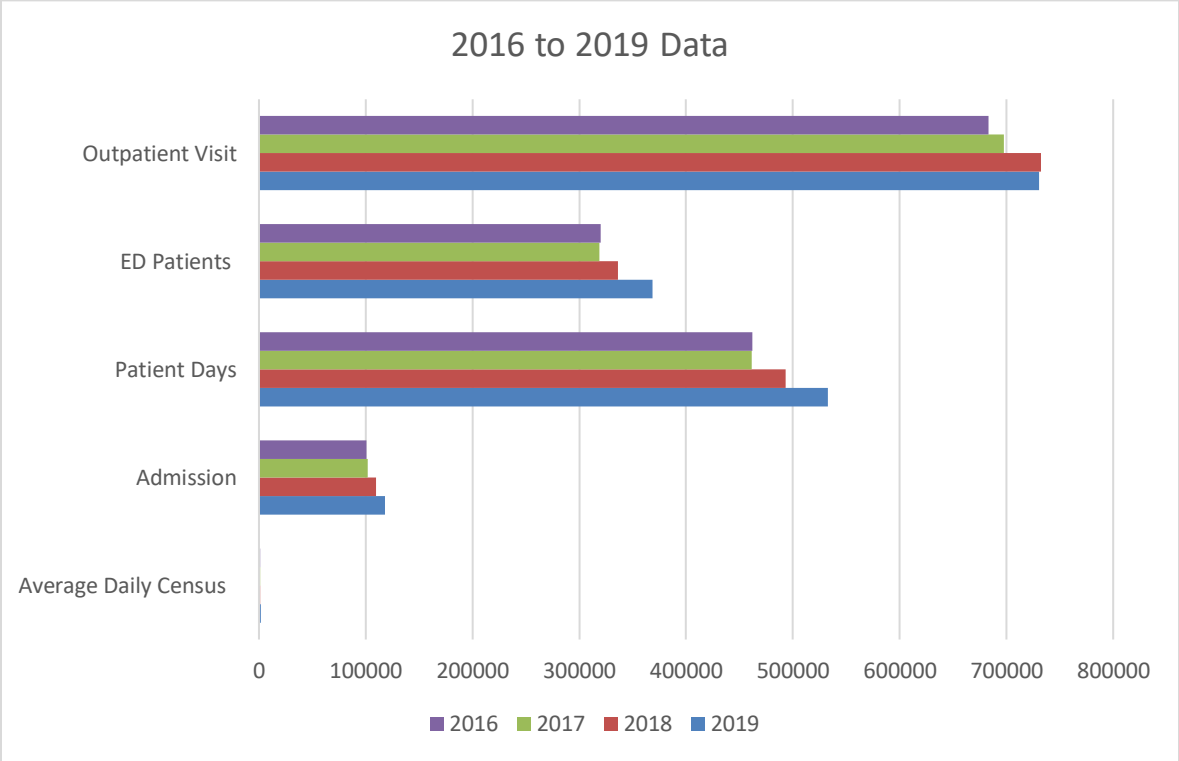


Figure 47: Chart comparisons from 2016 to 2019

Analytics tools assessment is part of the exercise to evaluate the capability of the tools and their cost. The assessment revealed that the HCO spends around $33M on analytics tools, and about $8.3M or about 25% of the total cost can be reduced by eliminating some of the analytics tools. The HCO spends about $9M in analytics tools with the right technical

requirements and fits innovation and advanced analytics. These tools should be retained. Less than half of the spending of $15M needs to be re-engineered and migrated for better use. Figure (48) show the chart of analytics tools that need to be kept, eliminated, or reengineered.

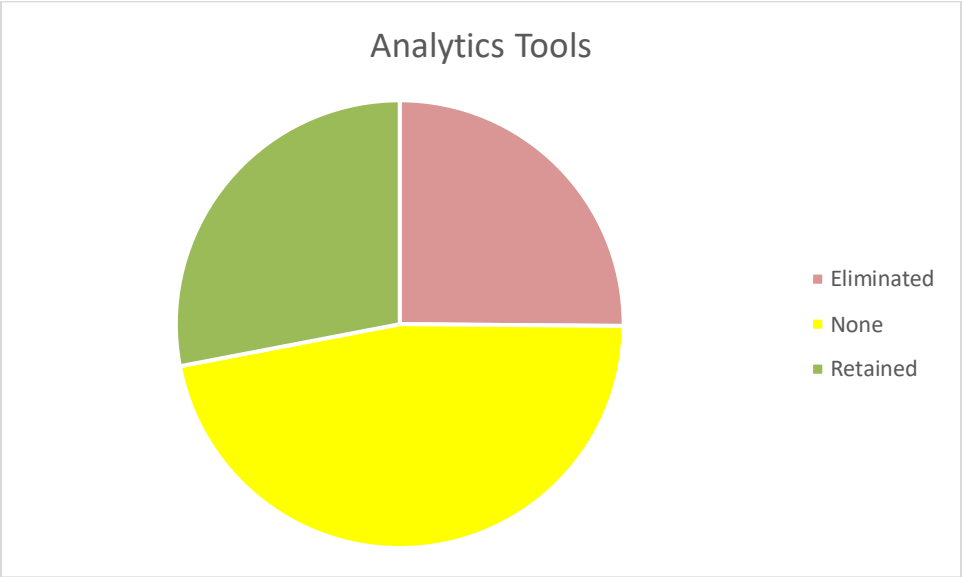


Figure 48: Analytics Tools Chart

Table (10) show the cost of the analytics tools and the areas of use.

Table 10: Cost of Analytics Tools

Row Labels	Sum of CALC_TIME_COST_TOTAL
Eliminated	**$8,251,980.92**
Business	**$6,001,980.92**
Reporting / Visualization	$6,001,980.92
Clinical	**$1,325,000.00**
Reporting / Visualization	$1,325,000.00
ERP	**$650,000.00**
Reporting / Visualization	$650,000.00
Other	**$75,000.00**
Reporting / Visualization	$75,000.00
(blank)	**$200,000.00**
Reporting / Visualization	$200,000.00
None	**$15,391,503.00**
Business	**$1,878,750.00**
Reporting / Visualization	$1,878,750.00
Clinical	**$1,304,375.00**
Reporting / Visualization	$1,304,375.00
ERP	**$1,373,000.00**
Reporting / Visualization	$1,373,000.00
Other	**$475,000.00**
Reporting / Visualization	$475,000.00
Productivity	**$1,683,500.00**
Reporting / Visualization	$1,683,500.00
Utility	**$725,000.00**
Data Science	$575,000.00
Reporting / Visualization	$150,000.00
(blank)	**$7,951,878.00**
Data Science	$0.00
Don't Know	$350,000.00
Reporting / Visualization	$7,526,878.00
(blank)	$75,000.00
Retained	**$9,199,462.52**
Business	**$3,100,000.00**
Data Science	$50,000.00
Reporting / Visualization	$3,050,000.00
Clinical	**$1,650,000.00**
Reporting / Visualization	$1,650,000.00
ERP	**$2,342,462.52**
Data Science	$1,967,462.52
Reporting / Visualization	$375,000.00
Productivity	**$25,000.00**
Reporting / Visualization	$25,000.00
Utility	**$7,000.00**
Reporting / Visualization	$7,000.00
(blank)	**$2,075,000.00**
Reporting / Visualization	$2,075,000.00
Grand Total	**$32,842,946.44**

In addition to getting recognized, receiving awards from different prestigious institutes and having cost reductions by retiring outdated and redundant analytic tools, the HCO's financial report accomplished an increase of net income of $112 million in 2018, being 52% more compared to 2017. This continued throughout the consecutive years, and in the first quarter of the fiscal year 2021, it was reported a revenue increase of $1.08 billion compared to $946.79 million in 2019. In addition, there was a 48% increase in net income of $313 million compared to 2019. As example, the usage of data analytics and digitalization in the emergency room visits has helped drive the increase of revenue by:

- 49% reduction in patients who Left Without Being Seen (LWBS) when visiting the emergency room helped gain revenue by eliminating missed income opportunities.
- 92% reduction in patients leaving against medical advice (AMA) that health insurance does not pay for.
- Identified 20% of medical cases in emergency rooms for faster discharge that increased patient satisfaction and helped free beds for more needed care to other patients.
- An average of 17.5 minutes of time reduction of discharging patients.

Data above was presented in a meeting that showed the new capability and innovation ("Presentation Strategy and Innovation, "2021).

Validations

The validations of the proposed framework in Chapter Four consist of two methods, a case study and SME interviews. The case study is employed to examine the framework in a large healthcare organization in Florida, USA.

SMEs also reviewed the framework in big data and analytics in healthcare and technology for further validation. Additionally, the qualitative approach forms a profound understanding of the framework while validating it rather than a quantitative method that deals with numbers. The qualitative validation deepens the framework's outcomes by including a multi-dimensional view and answering questions such as how, who, why, where, and when that provide accurate assessments to demonstrate good results (Leung & care, 2015).

Case Study

A case study is a research method that can be used to validate the conceptual framework. The questions of "how" and "why" are possibly being asked to direct the use of case studies to fulfill the principles of the qualitative method to define, comprehend, and explain a real-life execution of a framework (Yin, 2009).

The proposed framework was executed as a real-life case study covered in the Implementation of Framework section in CHAPTER FOUR. The validation of the case study is examined, and the result of adopting the framework is successful with the expected outcomes.

Interviews

The approach used in the interviews is a qualitative method that interprets the answer to open-ended questions in an organized way to form a complete picture of the valuation of the proposed framework. The truth of the answers in qualitative study relies on the trustworthiness of the interviewees with four criteria: internal and external validations, reliability, and objectivity (Lincoln & Guba, 1985).

Additionally, Whittemore, Chase, and Mandle (2001) listed primary and secondary criteria as well as the used techniques to validate qualitative study as follow:

Primary criteria:

1) Credibility: demonstrates the correct interpretation of the interviewee's answer
2) Authenticity: related to credibility for validation using descriptive and interpretive ways, and it allows including the different opinions of others.
3) Criticality: it is to appraise the influences and the biases of the study
4) Integrity: it is related to the interviewers to ensure he or she is self-critical. Both criticality and integrity are shown via recursive and persistent verifications of explanations of the answers.

Secondary Criteria:

1) Explicitness: is to review the explanatory works of the interviewers
2) Vividness: is to present the data with artfulness and clear ways
3) Creativity: is to be able to innovate to avoid the usual way of thinking.
4) Thoroughness: is needed to connect the dots between ideas for a full understanding
5) Congruence is the display of the questions of the study, the technique, and the outcomes.
6) Sensitivity: is referenced to the sensitivity of human self-respect of the interviewees

Techniques:

1) Design consideration: is to create a design of the study with sampling conclusions
2) Data Generating: is to convey the determination of the collected data
3) Analytic: is to illustrate data findings and discover adversary descriptions.
4) Presentation: is to describe the results with supportive evidence

Interviews are numerous ways to obtain information from individuals, getting their opinions based on their expertise in their area of knowledge. Certain techniques should be followed to avoid bias and persuading answers, such as appropriate wording of the questions and avert selecting inexperienced interviews. Consistency, dependability, and trustworthiness are essential aspects to attain reliable interviews with decreased variation (Teter, 2014).

Thus, the interviews are conducted with Subject Matter Experts (SMEs) by the interviewer to prevent different interpretations.

An SME has greater than normal education, training, and work experience in a specific discipline to verify and validate a system that requires giving judgment, according to knowledge and professional expertise (Pace & Sheehan, 2002).

In the qualitative research methods, the number of interviews does not need to be a large sample size compared to the quantitative research methods. Instead, it focuses on answering how and why questions about specific situations for a better understanding. Experts in qualitative research methods suggest a number between 5 and 50 interviews as a guideline. Still, the number depends on the question's answers and when it reaches saturation by getting repeating statements (Dworkin, 2012). Additionally, some quantitative research methods such as content validation recommend conducting at least six interviews and not going beyond 10 for the acceptable results (Yusoff, 2019).

Teter (2014) covered the “Interviewing Techniques” subject and the recommended guidelines for designing the Interviews in this section, as shown in Figure 49. He illustrated the types of interviews as: Structured Interview with the preparation of obtaining the information, or Unstructured Interview to collect a variety of outlooks on a topic from various interviewees, and semi-structured interviews that has more elasticity than structured interviews but more arranged than the unstructured interviews. Figure (49) shows the 3 phases for conducting interviews.

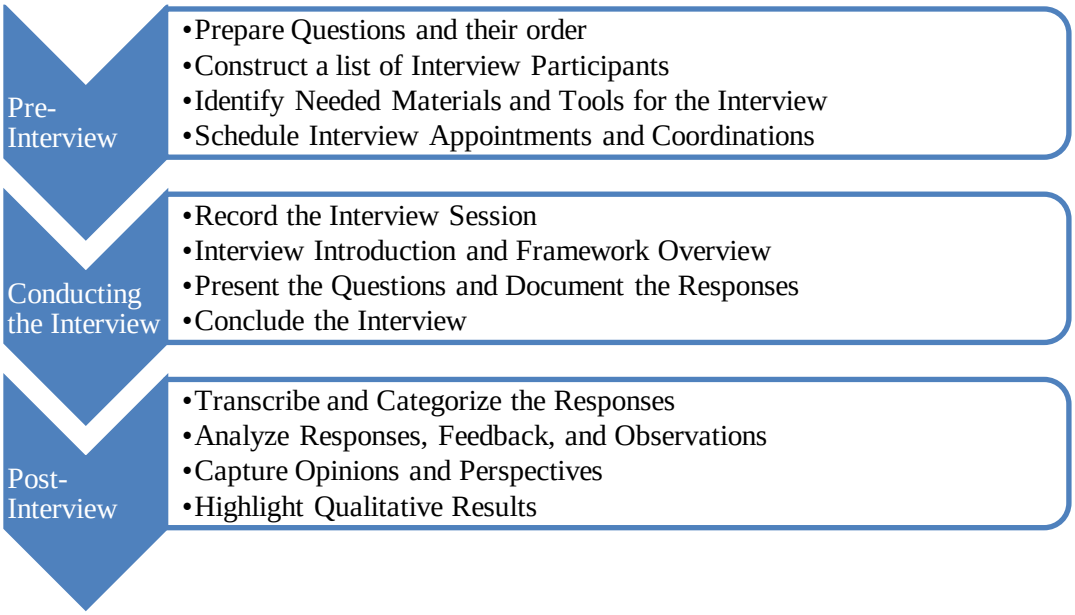


Figure 49: Interview Steps

Interview questions are designed to evaluate the framework's strength and gap by asking SMEs four questions to obtain their opinions. Professors reviewed the interview questions in the IEMS department at the University of Central Florida to ensure getting feedback on the questionnaires from the academic viewpoint and validate the methodology and techniques. It was suggested to pilot the questions of the interview and adjust as needed. Thus, piloting the questions happened with two SMEs, and the interview was changed from structured to semi-

structured interview. The four questions below satisfy the finding of the strength and gap of the framework.

Question 1: Overall, how well do you think the framework overcomes the challenges in big data Analytics adoption of healthcare?

Follow-up Question: And why?

Question 2: What does the framework provide healthcare organizations that they are NOT currently considering when implementing big data analytics?

Question 3: What step or function needs to be included in the framework to address any gap?

Question 4: In your opinion, what will be the most beneficial matter to healthcare organizations when the proposed framework structured approach is followed?

The selected SMEs' criteria are based on their education and number of years of experience in data and analytics technology in healthcare. 16 SMEs were identified and were initially contacted via text messages and phone calls. Only 12 agreed to participate with a 75% success rate. However, only 9 SME participants were needed as saturation was reached by receiving the same responses, and it was no need for additional interviews. Each participant received the instruction of the interview along with the framework before the interview for review. The interviews were scheduled individually for 30 minutes each, at the time and method of their choice and convenience, in a semi-structured format via Microsoft TEAMS and ZOOM online meeting application, except for one meeting that was in person. The meetings were recorded with the participant's permission, and a digital dictation with Microsoft Word was used to transcribe. The transcript of the interviews was tailored to clean the content for better understanding and to keep the participants anonymous. The interviewer spent 5 minutes in the

introduction and briefing on the framework and the purpose of the interviews, followed by asking the interview questions. The interview instructions and contents are included in Appendix A. The interviews were conducted with 9 SME participants with experience of 170 years (in total), with an average length of expertise of 18.9 years and ranging from 10 to 26 years of experience.

Analysis

The answers of the SMEs are analyzed to identify the successful capability of the framework and capability gaps or shortfalls. The responses of the SMEs reflected the reality aspect of using big data and analytics in the healthcare industry.

Thematic Analysis (TA) is used as the qualitative data analysis method of answering the interview questions. TA is a systematic technique to arrange, recognize, and provide insights by finding patterns and relationships of the qualitative data set. It also helps focus on realizing the noticeable meanings of the data as it has the flexibility and accessibility to find the theme (Braun & Clarke, 2012).

The meta-themes characterized from the responses of answering the four interview questions are valid. Thus, the proposed framework addresses the challenges that HCOs face when implementing big data analytics.

The interviews confirmed that using technology only won’t solve the challenges of implementing big data analytics for healthcare. However, the framework's approach to applying other components from the six layers is appropriate for successful results.

Many HCOs only focus on implementing the technology, and few apply some of the components in the framework. SMEs have concluded that including the components in the organization, people, and process layers is recognized. Listing the value of utilizing big data in

the outcome layer can help convince HCOs to invest money in big data analytics. Furthermore, treating data as a product and not as a project was very well taken, and assigning a product owner is crucial to control the process and deliver values.

The response from interviewing the SMEs identified the framework gap that includes formal training, innovation, using PDCA, and operational model, a model for continuous improvement.

Training can be part of the change management, but it is preferred to be its entity in the process layer. A self-taught attitude can lead to the discrepancy of competency and missed opportunities by not taking advantage of the full capability of the product. It will also delay making decisions by waiting on reports rather than watching insights in real-time.

Innovation is surely missed in the framework, and several types of innovation cover organization structure, processes, products and services, marketing areas (Gunday, Ulusoy, Kilic, & Alpkan, 2011). Therefore, innovation should be included as an entity that touches organization, process, and technology. PDCA is considered the core of Innovation ("ISO 56000 Innovation Website," 2021), which can be part of innovation.

The operational model is an essential element, and I recommend creating Concept of Operation (CONOPS) as part of the responsibility of the data product owner. This can be an area of future research. Table (11) provides a summary of the responses to the interview questions from the 9 SMEs.

Table 11: Summary of SME Responses

	Q1 Framework overcome challenges of adopting big data Analytics in Healthcare	Q2 Framework Components that HCOs NOT Considering	Q3 Gap and Missing Components in the Framework	Q4 Benefits of following the Proposed framework
SME #1 • Education: BS Computer Science • Years of Experience: 16 Years • Position: Data engineers • Professional Background: business analytics and business intelligence	• Yes, it covers it very well and includes the fundamental aspects. • It addresses the challenges. • It is not only the technology that is the challenge.	• The connections /interactions between layers. • Stakeholders' involvements. • Integration of systems	• Committee for selecting technology at the solution implementation level. • Adding more details. • The framework can be implemented as is.	• Allow looking at data at the enterprise level. • Impose alignment and eliminate different decisions. • Focus on business values and outcomes
SME #2 Education: BS Elec Eng. & MS Industrial Engineering Years of Experience: 17 Position: Principal, Enterprise Business Architect & Data Strategist Professional Background: data integration, data analytics, and advanced analytics.	• Yes, it utilizes the main pillars of business. • It addresses the challenges • It is not only the technology is the problem	• The bridge between technology and business • The bridge between people and process • The step-by-step approach	• The operation model is a gap. • It needs to have a road map for the future look and future state.	• Listing the outcomes identifies the value of adopting the framework. • It leads to innovation and benefits from IoT.

SME #3 Education: Ph.D. in Healthcare Management Years of Experience: 13 Position: Data and Quality Research Scientist Professional Background: research and development for the institute and comes through data, quality outcomes, working with clinics, working with the hospital, the physicians.	• This framework is what is needed • It clearly illustrates what HCOs need to implement. • Breaking down the components and interconnecting them is done well	• Defining the goal of creating metrics. • Defining the roles and ownership for data teams. • Data is a product, and it is not a project • The structured approach of the framework.	• The gap could be addressed as time goes on. • An implementation plan is needed. • Area of research and development for innovation. • Apply Plan, Do, Check, and Act (PDCA) is a good way to get feedback and continuous improvement of the framework.	• It is based on the business values. • Eliminating silos. • Eliminating inefficiencies and redundancies. • Defining the roles and responsibilities for measuring outcomes. • This structured framework is a foundation and starting point. It can be dynamic and make it live. • The framework will benefit HCOs and their patients
SME #4 Education: MS in Healthcare Administration Years of Experience: 10 Position: consultant with the value-based care for analytic Professional Background: data analytics, data aggregation for assisting any	• This framework touches and synchronizes the HCO and people. • Within the data area, it brings back IT to the business, rather than acting in silos. • It helps set the organization's goal and ties it	• Streamline connection between the organization, people, and process. • The life cycle of the data to exclude old data. • Using data completeness in real-time.	• The gap in the flow outcomes, process, and people. • Training of understanding the feature of the tools, standardizing the definition for all to speak the same language is a gap.	• Breaking down the silos • Business connects and collaborates with IT. • Will be able to add new business capabilities. • The ability to foresee and forecast the changes of the business

operational outcomes through measurement for data and helping in terms of reporting and forecasting in revenue cycle management to population health	back the data analytics. • Applying the framework should allow the implementation of big data analytics more efficiently.	• product, and it is not a project The structured approach of the framework.		• ROI perspective, operational efficiencies, reducing redundancies, and optimizing workflows.
SME #5 Education: Doctorate in Pharmacy Years of Experience: 24 Position: senior director for enterprise performance and quality goal solutions. Professional Background: corporate quality, clinician workflows, and clinical decision support.	• The framework is valid. • This framework expresses the complexity of the situation very well. • Address the challenge of the low confidence in our data quality • Make HCOs understand the scale and the complexity of the data that they preside over.	• Affectively implemented maybe half of those blocks or so • We have some elements, but we need quality, integration, ingestion, Metadata, and cataloging for maturity around data. • The life cycle of data. • A CDO was recently hired, and forming a data team is needed.	• The gap in educating hospital leadership on how to be good consumers of big data. • The flow outcomes, process, and people. • Training of understanding the feature of the tools, standardizing the definition for all to speak the same language is a gap.	• The patient outcomes drive the core mission of the HCO and financial success. • It helps streamline and makes things faster, instead of entering the report manually. • Having data lifecycle management and data governance realms. • Having a dedicated team
SME #6 Education: BS in Computer Science, MS in Information Technology	• The framework covers the most important aspect of building big	• Many HCOs do not consider Data integrity for	• I will say it's complete for the most part, but it is	• Patients' engagement by using applications on mobile devices

Management & MBA Years of Experience: 20 Position: Technical lead for the EMR/Supervisor of the database administrator/developer. Professional Background: database administration and database development, infrastructure, and enterprise architect.	data analytics for HCOs. • It addresses the integrity of the data. • It includes security and data privacy.	data accuracy. • Many HCOs do not consider the best practice of data integration for data completeness. • HCOs need to consider the other layers, not just the technology layer.	missing Innovation	• Ability to collect more data patient's data to improve their lifestyle.
SME #7 Education: MS in Management Information System Years of Experience: 26 Position: Senior data analyst Professional Background: EMR, Oracle database, and data analytics.	• The 41 blocks in the framework look very good. • Most of the data comes from big data, and data management is a challenge for big data analytics for HCOs. • Using unstructured data is a big challenge.	• Have metadata and cataloging, but we need data quality. • Have a data product delivery manager, other, but we need data management life cycle. • Defining data ownership and stewardship.	• The framework is missing the operational model. • Needs to add NLP in the technology layer.	• HCO is going to save money by adopting this framework. • The ability to leverage IoT and new technologies. • Identify the good payers from insurance companies. • Provide a better patient experience.

	• The challenge of big data is more technology source, but the framework still helps, and it is valid. • It shows the dependencies and the relationships between the entities in the six layers	• Good change management to inform data users with changes.		
SME #8 Education: Position: Enterprise account manager for solutions integrator Years of Experience: 24 Professional Background: data, AI, and ML pipelines with several healthcare organizations.	• Yes, all the base components are included in the framework. • The challenges of big data in healthcare tend to be massive and complex. • Data engineering is a challenge and is represented here in the framework • businesspeople do not get excited about data. Still, it shifts the organization's culture to a data-driven culture, and it	• HCOs do not consider the value of data, and the framework can provide the benefit to them. • HCOs tend to brush over IoT. •	• The Fusion team and the stakeholders, those two blocks can be consolidated into the center of excellence. • It is missing the audit for regulatory. • It is missing the quick feedback on itself. • Missing the minimum viable product (MVP)	• It will lead to culture change. • The ability to know the basic flow to get value from data analytics. • It is a good way to start the big data journey. • To be successful, you have to do most of these entities. Big data needs the other 40 blocks to get the desired outcomes and values.

	is good that the framework includes them as stakeholders.			
SME #9 Education: AS in Computer Science Years of Experience: 20 Position: solutions architect Professional Background: all aspects of data, AI and ML, modern data pipelines in large enterprises healthcare	• This framework is a great foundation to implement across the different pillars. • Challenges are disparate systems, age technology, and new technology that gets introduced on a somewhat regular basis to solve specific clinical problems. • Without a good, governed framework and the approach of just bringing standards, that's going to the complexity	• Many HCOs are not considering the alignment of clinician stakeholders or a business partner or IT • Applying metadata to improve patient outcomes	• Have a standard data dictionary or glossary • The framework would need to add data lineages into the framework. • The reusability of data is the gap. • Streaming component for real-time data. • Need to orchestrate the pipeline • Need to add data virtualization	• The framework has a tremendous amount of benefits, and the key is to improve patient care outcomes. • This will lead to predictive analytics and targeted healthcare based on that individual's condition. • Take advantage of the cloud software infrastructure.

Highlighted responses and comments provided by SMEs:

- Big data analytics in healthcare is different from other industries because it is more massive and complex.
- The design of the big data solution should be reviewed and approved at the technical and engineering level by cyber security, infrastructure, and cloud teams to ensure appropriate design.
- The challenges of big data analytics in healthcare are:
    - Age of data and systems
    - Disparity of systems
    - Data is not trusted to be accurate
    - Data has no ownership/stewardship
    - Business units are not aligned with the goals of big data analytics that causes silos and redundancy of tools.
    - Lack of standardization of term definitions and glossary of wording
    - Regulatory requirements
    - The focus in implementing the big data technology without considering other aspects
    - Big data technologies are evolving very rapidly
- Finance controls the decision of investing in big data and argues the value of adopting it. They only approved for a specific business case and a clear understanding of the value added.

- It is overwhelming to handle data analytics without proper training. Business leaders are not trained to use big data technology in real-time, and they rather rely on getting reports sent to them, despite the delays of getting the information.
- Cloud technology is the preferred method for big data and analytics implementation for agility, scalability, and elasticity.
- Data needs to be accurate, precise, and in real-time
- Big data is one pillar of digital transformation that builds the pipeline feeds for AI and ML.
- Data should be handled as a product, and big data should not be treated as a project
- Big data will change the cutler of the organization in many aspects, including decision making, process, and the number of needed employees.
- Research and development for innovation are needed in the framework.
- HCOs need to do more predictive analytics, make decisions based on data completeness in real-time, and not necessarily just paint something that they already know.
- Physicians use data are different, with a variety of skill sets. Some of them are exceptionally data-driven, and they are naturally very process-oriented, but this is not self-taught in most cases.

Both case studies and interviews are aligned on many points.

CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the overall findings of the research and ascertains the benefits of the framework of big data analytics in real-time for healthcare enterprises. The contribution of the framework is to acknowledge the structured approach taken by following the Industrial Engineering BoK to deliver a successful solution of enabling data analytics for the healthcare industry.

Summary of Research

Big data is one of the crucial components of data science. It is enormous data of 3 Vs; volume, variety, and velocity that brings value to healthcare and helps HCOs achieve their vision of providing a better quality of care with lower cost to their patients. The proposed framework answered the question, "how to introduce the right steps and technologies in a framework that consists of the right components of big data technologies to create the needed performance measurements reports to provide real-time and accuracy outcomes in healthcare?"

CHAPTER TWO presents the recent research in big data, analytics, metrics, and healthcare to develop the framework with an end-to-end solution that delivers value to HCOs. Again, the literature showed an existing framework, but all focus on specific use cases and not the enterprise level.

CHAPTER THREE delineated the necessary steps to accomplish the goal of the research and explained its idea. The research gap analysis deduced the need to integrate structured, semi-structured, and unstructured data before performing data analytics to ensure completeness of the information and offered a preliminary architecture for implementation.

CHAPTER FOUR presents the framework of big data analytics that fits the HC industry and synthesizes the stakeholders, source of data, metrics, and big data technologies to enable the business and deliver value. It identified the relation of data, quality, and profit and showed benefits to the framework. Additionally, it includes the cost analysis of conducting 1 TB of big data analytics in a public cloud.

CHAPTER FIVE covers the system architecture of the framework and the validation using case studies and interviews with SMEs.

Contribution to the Body of Knowledge

This study focused on providing an end-to-end solution with a structured approach by developing a comprehensive framework to assist HCOs in assembling their big data journey successfully. The framework is the answer to the research question "how to introduce the right processes in a framework that consists of the appropriate components of big data and analytics technologies to create value from the voluminous data and accurate outcomes in real-time for healthcare organizations?".

The framework was able to be implemented in a live hospital setting. Implementing the framework within an HCO brings to light the many obstacles and perspectives that other HCOs will face when adapting to big data technologies and analytics. Throughout the research, the work contributes to the healthcare industry that recommends a clear path for adopting the necessary components to show value and boost the bottom line.

Healthcare Organizations is complex industry that requires many different considerations and perspectives depending on the sector within the hospital. This study examines the perspective of the IT, Business, and Clinical departments. The proposed framework shows the necessity of the collaboration between these sectors to integrate their perspective and the big data

and analytic tools to produce a comprehensive structure for HCO seeking to utilize big data effectively.

Furthermore, the study also inspected a lighthouse project to test the feasibility of a proof of concept and associated cost. This highlights if the transition to using big data and analytic tools in the cloud is beneficial and compares it to the cost of implementation. Thus, bringing value to the business sector of the HCO by providing insight on the TCO and OpEx.

The proposed framework in this study offers the healthcare community an effective way to collect, store, and deliver data outcomes in real-time to help benefit the decision-making in the business and clinical sectors.

Conclusions

Big data and analytics have become a pivotal role in expanding structured, semi-structured, and unstructured data. Big data analytics takes BI of reporting, dashboard, and scorecards to the more superior level of analysis with descriptive, predictive, and prescriptive analytics. The value creation of big data analytics in healthcare came because of the forced demands in the industry to make a positive impact using innovation with data and new emerging technology such as big data.

As in any other industry, high-performing HCOs took the lead, exploiting the big data to their advantage and leveraging it as a strategic asset to improve care delivery. Low-performing HCOs that wait to adopt new technology are at financial risk. Other HCOs are absorbing them in merges and acquisitions because they cannot stay in business.

Instead of focusing on only one dimension of the implementation consideration, such as the big data technologies needed, this framework proposes a reliable structure approach for HCOs to adopt big data and analytics that includes 41 entities in 6 layers that integrate the

varying dimensions HCO has. The six layers include Organization, People, Process, Technology, Data, and Outcome as shown in figure (50).

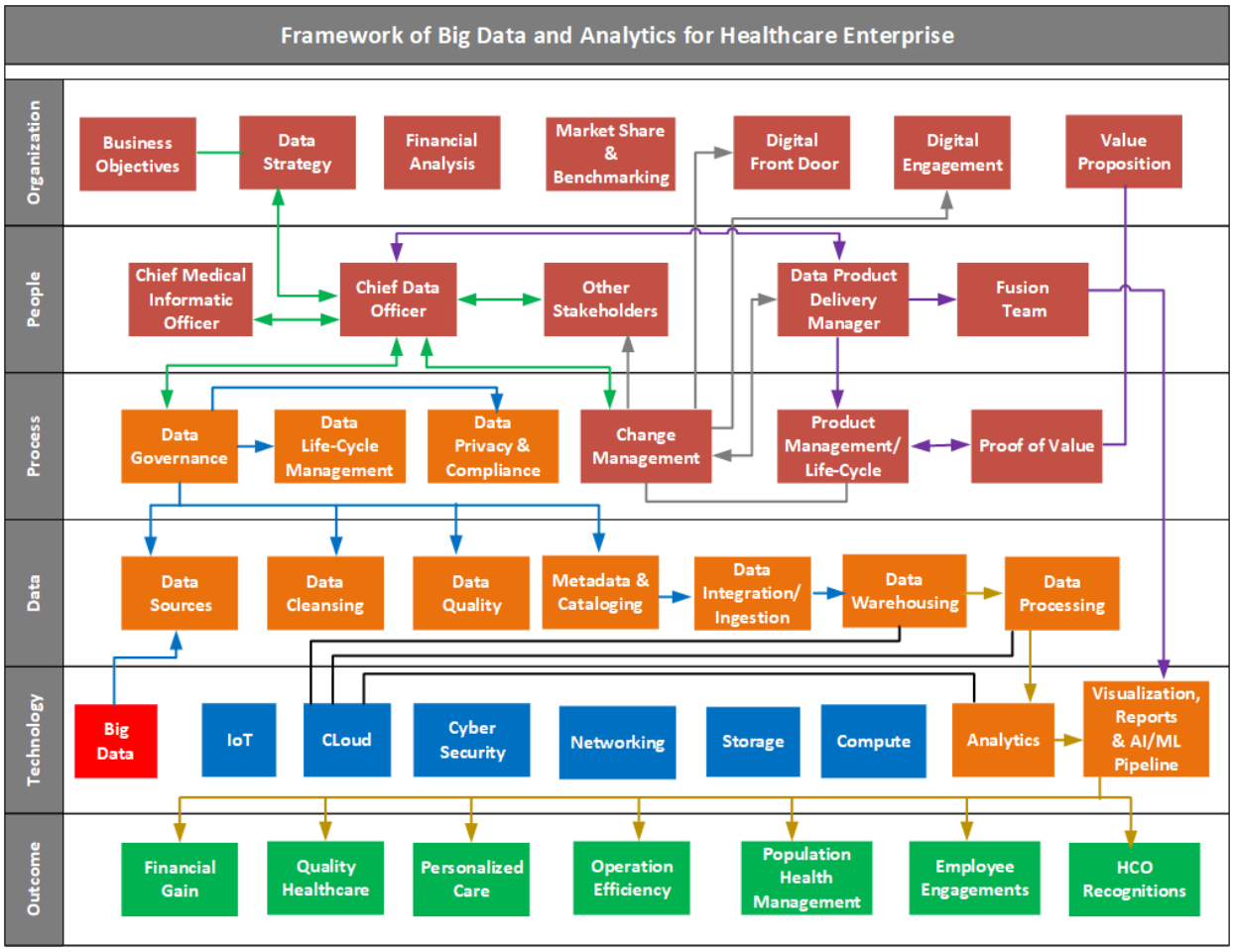


Figure 50: Proposed Framework

This study highlights the value of having a framework conducive to multiple sectors of a HCO, which can help elevate the care given to patients and allow hospitals to gain recognition and funding by providing better patient care by examining a case study performed at a Hospital in Florida. In addition, this case study validates the framework by demonstrating the benefits and value that adopting the proposed framework can bring to an HCO.

In addition, subject matter experts were interviewed, bringing an outsider's perspective to validate the proposed framework through a series of open-ended questions—the SME opinions aligned in how practical this framework can be if implemented in an HCO.

In conclusion, the proposed framework in this study is validated through a case study and interviews with SMEs to be a foundation for HCO seeking to utilize big data and analytic tools to improve the quality of healthcare and decision making.

Limitation of Research

The goal of this study was to fill the gap that did not consider the many facets of the healthcare industry. Therefore, a framework was proposed that incorporated 41 entities within six layers covering the many sections in HCO.

Using the case study and SME to validate the framework, it was found that the proposed framework can provide the necessary foundation that many HCOs lack. However, the findings would be strengthened if a direct comparison in utilized frameworks can be made. This can be done by comparing HCO with different framework structures.

In addition to having a direct comparison, the study suffers from many confounding variables such as other technological advances, new leadership, covid 19 pandemic, and continuous changing of government regulations. To what extent did the proposed framework benefit the organization cannot be fully examined due to the complexity of the industry.

To further analyze the benefits and value of the framework, a concrete ROI would be beneficial for HCO to be more interested in adopting the framework. Unfortunately, this study did not incorporate an ROI due to the confounding variables. Therefore, it could not be conclusive in providing an ROI.

Future Work and Recommended Research

This research study did not consider the quantitative method as a validating method for the framework. The goal is to find insight with open-ended questions in a non-numerical fashion for better comprehension of the challenges that HCOs experience when big data and analytics are implemented. In future work, it will be beneficial to approach the research using validation methods that include surveys when more participants in the field of big data analytics in healthcare meet the minimum survey sample.

This study introduces the framework to utilize big data analytics in real-time for healthcare enterprise performance measurements by incorporating all essential components. In addition, big data feeds other analytics technology such as deep learning, machine learning, and Artificial intelligence that can take performance measurements to more advanced outcomes. Thus, big data won't be useful without AI and ML, and vice versa. The AI, DL, and ML are not fully capable without taking advantage of data completeness, including big data. Figure 50 shows how big data is the pipeline of AI and ML.

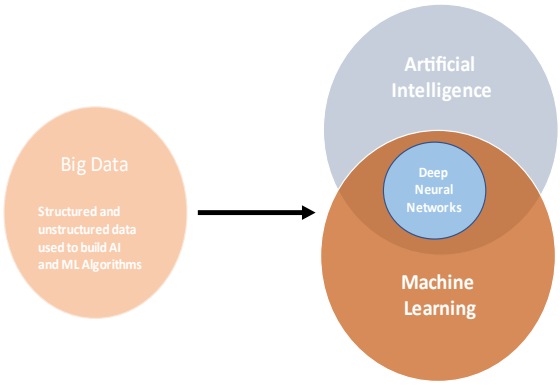


Figure 51: Big Data and relationship with AI and ML
Recreated and Modified from "Promise and perils of big data and artificial intelligence in clinical medicine and biomedical research"(Rodriguez, Scheinker, & Harrington, 2018)

Further research can explore how the inclusion of these technologies can improve the cost of healthcare. Especially, many retailers are entering the healthcare business through virtual care clinics and primary care that require digital capabilities to minimize the cost of treatment. This area of research can help the retailer deliver basic healthcare needs to customers by shopping health services with transparency and competition over price without the involvement of insurance companies.

NLP-based technology solutions suffer from having consistent false positive and false negative outcomes when it is used with unstructured data analytics. This is a gap that future research using NLP technology should be conducted to help improve this chasm's vital needs that help extricate precious information from big data. Additionally, outdated data creates noises,

and researchers need to explorer more effective data life-cycle practice to include the appropriate data when analytics is performed to avoid wrong results.

Block Chain is another emerging technology that is believed to be pivotal to make big data and exchange heath information data between HCOs and providers more efficient and controlled by patients, the actual owner of the health information. This should streamline the health information flow and comply with the government regulations such as HIPAA. Future research utilizing block chain to automate exchanging patient health information and control the patient, the actual owner of their health information, on who can access their health records in real-time with one click.

Additionally, I highly recommend future research to address health inequality because of race and social level. They have a higher death rate and use these emerging technologies to improve their healthcare, especially in rural and suburban areas.

Moreover, the Concept of Operations (CONOPS) for big data and analytics research is needed to provide guiding principles for supporting the operational model using the proposed framework as a foundation and guidance. The published "Concept of Operations: Data Capture and Management Research Data Exchange" can be used to reference (DOT:Website, 2011).

The recommended research areas above are related to the industrial engineering discipline. However, other research in the computer science areas can also help in big data and analytics technologies.

APPENDIX A TRANSCRIPT OF INTERVIEWS

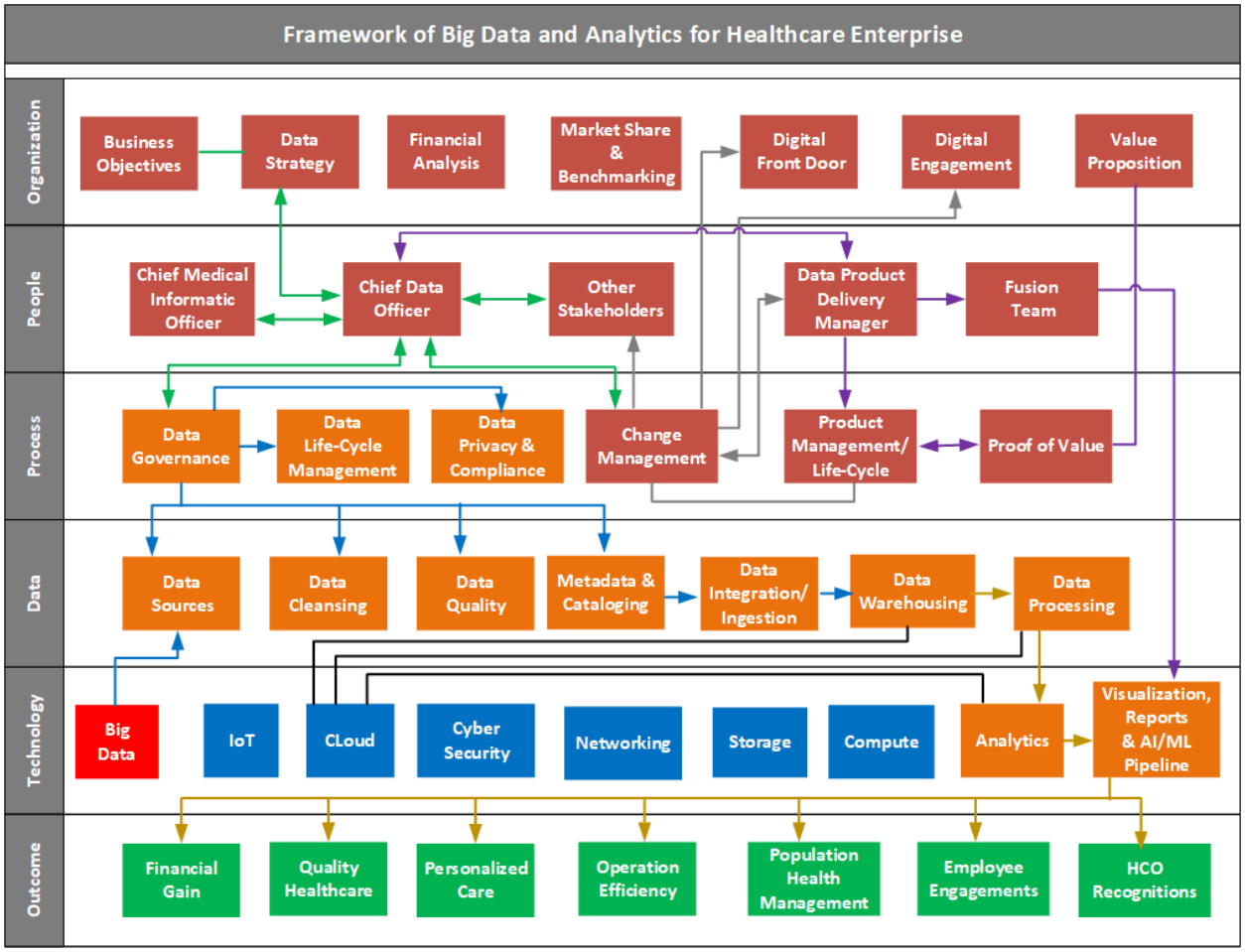
Interview Instructions

Interview Agenda:

- Interview Instructions – 2 Minutes
- Framework Overview– 3 Minutes
- Interviewee/SME Introduction and Background – 3 Minutes
    - Education
    - Professional Experience and Titles
    - Years of Experience
- Interview Questions – 20 Minutes
- Interview Conclusion – 2 Minute

Framework Overview

The proposed framework shown in the figure below includes 41 essential entities categorized in 6 different areas: Organization, People, Processes, Data, Technology and Outcomes.



Each entity represents a function, or a role and some entities are linked to one or more other entities. The standalone entities that are not linked to any other entity, but their presence is crucial to the framework for a complete view. The linked entities are connected with a straight line to show the linkage between them. The one-way or the two-way arrow indicates that there is an interaction between the two entities in one direction or in both directions. Additionally, each

color of the line or the arrow implies a relationship between several entities that interact together in several but similar channels.

The five colored blocks display the entities in groups to construct a cluster of a distinguish purpose. The big data entity in red block is the main component in the framework and it is under the technology area. The orange blocks are the core of the framework that place all data and analytics entities in one group, and they are distributed among processes, data, and technology areas. The blue blocks show the needed technologies that involve in utilizing big data and analytics. The brown blocks are the supportive entities in the framework that guarantee a successful result. The green blocks are the outcomes that show the expected value of adopting the big data analytics. The proposed framework is very refined, and each entity encompasses a huge amount of work and resources.

Interview Questions:

Interview's questions are designed to evaluate the strength and the gap of the framework by asking SMEs to obtain their opinions.

Question 1: Overall, how well do you think the framework overcome the challenges in healthcare big data Analytics adoption?

Follow up Question: And why?

Question 2: What does the framework provide to healthcare organizations that they are NOT currently considering when implementing big data analytics?

Question 3: What step or function does need to be included to the framework to address any gap?

Question 4: In your opinion, what will be the most benefiting matter to healthcare organizations when the proposed framework structured approached is followed?

Interview 1

Interviewer: Good afternoon, how are you today?

Interviewee: Good afternoon I'm doing pretty good how are you?

Interviewer: I provided the interview agenda earlier and it's going to be around 2 minutes talking about the instructions. It is recorded and we are using zoom. I will be on video, but you don't have to. It's going to be about 20 minutes for 4 questions with a total about half an hour. You're going to introduce yourself, talking about your background, education, years of experience in data and analytics field, and your experience in healthcare. Each of the four questions is going to take about 5 minutes to answer. I provided you the framework and it has six layers. It is based on people, process and technology model that was established in early 1960s and I built around. I added three different layers which, is the organization to talk about health care as a business, data which, is that healthcare organizations generate. And finally, the value or the outcomes of the framework. The framework has 41 components as presented in at 41 blocks. Some of the blocks are very much independent from each other's and others are connected because of the interactions. There are colors coded. Orange is very much related to processes, blue is for technology, big data is in red because that's the main component to introduce data completeness, the value which, is green and organization or the healthcare as a business is in brown color. I will stop here and if you have any questions before we continue.

Interviewee: No, that sounds great. Thank you for the introduction explaining what exactly needed. You know this is a very interesting subject, questions are always there, you know it's a huge subject and hopefully, I can provide valuable information and can help actually

Interviewer: Thank you. So, if you can introduce yourself and your background.

Interviewee: yes, absolutely so my name is as XXX. I've been working in this industry, data analytics software development for nearly 16 years and currently on my data warehousing is

my focus but obviously we act like engineers, business analytics, business intelligence evolves as well so we do coordinate work with each other. So pretty much we touch all different areas for the purpose of providing a complete solution for the organization.

Interviewer: your education if you don't mind.

Interviewee: Oh yeah sorry, so I have a bachelor's degree in computer science from the UK and I have a technical diploma in programming. I'm also Microsoft certified

Interviewer: I am very impressed with your background, and I really appreciate your time today to give me feedback on the framework. Before Starting with the first question, let me ask you first if you have looked at the framework.

Interviewee: It's very interesting actually. I looked at it and I reviewed the components. I've went through the explanations that you added on how these components interact, the purpose of the layers that you have there. And as a matter fact, it basically relates very well with what's going on in the industry and specially in the healthcare industry because analytics it's pretty much the same everywhere (other industries), but except when we talk about the healthcare industry. It becomes very much complicated for several reasons as you know, volume, velocity. The different in healthcare is that we need to answer in very quick and efficient manner. So, it seems very complicated and very cumbersome. A lot of different types of data that comes from everywhere.

Interviewer: The first question is "overall, how well do you think the framework overcomes the challenges in healthcare big data analytics adoption.

Interviewee: your proposal covers it very well and talks about all different angles. It's very much you know looking at the at the entire ecosystem, so it's coming a lot of the technicality of data big data and how healthcare can adopt them.

Interviewer: so why do you think it's the challenges what do you see the challenges and in health care and why is the framework overcome these challenges.

Interviewer: OK so again like I said, healthcare data in general is everywhere and it's across the board. It's more complicated in healthcare because of volume because of several things that goes around healthcare in general so the reason why, because it covers a lot. First of all, let's talk about the challenges are the fact that the organization as a whole looks at the data or the analytics only from one angle, which is technology that is being used to achieve the analytics only or to achieve those goals of the organization which, is not a right thing to do. It's very challenging and it causes a silo. The fact that there are a lot of moving parts and a lot of components that contribute to the success of this entire framework. So it's not only the technology, it's as matter fact lot of other areas like you mentioned in your proposed framework like the organization, people, processes, data, technology and so it's all together they all come together to basically solve those challenges and not we cannot look at it only from a technology perspective because it's always different technology every day and it needs to be selected based on your requirements and your need and also the needs of people that supports the technology is selected this order for this whole thing becomes successful as one unit so that's what was very important in your framework that is proposing. And I believe it covers all the areas.

Interviewer: Second question “what does the framework provide to healthcare organization that they are not counter currently considering when implementing big data analytics

Interview: Alright so what the framework provides is basically the connection between all the layers and between all these components that you have there because that communication is very much important to overcome the challenges that healthcare reliance. It would help the

healthcare puts on the implementation of big data analytics or analytics in general. So, when I talk about communication that means not only like communication among all different stakeholders among, but it is also for all different components and departments across the organization to achieve that goal and your framework here is basically mentions that very specifically and it addresses this problem from all the different angles. So, integration is very important, and the solution does not come only from one place, it comes from the contribution of all different entities in the organization.

Interviewer: question #3 “what step or function does need to be included in the framework to address any gap”. I'm trying to get the gap and missing in the framework.

Interviewee: Looking at it, it pretty much covers a lot of the important factors and from again like I told you as I've mentioned earlier that I'm one of the solution architects, so you know if I look at it from my perspective, I think I go more deeper into the components, but I understand that the framework covers a conceptual view at high level. The entire framework obviously you cannot mention every small detail here but what I feel is very important, is having specially in the people section or in the people area. what I think is important is having a committee for selecting the technology, architecture board review or data analytics council that will help the organization putting together the best architecture and technical solution for the whole thing. The reason why is because the ecosystem in healthcare for big data and analytics have all different technologies that are being available right now. Technical people all need to talk about what the right technologies to pick, they all need to talk very well with each other in order to complete this whole picture in order to be able to successfully achieve the outcomes.

Interviewer: so, what you are saying the gab is very much more technology specific or the way how things are selected.

Interviewee: The gap is more about stressing into the details. When it comes to the details, I think the selection process of the technology needs to be very clear in the framework. It's just as important as all the other component that I see.

Interviewer: Let me let me ask you about the gap point in this framework. can you implement it, or you feel it is just a conceptual framework?

Interviewee: it can implement it.

Interviewer: question #4 “In your opinion, what will be the most benefiting matter to healthcare organizations when the proposed framework structured approached is followed?” what is the value that your organization and maybe you can talk about if a similar framework has been implemented within your organization and when you saw the framework you say wow if we added this component, we will be able to address some of the challenges?

Interviewee: yeah, like I've mentioned earlier. I've been with the organization nearly more than seven years so I've lived the Upson downs I've lived the different decisions that were made to be able to come to a certain platform or to be able to provide our customers the best quality or the best of breed technologies and it's been evolving and out of common lift the struggles and the reason or what I've seen different was the fact that you know technology part was always separated from the entire organization. Meaning that it was the other way around, where we basically go and find the technology and look at the best of breed in the market for that the technology without considering what is it exactly. That without involving stakeholders, without looking at the overall outcome. It was more focused on very specific technologies and the cycle was. Take certain technologies will both flow through to finance for example, they're talking only about that specific technology that we're bringing to the organization. it's not going to add a value to which is very wrong I believe, and I believe this framework here helps that we

are heading towards this framework and that's the only one you're seeing or showing success in what we're doing and that's due to the several things. Again, the healthcare industry you know how challenging it is to collect all these data and all the information to put together, store it, and make sense of it. So, all that is you know needs to be considered and not only from a technology perspective like I mentioned it also needs to be considered from what matches our processes and what matches the vision of our people and the vision of our organization and how is that is in line with the outcome or with the overall vision of the entire organization. So, your framework is pretty much covering that and it's sitting on that and it's allowing the shareability and the integration that the focus overall versus the specifics.

Interviewer: So, I want to follow up with some other question, about costs and value. If you see the value in the outcomes. I'm not saying you implemented the same framework, but the value of data and big data completeness require you look at the whole entire value you look at the whole entire data. So the question here, have you looked at the cost and return of investment. It is investment in data and are you getting the value or the expected outcomes.

Interviewee: absolutely for the longest time it was like I mentioned it was the other way around so we basically like they say - we used to put the cart in front of the horse- what that means is basically we go to finance or you know people that control the budgeting and all the costs and all that stuff and talk about specific components and then discuss or basically argue the value of that specific component to the organization and the cost of that specific which is absolutely not the right way to handle that or to approach that. The reason why is because it's not about one software it's how we look at and how we handle it. We must emphasize that complete entire infrastructure and how much investment are we willing to put to see a better return of investment on the on the overall. So basically, big data after going through struggles and

understanding that what we were doing was not right for the purpose of technology that we were using. We were not in line with the amount of data we were capturing and was not supporting the things that we were asked for to deliver. For that reason, they were always paying extra money on storage, paying extra money on the technology itself, on the tools like the BI tools that are on top of these technologies and we always used to reach the end in terms of performance, in terms of agility and how fast we can deliver to our customers which we can give all the stuff and make the customer happy when we talk about things but when it comes to deliver, reality hits and then you know it becomes trust issue between us and the stakeholders because we start delivering in a longer or worst fashion that we used to order. So, this framework for big data adds that missing piece or basically help us recover that specific piece because then we will be able to define the way we are providing our outcomes to business and what that left is basically allowing the business to be pleased. But obviously using the benefit of the framework to allow the agility and all the stuff that we can do with this technology while we're also prepping and making sure all other components like creating the data providing predictive analysis, providing executive dashboards that basically compares things and numbers as well as allowing our data scientists to post models that will enhance the quality of care that will also enhance improve budgeting, improve the financial outcome of our organization. So that all together basically tells you the whole story.

Interviewer: I know we almost 6:00 PM now. One thing I want to ask do you guys use cloud because you mentioned agility.

Interviewee: We do use cloud and we are in the process right now of doing evaluation and again like I mentioned the software evaluation piece was very important because you know there's a lot of environments out there and a lot of offerings. It is very hard for us to go to a

specific one because we've been through this over and over again. We had big data on Oracle appliances which is very specific technology and what happened now Oracle decided to basically decommission this line of business which, led us to basically find another solution for us. That's the reason we're moving to the cloud to achieve several things. it is the ability to be free from being stocked in a specific technology. Whatever technology is out there so if one is obsolete, we can easily carry out the entire solution to another technology without going through what we're going through now. So, to answer your question yes, we are going or we're heading towards the cloud. We're not fully functional but we are in the process.

Interviewer: I note 6:00 PM. I promised it's going to be 30 minutes and I'd like to conclude by asking do you have any questions for me, any suggestion, anything you want to add.

Interviewee: My only question would be why you're building this framework and talking about looking at all different areas, is the technology piece of it is that something you're stressing on or you're diving deep?

Interviewer: yes, it is included in my dissertation, but the interview is very much about the framework itself from the system architect or from industrial engineering point of view. However, I included how you pick big data and analytics technology using the cloud and what the criteria of selection. What is available and like you mentioned things moving very quickly or very fast. Three years ago, the technology was much different from today so the technology keeps changing and you cannot keep up with it. Unless you are performing analytics in the cloud so you can use whatever analytics you need. I appreciate your question. Thank you so much I would like to thank you for giving me the opportunity to talk to you. I know you're very busy and I really appreciate your time and your expertise.

Interviewee: will always help and I always share knowledge. Thank you so much. And last thing I want to say, they say that data is the fuel nowadays. So, when data is used, we can take advantage of our fuel and not throw it away. I think your framework will make the best out of our organization and our business.

Interviewer: Again, thank you for giving me the time and enjoy your evening

Interviewee: I appreciate it thank you so much

Interview: thank you

Interview 2

Interviewer: Thank you very much for allowing me to do the interview hopefully you were able to review the instruction and the framework. As you saw the framework is 41 blocks and each block and it is based on people technology and process model. You can see here the organization, data, and outcomes are add to the framework so I will ask you four questions and each question you will have five minutes to provide your answer based on your experience. One thing I would like to do first if he can provide your background, your professional experience your education, your title, and years of experience. Then we will handle the questions that will take about 20 minutes. So, let's start.

Interviewee: my name is xxxxx. I have been working with data analytics since 2005 and the overall years of experience are 17 years and I worked in different business verticals like healthcare, clinical trials, financial services, insurance sectors multiple sectors to name few. My focus is with the data integration, data analytics and advanced analytics. And currently I also get involved in a lot of large-scale assessments of enterprises in terms of data strategy. I create data strategies for companies, and I advise CDO of organizations. From the education perspective, I have a bachelor's degree in electrical engineering and a master's degree in industrial and systems engineering. I work as Principal, Enterprise Business Architect & Data Strategist and basically, I work with a management consultant company. Prior to that I worked for a company in data strategy and digital transformation with the data centric kind of mindset

Interviewer: Thank you. So, let's start the first question. "Overall, how well do you think the framework overcome the challenges in healthcare big data Analytics adoption?"

Interviewee: Actually, what I like about the framework is that it's not a technology focused, and it utilizes we call the main pillars of any business. So, you look at the people's perspective, you

look at the technology perspective, you look at how the organization is structured, and you look at the processes involved. Of course, the data is the main kind of crude material here. So basically, what I like about the framework is it's covering all of the major aspects because it's not always a technology problem that you are solving and most of companies and organizations they really fail because they don't take care of the other pillars, so the challenges of healthcare and healthcare big data.

Interviewers: what do you think what the challenges is

Interviewee: OK yeah so there are different ways to classify challenges right. So, if we look at from the patient's perspective, they have usually to depend on moving healthcare records from provider to provider. There's no easy way to integrate an exchange the data and share the data across the institutions that the customer is treated at. It's not necessarily you get treated in one location so if you go to one hospital and then you move to another. it's it's really hardship for the patients to handle the data perspective. Customers of data, I call them usually stakeholders. So patient is one of the stakeholders in this process and the health care worker is another stakeholder in the process. It gets overwhelming for them to handle the data, to manage the data, and sometimes even for doctors to have all data in under their single view rather than making decisions without having the full data about their patients. So, it's really important to have accurate, precise and on time data with the high-quality for the healthcare provider to make the best decision in favor of the of the patients. There's another third stakeholder which is basically management right so without having a good robust managed data analytics, it's hard to manage without having the right analytics because then the healthcare management they will make decisions that are not based on data insight and that might introduce some errors and in these

decisions it might introduce some kind of gaps that might be covered by having to the right analytics.

Interviewer: question number two “what does the framework provide to healthcare organization that they are not currently considering when implementing big data analytics”

Interviewee: To me it's really the main thing that this framework provides is the integration of technology and business. It's very hard to for organizations to handle the business side without technology and vice versa. There's always that gap between technology and business. This framework bridges that gap and make sure that adoption of any data analytics solutions is going to be easier than what it is today because you are involving people and the process. Those two are important for adoption and the framework is really realizing that. I like it for that reason.

Interviewers: will talk about the gap in question number three “What step or function does need to be included to the framework to address any gap?” what you would add if you were building or creating this framework.

Interviewee: two things that I can come up with in regards of gaps or maybe enhancements that you can add to the framework. Number one is ops model, or operating model. So, the framework stated the pillars but how they are operating together in harmony and how they you harmonize all of these pillars and stakeholders. That's one thing which is the operation model for data analytics organization. The second is basically once you have the operating model, once you already have the stakeholders defined and all the pillars defined you need to really create a change management plan. So, change management plan is basically an actionable plan that takes into the consideration every single one of these stakeholders like or elements that

people, the process, how are you going to change the organization. It's always good to have the future look and the future state but do you really need to specify road map to get from current state to the future state in transitional way. So basically, it's a crawl walk run you can't just jump from zero to 100.

You have to crawl, walk and run and that should be in the change management plan and also it should be a clear road map clear and actual road map toward that.

Interviewer: Just as a note, you can find change management included in the processes layer. The operational and operation model is a great point how you can create or what to suggest about the operational model

Interviewee: so, the operating model is basically you develop it based on the organization that you have in your enterprise and how to develop, it basically has to be creating roles and responsibilities of every organization toward the journey of data analytics. So, you link and define the relationship between these stakeholders, who's responsible, who's accountable, and who has the responsibility to implement. Those are what I meant by defining the roles and responsibilities. And then everyone knows what to do, instead of just giving them the data analytics. You have to build something; you have to guide them through how they interact with each other. And I see the model is having blocks, but they need guidance. They don't understand like if you say for instance change management, what does change management mean in order for them to operate, not just what change management mean in itself, but how it relates to what they're doing today because you're not going to hire a completely new team to do all of this. You have to upskill the current existing staff and teach them how to operate, how to manage because each group has a responsibility and it's not the same. The details of each box it's another dimension

Interviewer: let me ask question number 4 “In your opinion, what will be the most benefiting matter to healthcare organizations when the proposed framework structured approached is followed?” So, what's the benefit or what's the value when they implement or adapt this framework would be.

Interviewee: I think the first value that I can suggest or say from this out from a successful implementation of this framework is increasing the quality of the human life and better quality of living for humans. The second is basically healthcare sector is that spending a lot of money and resources that most of the time they don't return value compared to the investment in money. So, with this framework, healthcare organizations can know precisely where they can save money and where they can improve the operation efficiency. It can save the money spent on the healthcare system, not just on the private sector but also for the government. Big part of the money spent in healthcare is coming from government organizations and government programs and that will help also saving the taxpayers money. A third value that they will get from here is basically innovation and being innovative organization. Innovation can be sometimes in successful clinical trials. Also, innovation in terms of being more proactive than reactive to the situations that the humans are encountering today in COVID-19. With data analytics, we can predict when there is an outbreak of certain disease ahead of time. So, the healthcare organization will be prepared to respond rather than react to an outbreak in any regions or any places. So, these kinds of things that healthcare organizations cannot do easily today but with this framework and with the right data analytics and algorithms, there will be far better than what we have seen before. So, innovation is really one of the areas everyone also talks about. Intelligence and automation are other areas where there's a system with the people. They are overloaded with tasks and the pandemic of COVID-19 is the best example where the

healthcare staff are exhausted. If you can automate and help augment their ability to handle more patients using automation or intelligence that will also be an advantage for the value of this framework.

Interviewer: so, let's talk about big data specifically, the value of the big data? Why big data has a value in this framework?

Interviewee: it is one box out of 41 but the other 40 are necessary but the main thing which is a big data, and which is innovation. So, when we say big data is basically it's not always the volume it's also the variety of data, so basically right now we have explosion of IoT devices. Healthcare is one of the major sectors that they are benefiting from IoT and also implementing IoT of personal wearable devices. They are collecting metrics and data day and night and all of the data today not really usefully. With big data stacked or ecosystem, you can leverage all of this delusion of data in terms of humans related or environmental related and then with the same techniques with AI and machine learning, you can also correlate this data together to come up with the best outcome for humans to improve human life. So, basically big data is equivalent to innovation so with the big data ecosystem including the algorithms built on big data along with the best data quality of the big data that we collect, we really can leverage this to take the human life to the next level which, is innovative way that helps the organization that adopts big data to be the leader and in their field rather than a lagger of other organizations and other competitors. So, that's what I see big data is only increasing and having prepared for big data kind of receiving data, managing data. Massage in curating the data you need a certain level of technology and an architecture that big data provides to handle all of this amount of data

Interviewer: while we are talking about innovation and emerging technology, do you see cloud used for IoT. Maybe you can give us some background on big data in cloud computing.

Interviewee: Big data is one of the pillars of digital transformation and an innovation right. Basically, to transform your organization or your business, you need different things, not just a big data. So, big data is really definitely important but at the same time cloud is another pillar for digital transformation, AI and ML are other pillars. So, all of these are work together. When we talk about big data, all of the technologies of big data, it's better to be on the hosted in the cloud for you to leverage innovation tools. What I mean by that is basically big data you need a huge computing and storage when you ingest the data but you're not ingesting data all the time, you're not analyzing data all the time. Having elastic compute, an elastic storage. it's really going to help. You are merging the good things from the cloud and the good things from the big data stack together to save the organization money and to innovate better and not to spend too much on capital cost. When you have elastic compute and power, you don't spend as much money as you do when you do all this on premise it's doable in premise but it's going to be you need more people to handle the devices and equipment to manage the architecture all that but if you put everything in the cloud basically you are leveraging your staff to focus on how to use the data not how to man manage the tedious tasks behind the scene like maintaining server and maintaining, troubleshooting errors in the operating system, patching. All of those you can outsource them to the cloud provider. Also, another element that we didn't talk about is security. so no matter how good you are how you secure your big data on premise you're not as good as the big cloud providers with their massive kind of security posture and the security support that they have so for that reason I I think I can't they're not mutually exclusive it's not like you know big data or cloud but in my opinion it's cloud big data in the cloud so it is it is very much that's what that recommendation of the framework. So, this is spot on. That's what I see in the industry in the real world and it's not just theory but most companies they are trying to push to the cloud

and there are like even some people they say, oh the cloud is not secure, it's not in compliance of our HIPAA but there are instances nowadays that are high highly classified in terms for secure data and secure hosting of the data. There's no kind of regulations that prohibit that.

Interviewer: I don't want to take more time out and I would like to conclude. Do you have any questions, any suggestion, anything you think?

Interviewee: I think it's really a great framework and it's very modern. It's up to date like it's taking all of the latest and greatest and it tackles not just the technology problems, but it also emphasizes of leads by the business needs. So, I think that's really awesome and great way of tackling the issue because most of the time, I see theory, or I see more focus on technology, but this framework is very balanced and balanced between the people, process, technology, and data and that's what it should be.

Interviewer: thank you very much for your time and I really appreciate it.

Interviewee: you are welcome good luck

Interview 3

Interviewer: good afternoon. Thank you very much for joining. This is a 30-minute semi structured interview and hopefully you looked at the framework before our meeting. You have 3 minutes to introduce yourself, your education, professional expertise, and years of experience. You, as a subject matter expert, would be able to help validating the framework or in another word evaluating the framework by answering 4 questions. It will take about 20 minutes, 5 minutes to answer each question. Then you have two minutes for conclusion or for any last thought. The framework has 41 blocks, in 6 layers. It is based on people, process and technology model that was created 1960s but three layers are added to the framework which is organization level, data and the outcomes layer. So, three layers foundation and three layers added. It shows how each block interact with each other. Each block has its own color, big data is one block is in red which is the focus in the framework, but it will require another 40 blocks for successful implementation. Please introduce yourself, education, professional experience and years of experience.

Interviewee: thank you for having me. good afternoon my name is XXXXX. I am a research scientist in health Research Institute. My background, I have been in the research and analytics space now for 13 years. I've worked in many aspects of the healthcare industry from administration and strategy, working on implementation of those strategies as well as analytics. I also work in the research side as well so using those skills and bring the same type of thing to the field which is analytics strategic development. I am also an adjunct professor in the health systems engineering program. Mainly, I work with those in the healthcare field and those with an engineering background in their understanding of healthcare systems and how to develop more efficient systems. So, in my current role, I am basically developing from the ground up. I am in the role of strategically developing and integrating all of the different stakeholders and all of the

different inputs and outputs that are essential for research and development for the institute and that comes through data, it comes to quality outcomes, working with clinics, working with the hospital, the physicians. We have a multi-tiered stakeholders where we have internal and external people that are working within the institute. So, we have a hospital group, private entity essentially, it's multi-tiered, multi-level and I'm kind of in the middle putting all the pieces together to develop a research program that connects with the education program. So, this framework is essential and when I saw it, it impressed me because this is something that I've been trying to explain to a lot of my counterparts within the organization of how we need to integrate things. I'm very excited to participate.

Interviewer: thank you very much and what education do you have.

Interviewee: I'll start at the highest degree level and kind of back into so my eyes degree terminal degree is a PhD doctorate in public affairs with a concentration in health services management and research interdisciplinary degree working with multitude of different public administration in different aspects of policy etc. I have a master's in health services administration and a bachelor's in business administration. I also have project management professional PMP as well as a certified scaled naturalist so essay and I'm certified clinical research professional, it is a designation for those in research.

Interviewer: very impressive credential. Let's start with question number 1 "Overall, how well do you think the framework overcome the challenges in healthcare big data Analytics adoption?" Interviewee: my first impression when I saw it, this is what I've been looking for. it's very hard to map out a lot visualizing and trying to explain these things. It is a very important aspect and I think that this framework would be additional areas that we've add. It really illustrates what it is that we need to be doing in order to make big data relevant and timely and

impactful. So, I see that the interconnections here and the identification of each one of the stakeholders at the different levels and what role they play in how they can act as either an input or output. It is very essential so that was my first impression from looking at the framework. In the detail that goes into when you look at the details and each box what you really feel like oh wow that's what we need that that's the one is going to solve the issue. So, overall is one thing but you look at more details I think the breakdown of the components is the thing that drew my attention to this framework. In a lot of cases whether it's in the professional world experience, working in the field, or even on the academic side what ends up happening is people go into roles and wouldn't talk about big data. They won't have a breakdown such as this where it's saying where things are coming in, where things are going out, and how there's inter relationships between each one of these components. So, what ends up happening is you'll have a "quote and quote" big data person or a team but there's no really defined roles of where the ownership lies. So, it's very hard than to understand how you're supposed to develop and improve each one of the relevant areas. When they combine just well, this is what this big idea. So, it's not really clearly there's policies, procedures, standards of practice that can be developed for each one of these aspects as individual and then connecting them all together. So, I think that what I see here in the framework is how you break it down is, it's very easy to understand and these are separate tasks and separate domains that need to be addressed but they can all come they all need to come together to then produce the correct input because you know the famous thing is garbage in garbage out. You can have a lot of big data but if the data that's being produced is not quality, it's not coming in, it's not relevant. Then the data is not going to be useful to the organization as a whole.

Interviewer: to question number two “what does the framework provide to healthcare organizations that they are not currently considering when implementing big data analytics”

Interviewee: there is a realization that there are very important domains for teams that we have that need to be in place in an organization. The trouble comes with how these things connect with each other and the reason that takes place there's lot of reasons but sometimes even developing an organizational structure for having people that have the knowledge to understand how these things interconnect really becomes a challenge because what ends up happening is you're given a role on job description you get it from people that are above you in the organization and what ends up happening is you become very focused on making that with the tasks or objectives within that role of success which is just needed. You're hired for a reason however what ends up happening is people then start operating in a back here where instead of looking at what are the stakeholders, what are the relevant parts of this, how do we communicate with the teams that are providing data, or what is the data, what is the downstream impact on the data that we're generating. It ends up being we're going to create benchmarks and we're going to create dashboards and we're going to create queries based on requests and the teams don't necessarily understand what the impact of that in the downstream. Does that mean that someone is going to hire or fire or you're going to increase resources? So, if the idea here why is this important and going into how it provides benefits to the healthcare organization. Again, it starts illustrating that it's important for these teams to develop what it is their objectives are, but then it's also very important that these are communicate and not just be a hop down or bottoms up approach. Essentially, you know people just fumbling around. It really needs to be a very structured some of it done organically meaning that you know someone is interested in learning something deeper they have that freedom and flexibility to do that but some of it is very

purposeful that the teams that are responsible for each one of these domains need to be in communication with each other and there needs to be a general understanding. It doesn't need that a person, or that's nonclinical, needs to become a subject matter expert on the clinical side however there should be an under a general understanding of the downstream impacts and as well what the reasoning behind the initial request is as well. If there's a vision, if there's that 360 view and understanding, those things, then what ends up happening is the organization can actually grow and develop progressing forward. So, you're talking about a portfolio rather than project where each person is focused on their project, and it's not seen as these projects all moving the organization towards a certain direction.

Interviewer: so, what you are saying many organizations are not currently considering the data as a product which is included in their products portfolio.

Interviewee: it is included in the framework, data is a product it is not anymore, a project or some of the tasks somebody asking. So, if that's one of the things you consider to be a product that is adding value. Whether it's through project managing or scale agile mindset, whether that is a big program, big product, there still should be elements of that value in added incrementally over time because we're in a fast-paced environment. So, even when it comes to the way that project management is viewing things, they're heading more into an agile mindset when it comes to these very iterations. So, moving from that as a project purpose for this specific team into this is a product, we're going to utilized it as a product that can inform our decisions across the board, not just financial decisions of what our expenses versus our revenues are, all the way down streaming even connecting what those things are outcomes and even demonstrating with big data to try to almost you may not be able to directly in some cases but you can start inclining yourself towards what is the impact on quality health care, what is the impact on population health as a

whole, when it comes to these things and I think that that's the missing, that's part of the missing piece of the data which is you have separate sectors which is quality population health access seen as a completely different team that reports to different part of the structure, and then you have the financial sites looking at the operational performance and those are seen as two different dimensions. So, what ends up happening is both sides end up suffering because in the quality aspect when you're putting in resources to look at and measure these. These are all expenses; these are not generally value directly or revenue generate. So, they're seen as an expense so sometimes there is a hard time or when tough decisions need to be made those teams or downscale while on the other side you may have a growing and growing team of those that are looking at the operational aspects of the expenses be performance the time in surgery, but they're not necessarily tie to quality. These things need to be balanced and without big data really connecting besides that should be integrated together. You're operating in a vacuum, it could have downstream impacts that if you start cutting, you say well we're spending too much money here we don't seem to be getting performance, but we don't see what taking away one FTE means on quality then we're operating plan and we're not using fake data.

Interviewer: So, what you're really saying is the wholeness of data or completeness of data is what you are looking for, not just part of the data.

Interviewee: yes, because when I said in meetings, and I've sat in many administrative meetings for the clinical site and most of it has to do with revenues, how many patients are coming in our clinics and are coming out. Then research or analytics this kind of seen as a separate thing done by an outside team so you may have physicians do you make people on the ground to have a very good understanding of communications were seen what the revenues are what the coding is that our views the relative value units they have ana understanding of that but

they don't necessarily have a really great grasp whether it's those on the ground or those that are the administrators that are supposed should be tying that. If we're spending this much time in patients, how many, how much of that time is spent on education so that the patient doesn't need to come back in for a re-admission or they have a complication so when you start tying both of those together, yes is not something that we're trying to say. This is something that's needed and there's not a desire. This is the number one thing when I change organizations and when we start having these conversations of what are the needs, the number one thing from these positions that are being passed in from his former ground case, I want an ability to collect data in an effective way and then be able to go back and see what are the impacts of the decisions that I'm making by change my technique or if I increase time or we increase education or any of these things that I can then track what the impact of that is and then what the impact is on our bottom line. We're being inundated all day with what's the bottom line what these things are, but I don't have a good grasp of these other things. So, their number one thing, this is across the board, across specialties, whether surgical or nonsurgical and what ends up happening, I mean this is more of a thing which is these things that end up getting relegated to research so that research becomes umbrella for people to collect data which is not the way it should be because research is very compliant, it's a very highly regulated thing. So, research should not be the first answer for collecting data, it should be a user, it should use the product on big data, it shouldn't be the precursor to the collection of data and the honest truth is when you're talking about it, all of these requests coming in through a research perspective is very difficult to nick transform that data into big data because it is very difficult to have all of these data points collected from that perspective. If it's not integrated as part of the system so that's my experience, I find that people have the desire for it they don't know how to end up going back. I want to find some way to do it

and then you end up putting Band-Aids on the collection aspect and then very little impact because these things take years to collect.

Interviewer: good deal moving to question #3 “What step or function does need to be included to the framework to address any gap?” So, what you think is missing in the framework.

Interviewee: I think this is a challenging question. I think that the scope here is large as it's. So, it's not something that I'm expecting to be done at this point. This could be done as time goes on but I think that what quality outcome looks like and how these we almost start defining and putting some interconnectedness and using this framework to establish some kind of structure because it's great to have frameworks, it's great to have structures, but if there are ways that this gap here is how does this get translated into a structure and can be used for implementation because like what are the rules for the job roles, how does an organization adopt this into its practice would that be the gap. An operational model or having the implementation so that's a gap.

Interviewer: any other gap you think it is there or you feel like it just needs to be added or missing.

Interviewee: this is the thing I don't see necessarily that we have like research and develop like. I don't see our specific box for research and development for innovation, with the broader sense of what research and development means, practices like optimization. A lot of those key terms in regard to how do we integrate into the framework, people that are responsible for that aspect of almost looking at it from the strategic level, is that one of the C-Suite folks or is there a need for teams that are on the employee level to speak that. Are stop driving force to make sure that these elements are adding value and that there’s clarity and that there is progression moving forward. Again with a lot of these teams in the operational model sort to

speak if it is where is the continuous improvement piece or feedback loop that is that is that enable to continue to move these forward once these have been set in place like how does this population health or the quality healthcare at the outcomes, I think those were to the two things the operational model and that feedback continuous the PDCA type I think those were the two gaps that I could see. Interviewer: question #4 “In your opinion, what will be the most benefiting matter to healthcare organizations when the proposed framework structured approached is followed?”

Interviewee: so, I think that model based in the business side of things is that linking those structures to developing the processes that make this the reality and developing those outcomes again. There's a lot of time, there's a focus for part of it is missing where we have the siloed approach where it's not just the teams even the structure process outcome is essentially silo. So, you have some people in an organization they are measuring and responsible for the outcome. They are maybe the ones that are interacting with the customer, and they may be dealing with the file. And then you have a structured framework organizational structure dotted lines and then some of the job descriptions, then the process is linked, and we talked about the cycle of how these are supposed to loop back into each other. So, I think that part of this really where organizations can benefit from. If they are able to take these structures, they can benefit by making this dynamic, making it live. This structure is a starting point, and the needs of an organization may be slightly different, but I think that is like a strategic vision of how things can connect and if you're talking about the operational model and then you have those feedback loops that you can continuously inform and update and adjust that I think is going to be the most beneficial thing. Because again some organizations do these things separately very well and some of them don't do it very well. Be great at one aspect and they are able to provide the best

healthcare for example, but they get to that point there's so many inefficiencies and redundancies that processes are not working well. So, how do you make sure that the structure, the process and outcomes are all of high quality and there it's very well-known and then it still if you have a way to visualize it when you bring employees on or when you're in a team meeting or when you're talking to another team, you have that imprinted in the back of your mind of where your place is and I think a lot of people in the newer generations they want to understand, they want some self-actualization, they want to understand what their goal is what their impact is, and sometimes when a person is so project or task oriented, they don't have an understanding of what does there report me to the greater organization at how does that impact the patient, how does that impact the community. So, if there is an understanding across the board in the organization of the framework is how we operationalize it, how do we continue to improve and what the impact of that is, and what everybody who works in that organization what their role is to make that happen, then I think that you've now revolutionized how organization works and now when a person is working, they're not just a number and they don't feel as if they're just doing these meaningless tasks. They feel as if their role is very important and what they're doing is going to be changing the lives of people down the line of their community. And ultimately, it will come back to benefit them all. These are going to be with this framework, it makes it measurable, it's not just a feel good like I feel that it is making a difference. You can inspire that feeling by being able to measure it well, so I think that in this dynamic of these feelings and emotions of it and I think that they can inform each other and if you do that then that's called listing up, you lift up the people that are involved with stakeholders and you're lifting up the organization and the customers.

Interviewer: to conclude you have two minutes for any last word or anything you want to share thank you very much.

Interviewee: I would say when I can use it. You know this is like I said, this is great, and I think I'm excited and with your permission once this is ready for prime time. this is something that I would like to utilize because like I said I'm involved in literally building something from scratch and this I think is an amazing way that you can take this to C suite and you can also take this to align employee and that they can then understand where things are coming in and out, understand the operationalization like I said often operational model, so I'm really excited to kind to this published so that I would be very excited to have this presentation slide deck or something that that I can share it and I can show because like I've been talking about a lot of these things but it's very hard against visualize plan through. So, many aspects different parts of the organization and I really want to use it. It keeps some of these aspects more streamlined

Interviewer: thank you very much for your time I really appreciate it

Interviewee: thank you and take care

Interview 4

Interviewer: good afternoon. Thank you very much for joining me and giving me the time to interview as a subject matter expert. Hopefully you reviewed the instruction and the framework prior to our meeting. I will go over the instruction, then we'll go over the framework. I will ask you 4 questions, and you have 5 minutes per question to answer. The total time of answering the questions is going to be around 20 minutes. Then you will have two minutes for conclusion or for any thought you have. Now, let me go over the framework. It's 41 blocks, in different layers. Organization, people, process data, technology, and outcomes. It is based on people, process, technology model. Each box has a color and as you see the main thing is the big data which, is in red. Including big data allows us to do get data completeness, or the wholeness of data. And the other 40 blocks are needed for a successful implementation. I know you are very much involved with the technology and outcomes. So, I look forward to hearing your opinion about the framework. Now, I would like if you can introduce yourself, what is your education, your experience, your title, and what you do. And as a subject matter expert, how you're handling data in your environment at work.

Interviewee: my name is xxxx. I have a background in finance and a master's degree in healthcare administration. I have been in the healthcare setting for 10 years with different operational roles from revenue cycle management to population health in a large healthcare system in central Florida. I have been working in data analytics, data aggregation for assisting any operational outcomes through measurement for data and helping in terms of reporting and forecasting.

Interviewer: very good and what's your current title?

Interviewee: I'm a consultant with the value-based care team and consultant for analytics and I'm also a product analyst on the EMR team.

Interviewer: let's start with the first question “overall how well do you think the framework overcomes the challenges in healthcare big data analytics adoption”

Interviewer: the big thing that the framework really touches upon is the synchronization between the people and organizations. I mean by that is in IT, or any sort of data roll, lot of times it's very siloed. So, you do your operational work in IT Implementation, standing up product, getting things live, then you get end of that. Then we move off the next kind of initiative, project, go-live or whatever it may be. When we're talking data and healthcare, lot of times we see the disconnect is different teams not understanding the operational objective or organizational goal and how it ties back. So, if you were to understand up lot of times but we see the disconnect is different teams not understanding the operational objective or organizational goal and how it ties back in order to stand up some sort of product that's going to be giving some sort of data analytics if those reporting are essentially not meeting the needs of the business unit, then that product is almost not as useful. So, what I'm seeing if he's reading through this framework is the fact that the people understand how you need to have a communication, understanding what the strategy is, what the footprint is going to be, what the objective overall, and then how does that tie back then down to the processes. And then what happens in data infrastructure from data technologies beyond the support that ultimately leads to the outcome. So, the framework illustrates the fact that all those things are important. It's not just having infrastructure in place or the processes in place but having the right people that get through sentence each language and be able to communicate back to any sort of staple.

Interviewer: And why?

Interviewee: It's breaking the silos and it is very much connecting things and I think breaking down the silos and really becoming kind of uniform and synchronizing the workflow

and projective walls and outcomes. If you're an IT personal or big data, you understand the languages of data and the technical aspects thinks but you need to understand the business side of things right. So, if you're able to understand your customers need or your business needs and you're able to refine the product to be able to meet that and breaking those doors down. Breaking the silos allows that to happen in more efficient way.

Interviewer: Move to question number two "what does the framework provide to healthcare organizations that they are not currently considering when implementing big data analytics"

Interviewee: I think really where that comes down to is back to the first question a little bit just having a sort of streamline connection between organization to people and process but ultimately what I also appreciate in this framework is the fact that the outcomes are listed very clearly. I think often when we hear big data or hear big data analytics, most organizations kind of right now when they're talking about data or dealing with data that's retrospective. So, when I am getting data from a payer or whatnot commuting data that's likely to 3/4 old so the picture that's being told when it's being illustrated is one that asked about making decisions on things that happened in the past. For lower risk and try to forecast, what can happen based just kind of on historical data. So I think with this framework we start to move into the space where we can start to look at data in real-time, to start to do more predictive analytics, start to make decisions based off data in real-time and not necessarily just paint something a picture that you already know. For example, the number of admissions-based office claims that those type of stats, they're nice to have the general list early to drive decisions or they don't drive any sort of goal completion. I think really where the next gap in opportunity comes to in taking data and being able to start to predict outcomes, start to forecast what your needs are and with this framework in place that

only happens when you have a good synergy between organizational objectives, people and infrastructure such the data technology in place to do that.

Interviewer: question number 3 “what step or function does need to be included to the framework to address any gaps” talking about the gap now

Interviewee: I think it's going to come down to the processes, but it seems the processes are mostly around data governance, lifecycle management, compliance but I would think you would need to have I'm sort of flow between outcomes and in-process and people what I mean by that is training in terms of understanding what the outcome needs to be. Defining that right because from what we experience is a term like readmission could mean three different things to three different people, depending on the context of the settings. So, some sort of training can sort of standardization in which we're all speaking the same common language. I think that starts to really drive into being able to trust and rely on any sort of big data sources and then being able to have something that typically stable in terms of outcomes.

Interviewer: So are you talking about standardization of things, and everybody needs to be on the same page or are you talking about operational models, so people know how to use the stuff.

Interviewee: I think from an operational model standpoint, it's the accessibility, the understandability. Understanding the features of the tools that are available to the operational stakeholders to utilize that technology to utilize data and then to some sort of data standardization. It is something that falls into data governance where we are able to let's just say identify the top under data element that is most frequently used across the organization. Having that is I think vital moving forward because they get it but we're not speaking the same language or we have a different interpretation of what certain data elements are, what we stand for. Then it

needs to those areas of risk in terms of the final product not meeting the business objective and the right expectation and I think I gave that example about readmission. If you're talking about readmission to value-based care, readmission to be defined based of whatever contractor agreement you have with a commercial pair and readmission in the hospital setting that definition put by CMS for different times. Is it 90 days period or 30 days period? So, you know it can get very tricky and add sort of layered. Standardization allows you to kind of set the term off the back and allow us to kind of keep the same language from engaged row for the finding process all the way through the execution

Interviewer: question #4 “in your opinion what will be the most benefiting matter to health care organizations when the proposed frame structured approached is followed” because your daily work especially in the outcomes, I believe you will be able to talk about the benefit to healthcare organization in more details

Interviewee: yes so, I think the most beneficial aspect of this framework will help organizations is really going back to breaking down the silos. Right now, with this framework there is an expectation of culture that defines that allowed it. Users to connect and communicate with IT subject matter expert. if both entities are on the same page, and both entities are able to speak to the same goals and understand what the end objective is. Then the time spent on developing and building the product is more valuable from an IT perspective. You're able to foresee, forecast, adjust, and adapt to the business environment of things are changing. From the business perspective, understanding the capability, the limit, weather be from the infrastructure standpoint of what can be done and what is something that has to be kind of developed and test it out. With that being said, ultimately if those two pieces are in place, it allows for a broader understanding from and operational end user end gate standpoint. The environment is constantly

changing, what reporting definition that we have today need, it changes tomorrow from a national perspective or from a state level perspective. There are new regulations in place things that come down the line that we as an operator have to adapt to design process, and having the silos breaking down for sort of common understanding, allows you to be able to take changes and start to brainstorm, how to incorporate from it, from an infrastructure standpoint lot of the work again as an organization that we've done it. You're looking at what is the standard workflow, what is end of foundation, what it offered and where can we meet halfway. What you know, what can we do to customize a lot of its goals and project development aspect. So, this framework here, it is again, it starts to break away from the fact that IT belong to its own corporate corner, and you only come to IT when you know the Internet is not working or when there's a data outage or call centers is down. What this tells me in fact that IT is in the table with the decision makers. IT aspect of things have to be on the table to provide their input because ultimately IT serves the customer of the organization. When you hear about new objectives, new acquisitions, new technologies being brought for the IT team for implementation. So, having that relationship matters as posed by this framework. I think it allows for a better conversation and more easily achievable objective.

Interviewer: I have a question do you think that framework going to expedite or it's going to delay things. Will it save money, or it will cost money?

Interviewee: I think part of that is really defining what that return on investment is right. I think we're trying to implement infrastructure to be able to appropriate big data which could lead to predictive analytics, machine learning, whatever you know. The data topic is a hot topic today. It is going to require some sort of capital upfront, but I think from ROI perspective we're looking at operational efficiencies, reducing redundancies, optimizing workflows, being able to deploy

the latest and greatest technology and mechanic out front. It allows for those outcomes to be achieved faster. When you think about it, if you have AI learning you will be able to kind of natural language processing where you could read dictation and be able to code right away stop that says a lot of time in terms of getting things coded getting billed out the door which means you know from accounts receivable AR upper things kind of drive up and down but definitely you know there's always the cost up front but when managed correctly, when implemented correctly, and put the right expectations that return on investment allow for that to be made up and when we're pushing down digital front where it's becoming more about having a digital presence being able to reach out to a wide variety of stakeholders and even a younger generation in which you know what you would think a traditional set of health care is not necessarily valued in the young generation. It is this type of framework that is needed to be able to drive to the future for Tele health to be able support geo location and being able to reach out to consumers and guide them to steer onto the right places over there.

Interviewer: you have like 2 minutes to give me your final thought conclude anything you want to add that we did not cover.

Interviewee: sure, I think the only thing I would add is I think the important for organizations trying to breakdown the idea that IT technology, clinical engineering living in their own world because again any sort of future footprint that you're trying to have as an organization when you're trying to capture you consumers is going to require the IT innovation side. So, the quicker the organization's culture can change in terms of how it being technology and who handles technology, I think you'll see a lot quicker adoption and better result in more official information I think the framework can help. what I read off is the first thing that be is organization layer and then I see people layer. So, to me that tells me we are talking about IT

people whether the IT leaders there's direct connections right off the bat, to business objective, data strategy, engagement. So, from top down, leaders will understand what where their business is going, they will be able to adjust their implementation standards and then drive it down through processes through management infrastructure all the way downtown.

Interviewer: thank you very much for your time and I know you have really a busy schedule and thanks for contributing to the healthcare industry.

Interviewee: yes, thank you I appreciate the opportunity

Interviewer: thank you and take care

Interview 5

Interviewer: good morning thanks you very much for joining and I really appreciate it. This is to interview you as a subject matter expert in data big data analytics to evaluate the framework. I will give you the instruction as an instruction and I will use MS word dictation and I will be recording the session. There are four questions, and you will have 5 minutes to answer each question. I will go over the framework and it based on people, process, technology model. It was created in 1962 and I added organization, data, and outcomes. The framework has 41 blocks and the main one is a big data which is a lot of healthcare organizations are missing for data completeness. You will have 3 minutes introduce yourself, your education, background, years of experience. then conclusion for two minutes yourself. You can start now by introducing yourself.

Interviewee: my name is XXXX. I'm a pharmacist by training completed my doctorate in pharmacy 1997 and I also completed a clinical residency. From there I practiced pharmacy primarily in the pediatric intensive care setting for five years and I was in a role clinical coordinator of the pharmacy department. When we first moved from paper onto computerized system, I became involved as an informatics pharmacist, and I always had a quality and clinical outcome as a real area of focus on my career. My interest and over the course of my career in informatics, I became more and more involved with the corporate quality department and clinician workflows and clinical decision support, computerized clinical decision support, and that sort of led me to my current role currently as senior director for IT performance and quality goal solutions. Which is a lot of words but basically, I form the bridge between our corporate quality department and our strategy towards clinical outcomes improvement for becoming a highly reliable organization from a clinical outcome standpoint and connecting that over to the leadership team so that we make sure that the organization is using the tools we provide to their

maximum potential. Then at the same time as we create road maps and strategies. We do so with the mind for the corporation strategic goals and our regulatory obligations.

Interviewer: You are a very good for the interview to get your feedback. Starting with question number one “Overall, how well do you think the framework overcome the challenges in healthcare big data Analytics adoption?” so the framework has 41 blocks only one blocks is big data, but you would need 40 other blocks or components for successful implementation

Interviewee: I think the framework certainly expresses the complexity of the situation very well. We have some elements of this in place, but I'd say as an organization, we've got a long way to go as far as maturity particularly around data quality, data integration, ingestion. Metadata, cataloguing. That whole data row I think we're weak on and then the processes as well around governance and data lifecycle management are also immature for us. I think we recognize the need for, we recognize that as a need organizationally and about a year ago when we brought a new chief data officer that was a major step forward, but he certainly has had his work cut out for him and in lifting us up to the next level.

Interviewer: so why do you think that the framework will overcome this kind of challenge.

Interviewee: I think that when I interact with quality folks specifically with industrial engineers, they have very low confidence in our data quality and it's primarily because we consistently failed to demonstrate back to them that we understand the data we're sitting on we particularly with the EMR, we often as an IT organization are unable to explain what specific user actions or workflows are producing a given piece of data and so they can't that renders them, unable to go in and validate the data you know against observation of the correct workflow. So, I

think the framework is valid, but I think we've probably like I said at least for us from a maturity standpoint, I'd say we've only affectively implemented maybe half of those blocks or so.

Interviewer: I see. so let me go to question number 2 “What does the framework provide to healthcare organizations that they are NOT currently considering when implementing big data analytics?” I think you mentioned you implemented only half of the blocks, so what the other half you are not considering.

Interviewee: I think health care organizations fail to understand the scale and the complexity of the data that they preside over. The EHR's are very transactional systems, and I would say we're pretty far behind other industries in turning our data into actionable knowledge. So, I think for us, I think healthcare had to recognize that they are not experts in big data and so healthcare administrators and physicians and other clinical leaders need to come to grips with the fact that they need to look outside of healthcare fields. It feels like engineering and data science and bring those experts in this, we can teach them the language of healthcare a lot easier than we can teach ourselves how to manipulate big data effectively.

Interviewer: so, will go question number 3 “What step or function does need to be included to the framework to address any gap?” so now we are talking about the gap, what the framework is missing.

Interviewee: I think I'm trying to articulate; I'm not coming with the right words, but I think that we need to one of the outcomes we need to drive this so educating our hospital leadership on how to be good consumers of big data. you know how to incorporate the right processes and tools into their decision-making process so that they are getting the most out of the information at hand and we have a lot of information at this moment, we have a lot of self-service analytical power at the fingertips of our leadership team and they're still looking to

middle managers within each silo to go in and manually pull information and send them a snapshot once a week. If they were agile consumers of the analytical tools, and we've got in front of them, they would be going into places like the executive dashboards and fueling their real time decision making with the real time analytics tools rather than continuing this sort of send a middleman over to collate a snapshot and bring it back. I think it goes back to trusting the data still we've got to grow trust in the process and trust in our data analytics capabilities among the senior leadership and then those senior leaders need to embrace the change and embrace the tools we got.

Interviewer: so, let me ask about physicians, do they still use reports. Also, are you saying senior leadership are not trained and they just relay on others to give them the reports.

Interviewee: OK, physicians who do clinicians use data are not uniform group, there's a wide variety of skill sets some of them are exceptionally data driven and you know in research oriented, some of them are naturally very process oriented and they're good at systems improvement but I think it's highly variable and in most cases it's self-taught by those individuals. They come from diverse backgrounds, they maybe have an engineering mind, or they have a population health or neuroscience background and they're very comfortable in the data but for the majority of them, I think they're self-taught so that leads to a lot of variety in approaches and a lot of variety and skill level.

Interviewer: are you saying there is no users training on basic usage of analytics tools and they cannot get things in real-time, and they rather wait for reports than going directly to dashboard. Interviewee: yeah, I think that’s covered in your in under employee engagement. I would maybe say employee slash leader engagement. When I use the term leader in healthcare,

I'm lumping in sort of the operational leadership along with the clinical leadership in the physician's community, I think those are that's a big bucket of leadership.

Interviewer: move to question #4 "In your opinion, what will be the most benefiting matter to healthcare organizations when the proposed framework structured approached is followed?".

Interviewee: for me there it all comes back to patient outcomes because those drive our financial success as well as just our core mission. So, the more that we can identify and standardize best practices and achieve consistently high reliable patient outcomes, all that hinge on our ability to function in this data space. If we want to target a very specific population with intervention by limited resources, we first have to identify that population if we want to find patients at risk of readmission or patients at risk of mortality within 30 days of discharge, we first have to find the algorithms and the documentation and to identify those smaller groups so that we can bring to bear highly specialized resources to intervene. So, to me the big benefit will be to healthcare's bottom line which is patient outcomes.

Interviewer: So, let me ask you will the framework helps generating reports that CMS require? I know you involve a lot with quality report sent to CMS.

Interviewee: that helps streamline and makes things faster and instead of entering the report manually from an operational efficiency standpoint it will certainly help where we are right now in healthcare is we're very rich in outcome metrics and we're very poor in process metrics. so I can tell you with a fair degree of reliability who developed a urinary catheter infection or who was re-admitted within 30 days of discharge but those are the outcomes where I'm trying to build our organization now is to get to those process metrics. We've counted the number of people with infections but where along the line was there a point where we could have

intervened and prevented that infection, what are the processes that we need to track so you'll have some knowledge but there are no action items behind that to tackle the issue so you have the insight but what needs to be done to deal with what you learn

Interviewer: I know it's 10:56 AM and I know you need to leave at 11:00 AM. We have 4 minutes so any final thoughts or any feedback that you want to have or share about the framework.

Interviewee: I think within that framework those are roles and in large part specially that the people layer, the process layer. I think in many cases we've tried to fill those roles and develop those processes with folks that were doing it as an add on to their day-to-day job and as we've grown as an organization in size and maturity, we realized that we need in specific places like the chief data officer, like chief medical informatics officer and even probably in the lifecycle management and the data governance realms. we need dedicated folks with expertise in the areas. We can't just have job description under other duties as assigned. This is so challenging and expensive. The good news is all those are included in the framework.

Interviewer: thank you very much, I really appreciate sharing your knowledge and insight.

Interviewee: thank you and any anytime

Interviewer: alright thank you and it take care. Have a nice day.

Interview 6

Interviewer: good afternoon, thank you very much for joining. First, I would like to take your permission to record the session if you don't mind. I will also be using Microsoft Word dictate while we are having the interview. I will go over the instruction. it's four questions, you have 5 minutes to answer each question. Then you will have two minutes 2 for conclusion. it's a semi-structured interview, so I may need to ask a different question. I have 3 minutes to go over the framework and then you have 3 minutes to introduce yourself, your title, your years of experience, your education.

The framework has 41 blocks, each block represents specific components or area needed and it is based on people, process, technology model and this model it was found in almost 1962. I added 3 other layers, the organization, data, and outcomes. Each color represents something, but I color coded. One important component in the whole framework is that big data and it's under technology layer and it is in red. The data components are in orange color, then we have brown color which is related to people and process and then the outcomes are in green. Hopefully you went over the framework prior of our meeting. You have 3 minutes to introduce yourself, your experience, background

Interviewee: thank you, my name is XXXXX. I am I have a BS degree in computer science, I also have two master's degree in business administration and information technology management. I have expertise in database administration and database development, infrastructure, and enterprise architect. I am certified in Microsoft reporting platform as a developer. I have an experience of 20 year plus in technology jobs. My expertise fits into data platforms. and I am the technical lead for the EMR product. I also the supervisor of the database team and also as a database administrator/developer. I have 20 years of experience in healthcare

Interviewer: thank you very much for introducing yourself. Starting with question number 1 “Overall, how well do you think the framework overcome the challenges in healthcare big data Analytics adoption?”

Interviewee: The framework covers the most important aspect of building big data analytics that will fit the needs of any enterprise, not just healthcare. This one is really going to work for healthcare specific. I could from observing the framework, it does have the main component in my opinion, one of them is security and privacy data collection this is important to be in place. It also addresses the integrity of the data which, is very important there is a saying about data quality and data integrity, trash is, trash out. So, if the data is not accurate and data integrity is not there, it basically does not help making any informed decision or it doesn't provide any good information. There is another one that is outlining most technologies where you can acquire data from Internet of Things, cloud, computing, storage, network. I mean all that stuff provide a platform where you need to get the data and how and what technologies do you use to bring that implementation. Then, there is the expertise and that's where the people come into place from technical and from management all the way to the to the top. so maybe I want to specify the challenges in healthcare when they deal with data. The challenge is coming from the healthcare data, it is all over the places, healthcare data is not in one location, health data difference from one area to another. That is the challenge working with data where security and privacy add to the challenge. So, the challenge is really how to deal with data and what kind of processes you should have. As an example, one of the biggest challenges in healthcare that is everywhere, and the framework tried to add with that or provide a solution.

Interviewer: move to question number 2 “What does the framework provide to healthcare organizations that they are NOT currently considering when implementing big data analytics?”

Interviewee: The way I'm looking at this framework, for me is 99% comprehensive the reason I'm saying that from my expertise, it has all the components that are needed so talking about this framework that you're presenting it does have all the components needed. So, from that perspective anybody doesn't have it and healthcare specifically don't have that integration and do not have a consistent way. Put it this way each one may be using different EHR systems, patient records not really the same across all platforms, so they might be missing the integrity, so they will think that their data is more accurate data. As an example, we need to kind of see if that's true or not the technology issues are observation, or some healthcare do not have the technology or the ability to do what they need to do so that would be also missing from talking to some other healthcare organization or you know from just interacting with them. That's what I can elaborate on that question.

Interviewer: So just to clarify do you think they only focusing technology and data?

Interviewee: they don't focus on the other stuff like processes, and it seems like everybody only think about their area and instead of looking at the big picture

Interviewer: and you think that the framework will get this big picture, so people understand that it is not just that a technology need to take care of the data, it's more than that.

Interviewee: that is correct and in that framework list all of them the vision the teams and technologies and how. and of course, the vision is going to be really well outlining group in the framework. As the outcomes, what is the result, what is the added value.

Interviewer: so, we talked about the benefit, now we're going to talk about the gap in question number 3 “What step or function does need to be included to the framework to address any gap?”

Interviewee: for me it's unknown but looking at where the world is trending and where the world is going, I will say innovation is a big one. To be good enough and in the framework, I will put it right behind or before the outcome may be between the technology. And as an observation from my end because sometimes you need the innovation to get to that next level. The technology will provide it, the innovation will make it happen, and then the outcome will happen because of that. So, I will say that maybe something missing in the framework is it a layer or a component. For me is a little bit then innovation is not a technology. I think is going to be a layer if I want to put it into perspective it should be between technology and outcome because somebody have to basically make it happen.

Interviewer: So, what you are saying is, there should be an innovation layer to connect technology that will be used to the outcomes and there's some kind of innovation need to involve.

Interviewee: yeah, the technology will be the way we can look at collecting the data. And everything else according to the outcome that's basically say what the next step is that's the execution layer. So, we possibly don't have what we need so that's where innovation is going to come in. We need to create this anyway this to allow us to get the desired outcome. that's what I mean by innovation involved with new capabilities.

Interviewer: anything else

Interviewee: I just want to mention which is more of an artificial intelligent and machine learning. That is part of technology but that could be used to accomplish the outcome that's but it is kind of like technology and innovation. where will it fall, like a dotted line in the technology. It is not considered a technology; I consider it as a data

Interviewer: with you experience in data analytics, what you see the gap or what is missing in this area to get data the wholeness of the data or data completeness.

Interviewee: the bullet point is basically this is a high level. it's really nothing is complete in this world, but I will say it's complete for the most part. The source of data is going to be really the major thing but not defining what the sources are that's going to be something. There could be gaps and that's really all depends on the source itself. So, if there is any gap, I want to outline in here is really nothing.

Interviewer: question number 4 “In your opinion, what will be the most benefiting matter to healthcare organizations when the proposed framework structured approached is followed?” so about what you see the value of the framework.

Interviewee: if you can look at the data and you have something in mind and when you start looking at the data and what is in it you will get 500 different ahah because of just looking at numbers or names codes for whatever the case might be so the benefit is at least you will get a very large set of data that can let you think differently let you think about different things of the benefit. One of the benefits and I can name few here is allow patient using application on mobile device that allows to collect data from all your visits and everything else but that data could be used to tailor some kind of plan for the patient. So if you really think about it that big data we have is basically can provide a lot of different points because we could look at and all information about this patient and saying, these are the habits you have, these are the lifestyle you have, this is what you always eat, this is according to what the patient providing for the providers or the hospital that data could be out voted to me apparently quickly done to provide some feedback to the patient to better their life as an example. I think about the Apple Watch, smart phones that collect a lot of information your heartbeat, collect how many steps you took,

how many levels did you climb, it has a lot of data if we can tap into that information that data could be provided in a more organized fashion for the patient to allow to take care of themselves then going to a provider for example given that image of what am I doing wrong or what am I doing right so that's what big data can provide to healthcare provider healthcare which increase quality and quality of care. Another one is providing a correct timely treatment an appropriate treatment it also allows the collaboration between caregivers when you have one coherent system big data basically all the information about one patient would be one place so there is no duplication or redundancy of treatment because you basically follow and what everybody else is doing and you don't have waste time on complicating some others so just few things that I can mention about it. It just gives that much value not mentioning the basic stuff not wasting time not, eliminating a lot of unnecessary things, making patient happy quality.

Interviewer: let me follow up with the question do you think that the framework helps to expedite things, make things faster using? so there is a benefit for using the framework because it going to cost money to implement.

Interviewee: to make money, you have to spend money. So, money is always going to be an object. Technology is the most consuming from a cost perspective but from a return on investment, it actually can generate more revenue. Patients will establish that loyalty because I'm providing them exceptional healthcare. Just the simplest thing, the end result you have to look at the investment and basically say if I spend that much money my return of investment is going to be quiet and some of these return investment does not have to compute in a money. value like I mentioned, loyalty for example, because that can bring other people and that word of mouth or whatever you want to call it but yes definitely it is going to cost money, but you have some company spend money to generate new market share.

Interviewer: So, do you think the framework will help the organization to grow.

Interviewee: yes financially, market share exactly because that data is going to give them the forecast for the 10 years coming.

Interviewer: last thing, you have like a few minutes for conclusion or to bring we haven't discussed.

Interviewee: I think we discovered a high-level framework, but I can add that the focus here in health care person but different it could be used anywhere else might. In conclusion, data is the big thing right now that is what's going to make businesses make it or break it, without that data and statistics you cannot survive.

Interviewer: thank you very much for your time and I really appreciate it.

Interviewee: thank you

Interviewer: take care

Interview 7

Interviewer: good morning, thank you very much for joining the meeting and I really appreciate that you allow me to interview you to validate the framework. Hopefully you had the opportunity to look at the instruction and the framework prior to our meeting. I will take 2 minutes for the instructions, 3 minutes to go over the framework. You have 3 minutes to introduce yourself, your background, education, professional experience, title, and years of experience. You have 20 minutes to answer four questions, 5 minute to answer each question and then you have two minutes for conclusion. I am using Microsoft TEAMS for this meeting, and I will be recording the session. I am also using MS word dictation to allow transcript. The framework it is 41 blocks in six different layers organization, people, process, data, technology, and outcomes. It is based on people, process, and technology model which I added another three layers because data is an important layer and big data is one block under technology in red color. Please introduce yourself.

Interviewee: my name is XXXXX. I graduated from University of Central Florida with a degree in management information systems as well as computer science minor. I also got a masters in 2018 in project management in information systems from Florida Institute of Technology.

I've worked at various places. I've worked at Oracle doing support for their database product. I've also worked at a healthcare organization in Orlando managing their implementation of their EMR system for five years and I was managing their Oracle database at the time. I've also worked for a small startup for plastic surgeons just managing small practices and small medical offices. My current job now is with a large healthcare organization in Orlando as a senior data analyst for the last 3 1/2 years. My total experience is about 26 years in general and 10 years in healthcare

Interviewer: let's start with question #1 “Overall, how well do you think the framework overcome the challenges in healthcare big data Analytics adoption?”

Interviewee: so, healthcare we have a lot of challenges. In review of the framework that shows the 41 blocks, but I see personally with big data in the implementation that you proposed here this looks very good, but I think some of the challenges that we're going to have in big data. You have structured data and the majority of the data that comes out of big data is going to be unstructured. So, one of the biggest problems I think we have here is identifying when a system is implemented, we have a start service and then the end of service. When that data commences and when it is decommissioned. And the source of the data when we stopped putting data into that particular system having noted that I think the only concern that I would have is I had technology outcomes that there might be some boxes there which probably have to be drilled down because there's other players that come that maybe like the people that were providing healthcare.

Interviewer: I think that could be discussed in the gap of the framework. To clarify, do you think that framework overcome some challenges of the health care.

Interviewee: absolutely yeah.

Interviewer: what then are the challenges of healthcare with big data and how you see the framework is going to address these challenges or fixing them?

Interviewee: I think the framework is designed just fine. I think the challenge is going to come in with how that the unstructured data as well as the structured data. how we integrate that into the big data. So, the challenge of the healthcare today is exactly with data. The challenge of that is the unstructured data and the disparate data that we're capturing.

Interviewer: and do you think the framework will allow or will help correcting or addressing this challenge?

Interviewee: I do believe that the challenge with data is going to be more technology source of play in and it will be irrelevant to the framework. But I think it's still valid and will help absolutely yeah.

Interviewer: there is a follow up question, “why you think the framework will help healthcare organization utilizing their data in a better way”

Interviewee: because what I see here is it shows the dependencies and the relationships that are between the six layers as well as how they come into play with all parts of those entities. The organization, the people, the processes, the data, the technology, and the outcomes I think what will show relevant and what is dependent on, versus what’s not related but still dependent upon a particular layer. I think one of the problems that we have in our organization data ownership but is that a governance? So, data governance will allow identifying who owns the data. So, there should be that ownership should remain and it should not defeat at any time. if you know what I mean, or it needs to be handed over to somebody else before another person releases over languishes their ownership of that data. If you understand where I'm going up. Yes, business continuity processes maybe that's something that we're missing in the framework.

Interviewer: OK so let's have that in the gap question. I want to go back to the challenge, and I want to be clear that you are talking about the healthcare challenge with big data analytics.

Interviewee: we have a lot, from what I've experienced, and I've read that 80% of healthcare data is unstructured so that is a big challenge. yes absolutely, trying to capture that data and making it useful would be beneficial to HCO. I almost feel like in our setting right now, we have so much data in our warehouse that I'm managing but I don't think we're leveraging it

and using it like we should. there are many ways that we can leverage it further. The challenge is wasting of data and there's no value added. it's not covered and the third-party tools as well as the third-party vendor or vendors or then we're not getting our value out of it.

Interviewer: question number 2 “What does the framework provide to healthcare organizations that they are NOT currently considering when implementing big data analytics? So, you may look at what is your health care organization is not considering from all of these 41 blocks.

Interviewee: I like what I see here in this framework. I think everything is valid and it will be ideal in the HCO setting. I think for the most part, we do data product delivery manager, other stakeholders. Product and data management life cycle maybe we need. Product management lifecycle, I think maybe that is an area that could be improved upon in the product delivery manager lifecycle. Right now, what I see is like we don’t. We just did sunset a source of data from a system and having noted that there is never a clear understanding when the data will be sunset. There should be a product manager or a team that says OK this is our cut off the data that we have here. Will be to the effect like where will it be transferred to? will it be transferred to a new system? and how much of it will be transferred? I saw the notes that we had but when we do actual analytics you know when we grab that data into the data lake or the warehouse we can identify where it's sourced from, but we don't really necessarily know when that stop occurred. And for which facility edit occurred.

Interviewer, So, this one of the things that is missing.

Interviewee: yes, it's nobody owns it. I think we do own it, but we can say that for one facility they've had the data that exists in the warehouse will only be valid up until certain point when we stopped that particular product or that EMR tool for example.

Interviewer: do you do like change management for data product to inform and train users and others when there are changing in the system or when they are adding new

Interviewee: I mean it's all dependent on what we grabbed from live systems. alternately we need to do a better job of saying when they have that data valid for and when that changed over. If you will end the warehouse. The good example would be like when we do that in big data when we grab all that data, you are adding new capability and new insight.

Interviewer: but that may change your policy or procedure or processes and when you change the processes do you have change management allow users to take advantage or train with the new way of doing business based on the new capabilities.

Interviewee: I'll give you really a good example really quick. we had an easy explorer product that was provided to us that we worked on with a data warehouse vendor that was being leveraged heavily amongst all the facilities but once we went to a new EMR. Some of those facilities no longer work capturing data into that visualization dashboard. So, and people were at that point a little frustrated because we had some facilities that were working that still had access to their data but the facilities that that went to the new system first are no longer have that ability to look or see and there was never any information of what's the new way or what's the new format of how to get that data they were looking at previously.

Interviewer: that's a change management so you are changing things, but you are not making users aware of the change and train them.

Interviewee: yes, correct and you are absolutely right.

Interviewer: So, what other think you think it is missed in your organization

Interviewee: the framework is good, and I'll let you know one other thing

Interviewer: let's talk about the gap question #3 “what step or function does need to be included to the framework to address and gap?” So, if you can you cover the business continuity planning for example.

Interviewee: let's say we just made a discussion about how if anything would happen to the EMR and if the systems went down. what is the business continuity planning? how do we continue operation with an EMR that's down from the data side? Would be our business or clinical analysts that review the dashboards if they no longer have data for a facility. where can we go and capture that data, so you don't lose operational insights which are critical for a lot of these facilities. Without us getting involved and making sure that they continue to have their data that they need that's relevant for their jobs and reporting.

Interviewer: I want to be sure I'm capturing your thought in the gap and business continuity planning. But is it part of the operational model? Do you have a downtime system?

Interviewee: correct it could either be operational. I mean we can apply to both but ultimately, we don't want somebody to say OK we have it before and now we no longer have it on the next day because of a transition to a new system. We need to be prepared and tell the users how to get that data on the new system before they actually go to the new system.

Interviewer: I see that's a change management because change management will address all of that who are the users and how they will be affected and what the new way of doing business. Interviewee: BCP could fall under change management. It could be a sub block. I think ultimately that's going to help us to be more resilient in our working environment.

Interviewer: what else you think is it is a gap

Interviewee: I see artificial intelligence and machine learning. I don't see anywhere you put in here NLP. It's important technology to have especially for unstructured. I think it's just

falls under probably big data that's unstructured. EDW team works on quality healthcare. So, it's just included. You have financial gain and personalized care, a lot of the data that comes out of the warehouse the warehouse that I support is going to be for clinical emphasis and clinical quality. So that would fall under the quality healthcare but then looking at your outcomes financial gain, quality healthcare, personalized care, operational efficiency, population health management, employee engagement, and HCO recognitions. I think there's one item that might be missing. It would be reporting to external agencies or government. You could add external agencies reporting.

Interviewer: let's talk about question number 4 “In your opinion, what will be the most benefiting matter to healthcare organizations when the proposed framework structured approached is followed?” I think saving money. yes, if they adopt that framework what benefit the HCO. I personally think if this framework is adopted that not only would be saving money, but we would be providing a better personalized care. We leverage the data that comes from IoT’s and we're going to start doing predictive analytics before people get ill. We would be identifying those who the better payers and who is not. So better payers are for insurance who we can leverage to do business with, who we can avoid doing business with all that is going to encompass what this framework with this platform will benefit to HCO. We're going to reduce costs clearly. We're going to find better ways to provide a better patient experience if you will. What you want to analyze and what we can do to take that data and act on it.

Interviewer: very good so we have 2 minutes to conclude any final thought or anything you want to add.

Interviewee: I think this is a great framework. I like what I see and how this could be beneficial but once it's implemented, I'm sure that we will start seeing where things might have

to change, or you might have to improve an area or maybe add another block or maybe reduce the block.

I think we have some boxes of the framework in our environment. We have metadata and cataloging that's part of the EDW platform that we use right now. Data quality I don't think we have good data quality; we're improving data cleansing. if we implement this framework, I will definitely see a big change in how we operate as a healthcare organization absolutely.

Interviewer: that's great information and thank you very much for your time and I really appreciate it. I want to respect your time and we are at 30 minutes. Any questions before we end.
Interviewee: sure, absolutely when can we expect this framework to be adopted. It looks good honestly, I think you've done a wonderful job and depicting all the entities and how they're related. It's a great framework to follow

Interviewer: thank you very much and I really appreciate your time

Interviewee: thank you for allowing me to give you my thoughts thank you

Interview 8

Interviewer: good morning. thank you very much for joining and I really appreciate it. the agenda first for our 30 minutes meeting is, two minutes is for the instructions, three to go over the framework, then you have 3 minutes to introduce yourself, your background, professional experience, title, and years of experience. There are four questions, and you have 5 minutes to answer each question. Then you have two minutes to conclude. I will have Microsoft dictate and I'm also using Microsoft TEAMS for this interview. The session is recorded.

Interviewee: OK

Interviewer: Let me go over the framework, hopefully you reviewed it prior to our meeting here. The framework consists of 41 blocks, big data which is the main block of the framework, it is one block, but it requires another 40 other blocks in different layers to make the implementation successful. The framework is based on people, process, and technology model which, was created in 1962. Three other layers are added: organization, data, and outcome to the model.

Now you can introduce yourself so please.

Interviewee: My name is XXXXX. I am an enterprise account manager for solutions integrator based in Irvine CA and I am responsible for business relationship and go to market throughout the state of Florida for our large enterprise accounts. From a historical perspective technically, I did not attend college, but I have been in the technology space for the last 24 years and focused on data and AI and ML pipelines, working with several healthcare organizations to build their AI/ML pipeline.

Interviewer: alright so let's go with the first question “Overall, how well do you think the framework overcome the challenges in

healthcare big data Analytics adoption?”

Interviewee: I think that the overall components that the base components are all here. I think that the challenges in healthcare tend to be massive and complex. I think one of the biggest challenges is the age of systems and some of the disparity between what those systems do. I think that regulatory requirements that continue to change audit tend to cause a lot of challenges but as well as just the transformation of update and what does data mean and where does it come from and in what format and alternately where does it need to go in and in what format to be consumed appropriately. So, I kind of seeing that the is the biggest challenge in healthcare when it comes to data is really that data wrangling, data engineering aspect which is represented here mostly in the orange color and with a little bit in the blue. I think from a people perspective, this looks really good, but I do have some comments around things that I see. Maybe are not missing that might help

Interviewer: that's going to be under the gap question, so overall you think the framework address the challenges and provide a solution.

Interviewee: yeah, I think the main components are here and it depends on the use cases

Interviewer: So, you're saying the main components are there

Interviewee: well, I think I think depends on what we're trying to achieve. I think it's very difficult to get businesspeople and particularly marketing folks to get excited about data right and to really shift the culture of the organization to a data driven culture. So, let's make some of the things that some insert connections are kind of where things feed and where things interact. I think what we've seen with customers that have been built is the positive feedback loop of like where things start versus where they were finished and so the one thing, I'm not seeing here is how do these connections feedback on themselves. For instance, the chief data officer or the

other stakeholders or specifically data strategy are getting the input back from the efforts so that they can move and adjust it as part of a cycle.

Interviewer: So, what you're saying there are some missing arrows or something like that. We can cover it under the gap question. For this question, I want to know if the framework overall addresses the challenges.

Interviewee: So, I think the positive things are around the governance, the lifecycle management and the data privacy that kind of feed into some of the data engineering components like cleansing, quality, cataloging. Those types of things that stuff is from a flow perspective looks great. That's the part that I see the most value there. And from a red blocks perspective feeding down into those visualization reports, I feel like that's kind of a two-way arrows but those would be the stuff in orange is the meat of data engineering and data science. That's where success comes from. So, this is the most positive thing for me.

Interviewer: question number two "What does the framework provide to healthcare organizations that they are NOT currently considering when implementing big data analytics?"

Interviewee: I would say I think that healthcare tends to brush over IoT a little bit. I think that there's a lot of devices in healthcare organizations that can be monitored and then there's a lot of data that could spin off those. That helps the patient, and it helps the charting, it helps the IT, it helps the delivery of care. But I don't think a lot of that is captured to be useful. How to get even better care. So, I think that's a big challenge for healthcare, there's a lot of data out, there lot of devices out there. There's going to be more of them and better connected all the time. Particularly, inside humans so I would say that that's probably one thing because it's so challenging and it gets ignored. Healthcare organizations that want to consider data as an asset but actually getting to the point where it's like wow that data has value. I know exactly how

much that data is worth, and I know how to extract the dollar value. When you go to the CFO and those types of things the CFO looks at that like an MRI. From a data perspective, where its value. Because I know how much I charge every time the somebody goes under the scanner, but I would say that no I don't think most healthcare companies view data as an asset in that fashion from like a capital perspective. But when I say data is a product, then it's a product has a lifecycle you need to manage it, and it has to have an owner. One of the things they are not adapting, and the framework can provide value to them based on that would be the number one is the value of data. The framework can provide is shifting the culture of the people one of those three-legged part where their culture is data focused and they understand t that data is a capital asset. Fusion team and the other stakeholders, I think that those two blocks are kind of the same way, or they should be, the reason why I like, it where it is successful, we call like a data analytics center of excellence. So, kind of call it whatever you want whether it's AI/ML User or manipulate that data for reports or business decisions is important, but I feel like that center of excellence is folks from all three colors of blocks, not just people in the red block. I would say there's a lot of consolidation in the healthcare industry, we're going to continue to see that I mean certainly that's a business objective of large providers is acquiring other clinics or practices or building new hospitals or merging with someone else. And you have a brand-new issue when it comes to data and as you're starting to build out your data pipeline in your data engineering or just give an example, add like brand a right made in their data governance. They may say that red now. They may pick up brand be doing acquisition and that brand may from color perspective their definition of bread is actually Crimson bright and that's a big deal when it particularly comes to marketing. And in those types of things and so just the disparity of like OK, how can I scale while still keeping my data coherent as I go through acquisitions, but I think that when it

comes to data that's something that the healthcare does not pay too much attention to consolidation brands, they have lots and lots of brands, so something that they've become fairly adept.

Interviewer: so, let's move to question #3 “What step or function does need to be included to the framework to address any gap?”

Interviewee: what is the gap in this framework is audit. It needs to be center stage when it comes to data strategy. One of the challenges that we've seen other organizations come up across in particularly in healthcare as they start to build out their pipelines and build out their data and kind of get their engineering done and hopefully it ends up with a cloud data warehouse or at least it's centralized when they can actually do some transformations. The regulatory guys are like that's not going to work and so one of the things that we've seen the earlier that you get regulatory into the data. Word regulatory is kind of feeding data strategy that can accelerate your journey.

Interviewer: what do you mean a regulatory? Do you mean government agencies?

Interviewee: like for healthcare this would be like HIPAA and some of those types of things but also don't bet on it right where General Data Protection Regulation GDPR would be another one and then you also have the finance side. I mean when you deal with people's credit cards and that that personal identifiable information even outside of HIPAA, healthcare organizations kind of get hit by every regulation out of Washington more than other industries just because of the the risk involved with providing care. So, I would say that in healthcare, it's even more important to have particularly internal audit folks and who's responsible for adhering to the regulations involved in data skin data strategy.

Interviewer: so, there is a box called data privacy and compliance, don't you think regulatory is part of that?

Interviewee: I do right, I do just think it should be up and as part of data strategy. I don't think data strategy should feed data privacy and compliance. I think data privacy and compliance should feed data strategy.

Interviewer: The strategy is between leadership and chief data officer, then the chief data officer interface with the data governance, and then data governance works on the data privacy and compliance. That's the flow, don't you think that the right flow.

Interviewee: yeah, I mean, I think it just needs to be more integrated. I guess you know like even governance, in my opinion, would recommend that there is a data governance like a red block. Honestly, because you know there's so many different groups and there's so many different use cases and sources of data and particularly again you come to an acquisition right where it's kind of like OK what is our data governance policy that should really kind of need to be at a higher level as well in my opinion. This where a business leader says hey, I want to have objective, then goes to strategy on how you are going to use data to help me achieve my objectives. Those guys come up with some ideas, kind of try and flow that through and I think that having the data inform the objective and having the data inform the strategy is a better flow for success.

Interviewer: so, what you're saying is the gap here regulatory? what else you think it's a gap. Interviewee: I think the other gap is how does this feedback on itself. It is a data journeys, it's really about that journey where you should be able to move quickly and fail fast right and when you do that particularly when it comes into the technique that the orange blocks as well as the blue blocks, you're going to try whole bunch of stuff just not going to work for whatever

reason and then like how does that, how do all these things kind of feedback to then inform so that the center of excellence right up at the higher levels.

Interviewer: is that's proof of value block for? And doing lighthouse projects first.

Interviewee: yeah, I mean I would say that proven value kind of ties to value proposition but ultimately there's the connection to the outcome right and I think that outcome really needs to be decided first. That's kind of part of the business objectives and the data strategy. And then I think that needs to be kind of broken down at a higher level because one of the other things that we've seen, particularly, in the data space these days there are a billion companies and there's a new one every day and I'm not getting that. look like one of these blocks that fixes one thing or maybe and accelerates two or three things but then they all have gaps. So how do you kind of put that all together and build that connective tissue and what happens a lot in organizations particularly they're starting out trying to kind of just like boil the ocean. So, having a singular focused use case that you can iterate on you can build like a minimum viable product MVP very quickly and you can succeed or fail right and then iterate through that you know for that very specific finite well defined use case. I think it is the most important thing because once you have a successful process then you can say that you go to the next use case that has a lot of commonalities from connective tissue perspective. Then you can feed through your pipeline.

I just think that value has to be super finite as you know and super focused and really identified back to the business objective like it's hard to go and say you know our outcome is quality healthcare like OK, well there are 300 million ways to have quality healthcare which one? What are we going to pick first? Are we going to do our proven value on first that we can ensure our entire data lifecycle pipeline to that's repeatable?

I think that that's got to be a core key component for success in any data driven organization there's got to be that core group of experts right then they go across all of the different colors of blocks you know their work cohesively because that goes to scale right. I mean at some point once you get good at data in one use case you got skill into four and then you're going to have four more and more people involved, and they've got to have like in a place. So, I would say that was something like fusion team is center of excellence. We are talking about other stakeholders in fusion team and make them the center of excellence. Somehow, I like that. Maybe we define it in a different way, member of that team who are those stakeholders right that they go across all the different responsibilities and personas that are involved.

Interviewer: question number 4 “In your opinion, what will be the most benefiting matter to healthcare organizations when the proposed framework structured approached is followed?” So, the proposed framework is what you're providing a validation. Is the framework valid? if yes, it is valid, what is going to benefit health care organization that's what this question is about.

Interviewee: yeah, I think the number one benefit is called this culture change, I mean ultimately if you know if you can get everyone across the organization to be data driven as in culture, I think that that's the hardest thing to do. Having a framework like this where you made some decisions right you can know what your basic flow is and you've got to go and discover that's where that's where it starts and so that's the really the biggest value is enabling culture change

Interviewer: so, do you think the framework is valid and can be implemented?

Interviewee: yes, I think all frameworks are valid and can be implemented. I think the thing I like about this one in particular is that the foundation is there. I think that if we start having a place to start like this, is awesome, and it's really good. I would be really interested to

see what this diagram looks like a year down the journey, two years down the journey, three years down the journey. I think it will at its core be very similar but then in reality it would be just different depending on kind of the success and failures. So, flexibility in this model needs would be, maybe another thing that maybe I would I recommend. How can this model be flexible right as the journey continued and as the discovery happens?

Interviewer: you have two minutes to conclude and anything you that we didn't cover.

Interviewee: understanding that each of the blocks particularly go the blue blocks and then like you said that the big data needs another 40 blocks underneath the tray. I would recommend to lineages data history which kind of goes along with metadata and cataloging. Definitely relative to data sources particularly in healthcare. One of the things that we see be in the latest data platforms is that lineages can get lost, particularly if the data sources are feeding like enterprise data warehouse and then once that lineages lost now you have regulatory challenges, now you have on a challenge is an older data that can lead to users not trusting the data. That's another big challenge that we see which is kind of like hey we did this big data pipeline; we ran it through the AI model or an ML model and here's your answer and then people are like I don't know about that. I don't trust the data, then they fall back to their excel spreadsheets and so I think lineage should be kind of the center in the orange. One thing I want to emphasize here, are you cannot just underwater framework, is all about you cannot just implement what they call big data, and you think is going to be successful you have to do all of this. No doubt about having a lot of data is interesting, doing something with it is hard

Interviewer: I really appreciate. I want to be respectful of your time. Thank you very much

Interviewee: Thank you, absolutely anything

Interview 9

Interviewer: good morning, thank you very much for joining. The interview agenda is going to be the instruction for two minutes, the framework overview 3 minutes, and then you're going to introduce yourself as a subject matter expert, education, professional experience, title, years of experience. Then I will ask 4 questions and you'll have 5 minutes to answer each question, and then you will have two minutes for conclusion. I will be using Microsoft dictation. We are having this interview over Microsoft Teams, and I am recording the session. The goal of the interview is to validate the framework. It is about big data analytics and big data is one box or one block in the framework. The framework consists of 41 blocks and there are 40 other blocks or components needed for successful big data implementation. The framework is based on people, process, technology model that was created around 1962. In the framework, I added another three layers organization, data, and outcome. The boxes have different colors, brown color is for organization and people. Anything related to data is in, big data is in red, and technology is in blue color and the outcome boxes are in green. The boxes are connected with two ways arrow, one way or just connect. Other boxes are not connected at all, but they need to be in the framework. Hopefully, you reviewed the framework prior to our meeting, and I would like you to introduce yourself, your education, professional experience, titles and years of experience.

Interviewee: yeah, absolutely thank you so much my name is XXXXXX and I am a solutions architect. I have been working in the IT and IT healthcare space for the better part of 20 years. From an educational perspective I've got a degree from Riverside Community College and that's in in math and computer science. I have focused specifically on all aspects of data, AI and ML, modern data pipelines in large enterprises and specific and healthcare.

Interviewer: thank you very much for the introduction that would take us to question number one ": Overall, how well do you think the framework overcome the challenges in healthcare big data Analytics adoption?" this question here it's about the challenges of healthcare with big data and data in general and if it overcomes these challenges.

Interviewee: I think that this framework is a really great foundation to implement across the different pillars that you've outlined here. The reason for that is in healthcare there is a lot of disparate systems, lot of age technology, a lot of new technology that gets introduced on a somewhat regular basis to solve specific clinical problems and so whatever challenges from a healthcare perspective is that because we've had so much disparate data in the past and those systems are not heavily interconnected with one another that that we've lost sight of what we can actually get out of the value of that data that's being created. I think a framework like this that really addresses the additional colors, I love the pillars that you've included in this I think that starts to put a good foundation in place because from a clinical standpoint as well as a business outcome standpoint. Unless we have a framework, we're not going to be overall successful and really turning the corner from a healthcare standpoint and using data as an asset to drive better clinical outcomes in business decisions.

Interviewer: as a follow-up question why? I think you covered some of the why.

Interviewee: I can go a little bit deeper on the why because we've got a lot of so take mergers and acquisitions just in the healthcare space just alone integrating those disparate data sources into existing processes is required. Suppose we're going to continue to merge additional types of datasets into our healthcare system if you will and so that's one aspect. Another aspect is going to be just around improvements to overall patient care. A good, governed framework to do that, without a good framework that's going to complexity and it's going to going bad and

continue the chaos. So, I really love the approach of just bringing standards in a framework and so yeah that's a little bit deeper on the why there.

Interviewer: second question "What does the framework provide to healthcare organizations that they are NOT currently considering when implementing big data analytics?"

Interviewee: so there are few things in the framework are you know not all healthcare organizations are considering, they just want to implement big data and attempt, sorry don't mean to laugh, so from a healthcare perspective again it goes back to an extremely diverse amount of data, so how do I say this the very first step from us oftentimes the mistake that healthcare providers and health systems in general will make is that they think that they're good, they can go implement or bring a new data culture in to the organization and there are oftentimes what I see is that there's not alignment from a clinician perspective and so folks will build things and expect that, that thing that they built, is going to have a good outcome without actually deriving what those things are working hand in hand with the various different clinicians. So, on here at under the people section, I see other stakeholders listed and I view those other stakeholders a largely as the clinicians that are actually responsible for systems say like PACS (picture archiving and communication system), so we've got PACS, and we've got an EMR solution. How do we start to emerge and drive value and what types of value can I actually get from merging those two different datasets, data types together? Applying metadata to that how can I go about improving patient outcomes and what questions can I begin to ask the data. I see folks again they just try to build things and expect that folks will come to use it but the truth is, particularly in big data and analytics is those data structures actually have to be structured to be formatted in a specific way to create re usability of the data across all the various disparate clinical systems that a particular system might be running and I think that's one of the largest

mistakes that folks make and so a lot of it is going to be in the people area and then the process area overall.

Interviewer: So, you think that people and process areas many organizations don't consider implementing it.

Interviewee: Yeah, they don't and constantly, what I've uncovered is depending on who you're talking to with inside of the organization so if it's going to be a clinician stakeholder or a business partner or somebody who's on the IT side of the house right, you're going there's this there's definitions of words that are not consistent across that board so you talked about data you know kind of data governance in cataloging quality and things like that so you've got to have a data dictionary and create standards within standards with inside of your data set. Have a standard data dictionary or glossary, if you will, make sure that the same terms are being used and have the same definitions when it comes to the data space. Make it because there's a good example that what time did, I discharge the patient, OK so does that what time was that was, when I actually keyed on the entry to initiate the discharge or is it when the patient actually left the hospital. There's a there's a lot of different ways that one value could be interpreted by various different systems and that really starts dovetail into the quality of the data. And then you need to backtrack that even further and this is particularly important in healthcare and it's exceptionally difficult to achieve which is ensuring that you've got good data lineages across the board so that you can actually track things that have a system of record and be able to bring history into things that you've listed on here like data warehousing. how do I bring all of the full history from those disparate systems and track that lineage and produce a quality score of that actual data that now I want to do data processing on and things like that really think that is overlooked a lot of health systems today.

Interviewer: do you think that this framework addresses that or include that.

Interviewee: I think the framework would need to add data lineages into the framework. I think that's a really key component. I would shift data cleansing over into a different area because there's a lot of things that can go into data preparation and one of those key things that I talk about a lot is going to be the reusability of the data. I could have multiple data scientists or data engineers really of doing their own thing because one person focuses on all things blood related and pathology, and another one focuses on all things packed than imaging. So, there's value in bold individual datasets and they might have to tap into common data. I want to make sure that all data is reusable and right features so that I can produce those data engineer features to create that reuse ability layer and that's something that I would probably add to this framework. The reusability of data. The common way to do that today, so this has been about 18 months old or something and there's a lot of motion and engineering happening around it, is the concept of something called a “feature store”

Interviewer: question #3 “What step or function does need to be included to the framework to address any gap?” what you mentioned earlier are considered gap.

Interviewee: correct it's again, so data lineages is a gap, reusability of data is the gap. Any other thing. I don't see on here that you could probably add but something high, there are two different ways that data is ingested it through batch integration or streaming. So, when you wanted to real-time analytics that's the streaming component and I don't see a real-time streaming component on here. I see in the data space, data integration/ingestion and on end of data warehousing which is not real-time, so somewhere in there or I would probably add streaming is a key for correct.

Another thing is very important because this touches the people, the process, the data and the technology aspect. It is going to be orchestration. Let me share what I mean by orchestration, so it's a service that a technology so there's components or tools out there, like giving example like Apache airflow or orchestration automation to be able to orchestrate the pipeline end in to and do scheduling around specific thing. So, if I have batch integration and I need to schedule something that's coming in from disparate data sources, I need to orchestrate the scheduling of those data sources and I need to be able to track the performance of those things. If that was going to be something that was done, and I need to have insight into the way that that data is being scheduled and flows throughout this entire framework that you've put together and you want to do that with orchestration software tools that allow you to get those levels of insights into what's actually happening inside of the overall infrastructure. So, I don't think changing management addresses that I think change management component of governance in some way, but I orchestrated orchestration should sit around all of those middle pillars that you've outlined. Nothing else in the gap other than just say that I know that there are some others you know like blocks that might not be represented that's it you've mentioned that kind of sit in that big data box. There is this concept of data virtualization that I don't actually need to be called out on this framework per say but I think that it's an important component to achieving some of the outcomes that you've listed at the bottom so I don't know how it would fit. For data virtualization, let me say that I've got a data lake that has formatting that I've got structured data, I've got semi structured data, and I want to be able to run analytics across those two disparate types of data sources. I have the ability to bring in a data virtualization layer to allow me to virtualize that data so that I can then start deriving the insights out of those multiple sources and I

think that that's an important component to an overall framework. I just don't know where it fits on here.

Interviewer: question #4: In your opinion, what will be the most benefiting matter to healthcare

organizations when the proposed framework structured approached is followed?"

Interviewee: this is going to be a difficult question to answer in a short period of time because I think that is tremendous amount of value and I don't know that it could be quantifiers in boiled down to just some basics but at the end of the day from a healthcare perspective what we want our charter is every single day is to improve our patient outcomes. So as many different facets associated with it in it could be the wait times in the ER all the way through the clinical experience that they've got in the hospital as well as the outpatient care outside of the hospital. Being able to pull and track the patient throughout the patient's life cycle and engagement with the system is going to open up a world of opportunities in my mind that then lead us to other questions, can we ask of the data how we can continue to iterate to improve that successful outcome and also which ways can we begin to track that patient to really help to get. Instead of just this world of prescriptive analytics right how, do we start to get more into the world of predictive analytics and really start to do targeted healthcare based on that one individuals' conditions or abilities etc. and really start to improve that patient's life overall. I think that that is one of the things that out really get from laying out a framework like you have. I think that that is what the charter of healthcare is all about because that dovetails into reduced cost bold from provide patient perspective an efficient use of time operationally and just I think the use cases at that point in time or are near infinite.

Interviewer: you have few minutes to provide any thought in feedback or to conclude.

Interviewee: I see IoT sitting out here a box that's not connected, I would say IoT/IoMT/OT, so Internet of Things, Internet of medical things, and operational technologies. I think all of those things but that is going to be a fantastic data source. I mean gosh medical devices already are a fantastic data source and so I would definitely I would definitely link that in there somehow and then one of the things that I think t that you could simplify in this is, we've got compute network and storage and cloud and cybersecurity, all kind of lumped into blue. I think that we could probably peel security out and make it its own but take cloud, networks, storage and compute and have that just the infrastructure services. I think that would simplify and that's just IT infrastructure.

Interviewer: so, you're saying remove things and just have infrastructure including cloud. Could be any of three things could be SaaS, PaaS and IaaS, data processing and data warehousing are using the cloud, software as a service

Interviewee: I think the reason that I poked out in my mind is I've been working hand in hand with one of my data scientists lately and he has really helped me to view cloud is what it is, it's just a tool in your tool chest and he calls it cloud software infrastructure. if it's just the Lambda function, he uses it as infrastructure because it's just an underlying means to an end. From a data pipeline standpoint that's why that's it that kind of popped out so it's just not correct and I agree with you, when you deal in real like you adjusting the building blocks, so the boxes some of them will fit and how you define things, and how you connect things.

Interviewer: my main question, do you think that framework is valid? and yeah will it benefit healthcare organization?

Interview: I want to conclude here.

Interviewee: Yes, I do think the framework is valid and I think that will have a tremendous number of values to health systems.

Interviewer: thank you very much for your time and I really appreciate it.

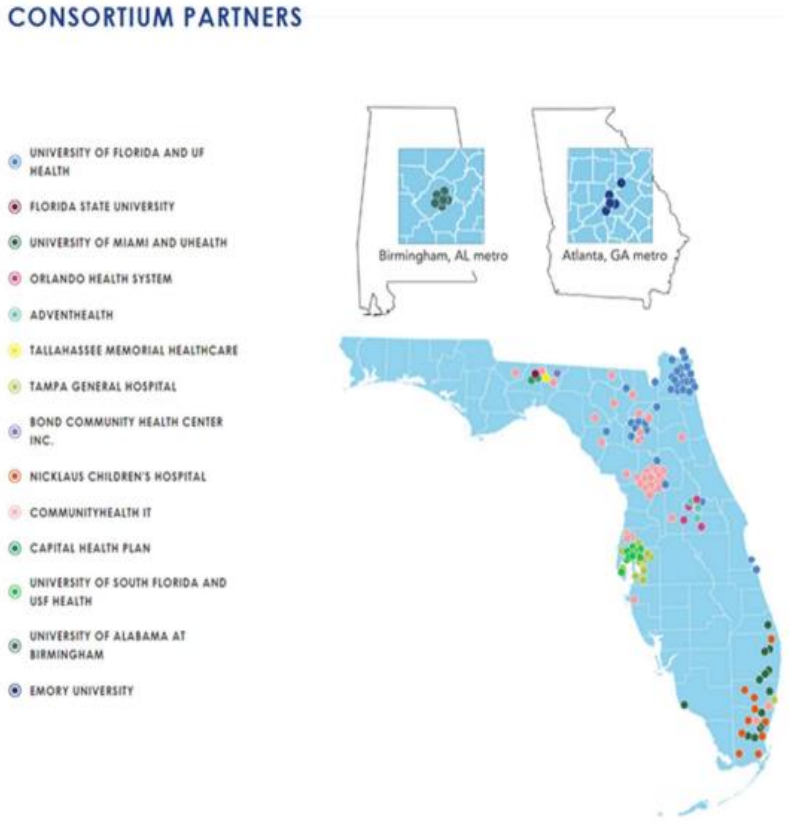
Interviewee: thank you and you're welcome.

APPENDIX B PERMISSION TO USE FIGURES

From: Chen,Yujun <yujunchen@ufl.edu>
Sent: Sunday, October 17, 2021 4:37 PM
To: Mohamed, Ahmed, OneFloridaOperations <OneFloridaOperations@health.ufl.edu>
Cc: Norton,Jane-Ann D <janeann@ufl.edu>; Chisholm,TaJuana <tchishol@ufl.edu>; Galvan-Miyoshi,Yankuic M <yankuic@ufl.edu>
Subject: RE: i2b2

Hi Ahmed,

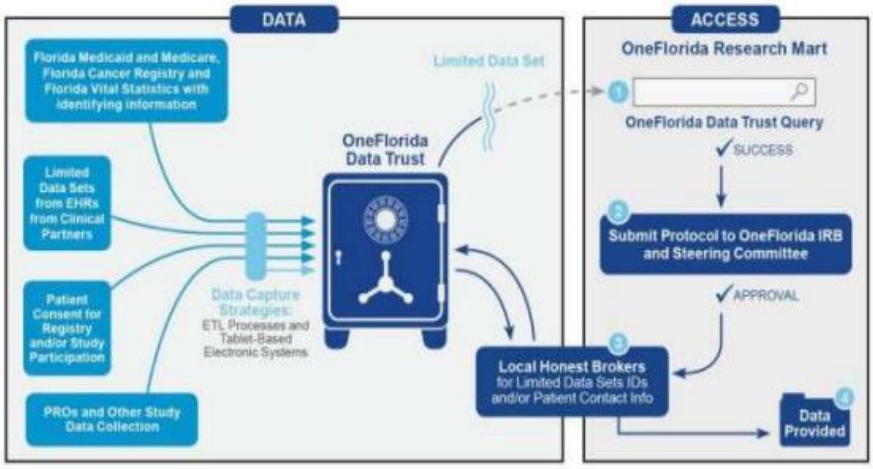
Yes, please feel free to use the information we shared with you. Meanwhile, please notice that oneflorida consortium partners have increased to 14 as shown below.

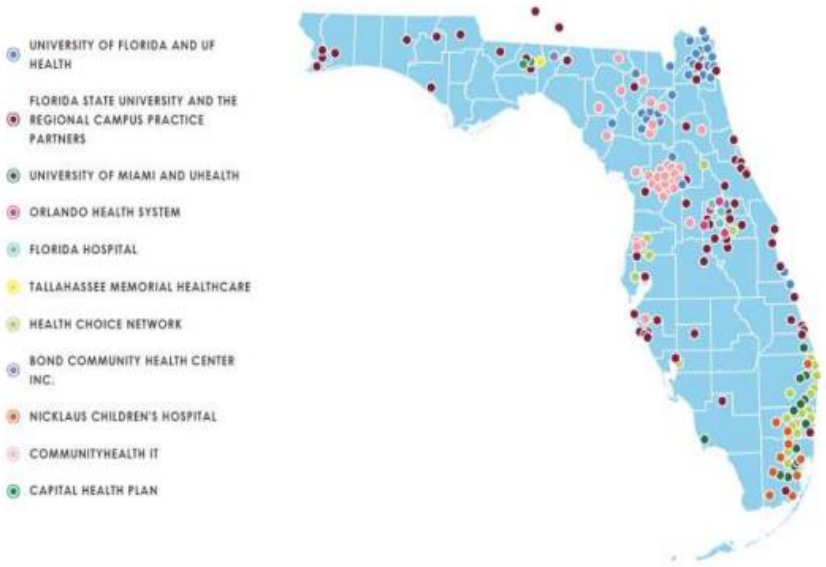


Regards,
Yujun

From: Mohamed, Ahmed
Sent: Sunday, October 17, 2021 12:53 PM
To: OneFloridaOperations <OneFloridaOperations@health.ufl.edu>
Cc: Norton,Jane-Ann D <janeann@ufl.edu>; Chisholm,TaJuana <tchishol@ufl.edu>; Galvan-Miyoshi,Yankuic M <yankuic@ufl.edu>; Chen,Yujun <yujunchen@ufl.edu>
Subject: RE: i2b2

Hello, Do I need a permission to reprint both figures below in my research? If yes, please let me know how I can obtain it. Thx





APPENDIX C IRB APPROVAL

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

NOT HUMAN RESEARCH DETERMINATION

March 16, 2021

Dear Ahmed Mohamed:

On 3/16/2021, the IRB reviewed the following protocol:

Type of Review:	Initial Study
Title of Study:	Big Data Analytics in Real Time For Healthcare Enterprise Performance Measurements
Investigator:	Ahmed Mohamed
IRB ID:	STUDY00002801
Funding:	None
Grant ID:	None
Documents Reviewed:	• Ahmed Mohamed HRP-251- FORM 1.0 signed.pdf, Category: Faculty Research Approval; • Ahmed Mohamed - HRP-250-FORM- 1.0 Intiaialed (1).docx, Category: IRB Protocol; • IRB _GPRO_PREV_SupportingDocument.xlsx, Category: Other

The IRB determined that the proposed activity is not research involving human subjects as defined by DHHS and FDA regulations.

IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities are research involving human in which the organization is engaged, please submit a new request to the IRB for a determination. You can create a modification by clicking **Create Modification / CR** within the study.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Katie Kilgore
Designated Reviewer

APPENDIX D IRB APPROVAL ADDENDUM

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

NOT HUMAN RESEARCH DETERMINATION

April 23, 2021

Dear Ahmed Mohamed:

On 4/23/2021, the IRB reviewed the following protocol:

Type of Review:	Modification / Update
Title of Study:	Big Data Analytics in Real Time For Healthcare Enterprise Performance Measurements
Investigator:	Ahmed Mohamed
IRB ID:	MOD00001813
Funding:	None
Grant ID:	None
Documents Reviewed:	• Ahmed Mohamed - HRP-250-FORM- Request for NHSR 3.docx, Category: IRB Protocol; • Questions to interview SMEs.docx, Category: Interview / Focus Questions

The IRB determined that the proposed activity is not research involving human subjects as defined by DHHS and FDA regulations.

IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities are research involving human in which the organization is engaged, please submit a new request to the IRB for a determination. You can create a modification by clicking **Create Modification / CR** within the study.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Katie Kilgore
Designated Reviewer

LIST OF REFERENCES

Abinaya, Kumar, V., & Swathika. (2015). Ontology Based Public Healthcare System in Internet of Things (IoT). *Procedia Computer Science, 50*, 99-102. doi:10.1016/j.procs.2015.04.067

Adjerid, I., Adler-Milstein, J., & Angst, C. J. I. S. R. (2018). Reducing Medicare spending through electronic health information exchange: the role of incentives and exchange maturity. *29*(2), 341-361.

Alghamdi, A., Baz, A., Alsubait, T., & Alhakami, H. (2021). Healthcare Analytics: A Comprehensive Review. *Engineering, Technology & Applied Science Research, 11*(1), 6650-6655. doi:10.48084/etasr.3965

Almotairi, M. (2009). *A framework for successful CRM implementation.* Paper presented at the European and Mediterranean conference on information systems.

Alshehri, S., & Benedicenti, L. (2013). *Using the analytical hierarchy process as a ranking tool for user story prioritization techniques.* Paper presented at the Proceedings of the 8th International Conference on Software Engineering Advances.

Ambinder, E. P. J. J. o. o. p. (2005). Electronic health records. *1*(2), 57.

Archenaa, J., & Anita, E. A. M. (2015). A Survey of Big Data Analytics in Healthcare and Government. *Procedia Computer Science, 50*, 408-413. doi:http://dx.doi.org/10.1016/j.procs.2015.04.021

Baldominos, A. a. i. u. m. e., De Rada, F., & Saez, Y. (2018). DataCare: Big Data Analytics Solution for Intelligent Healthcare Management. *International Journal of Interactive Multimedia & Artificial Intelligence, 4*(7), 13-20. doi:10.9781/ijimai.2017.03.002

Baro, E., Degoul, S., Beuscart, R., & Chazard, E. (2015). Toward a Literature-Driven Definition of Big Data in Healthcare. *Biomed Research International, 2015*, 1-9 9p. doi:10.1155/2015/639021

Belle, A., Thiagarajan, R., Soroushmehr, S. M., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big Data Analytics in Healthcare. *Biomed Res Int, 2015*, 370194. doi:10.1155/2015/370194

Berry, D. (2018). Advancing the IE Body of Knowledge: Years of work on online, open-access document to be showcased at Annual Conference. *ISE: Industrial & Systems Engineering at Work, 50*(4), 37-40.

Bititci, U., Garengo, P., Dörfler, V., & Nudurupati, S. (2012). Performance Measurement: Challenges for Tomorrow. *International Journal of Management Reviews, 14*(3), 305-327. doi:10.1111/j.1468-2370.2011.00318.x

Bottles, K., Begoli, E., & Worley, B. (2014). Understanding the Pros and Cons of Big Data Analytics. *Physician Executive, 40*(4), 6-12.

Bourne, M., Neely, A., Mills, J., & Platts, K. (2003). Implementing performance measurement systems: a literature review. *International Journal of Business Performance Management*(1).

Braun, V., & Clarke, V. (2012). Thematic analysis.

Brennan, N., Oelschlaeger, A., Cox, C., & Tavenner, M. (2014). Leveraging The Big-Data Revolution: CMS Is Expanding Capabilities To Spur Health System Transformation. *Health Affairs, 33*(7), 1195-1202. doi:10.1377/hlthaff.2014.0130

Brugha, R., Varvasovszky, Z. J. H. p., & planning. (2000). Stakeholder analysis: a review. *15*(3), 239-246.

Camilovic, D. g. E. r., Becejski-Vujaklija, D. b.-v. d. f. r., & Gospic, N. n. g. s. b. a. r. (2009). A Call Detail Records Data Mart: Data Modelling and OLAP Analysis. *Computer Science & Information Systems, 6*(2), 87-110. doi:10.2298/csis0902087C

Campbell, R. (1998). Data warehouses vs. data marts, 32.

Cates, J. E., Gill, S. S., & Zeituny, N. J. I. J. o. B. I. S. (2005). The Ladder of Business Intelligence (LOBI): a framework for enterprise IT planning and architecture. *1*(1-2), 220-238.

Chan, J. O. (2013). An architecture for Big Data analytics. *Communications of the IIMA*(2), 1.

Chawla, N. V., & Davis, D. A. (2013). Bringing big data to personalized healthcare: a patient-centered framework. *JGIM: Journal of General Internal Medicine, 28*, S660-665 661p.

Cortés, R., Bonnaire, X., Marin, O., & Sens, P. (2015). Stream Processing of Healthcare Sensor Data: Studying User Traces to Identify Challenges from a Big Data Perspective. *Procedia Computer Science, 52*, 1004-1009. doi:http://dx.doi.org/10.1016/j.procs.2015.05.093

Cunha, J., Silva, C., & Antunes, M. (2015). Health Twitter Big Bata Management with Hadoop Framework. *Procedia Computer Science, 64*, 425-431. doi:10.1016/j.procs.2015.08.536

Darwish, H., & van Dyk, L. (2016). The industrial engineering identity: from historic skills to modern values, duties, and roles. *South African Journal of Industrial Engineering*(3), 50. doi:10.7166/27-3-1638

Davenport, T. H., Barth, P., & Bean, R. (2012). How 'big data' is different. *MIT Sloan Management Review*(1), 43.

Davenport, T. H. J. H. b. r. (2013). Analytics 3.0. *91*(12), 64-72.

DOT:Website. (2011). Concept of Operations: Data Capture and Management Research Data Exchange Retrieved from

https://www.its.dot.gov/meetings/pdf/ResearchDataExchangeConceptofOperations_August2011DRAFT.pdf

Dworkin, S. L. (2012). Sample size policy for qualitative studies using in-depth interviews. In: Springer.

Eckerson, W. W. J. E. t. V. o. Y. D. W. I. T. B. P. R. (2007). Predictive analytics. *1*, 1-36.

ECRI Website. (2021). Retrieved from https://www.ecri.org/

Fang, H. (2015). Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem. *2015 IEEE 24th International Symposium on Industrial Electronics (ISIE)*, 820.

Fanning, K. (2016). Big Data and KPIs: A Valuable Connection. *Journal of Corporate Accounting & Finance (Wiley)*, *27*(3), 17-19. doi:10.1002/jcaf.22137

Florida HIE Website. (2021). Retrieved from https://www.florida-hie.net/participants/

Frederiksen, L. (2012). Big Data. *Public Services Quarterly*, *8*(4), 345-349. doi:10.1080/15228959.2012.730411

Garets, D., & Davis, M. J. P. w. p. C., HIMSS Analytics. (2006). Electronic medical records vs. electronic health records: yes, there is a difference. 1-14.

Gartner. (2019). Fusion Teams: A New Model for Digital Delivery. Retrieved from https://www.gartner.com/en/documents/3955908/fusion-teams-a-new-model-for-digital-delivery

Gunday, G., Ulusoy, G., Kilic, K., & Alpkan, L. J. I. J. o. p. e. (2011). Effects of innovation types on firm performance. *133*(2), 662-676.

Häyrinen, K., Saranto, K., & Nykänen, P. J. I. j. o. m. i. (2008). Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *77*(5), 291-304.

HealthIT.gov Website. (2020). Retrieved from https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/what-hie

Hu, J., & Nerenz, D. R. J. J. o. H. M. (2020). Performance of Multihospital Health Systems' Flagship Hospitals in the CMS Star Rating Program. *15*(7), 407-410.

IISE BoK Website. (2021). In.

ISO 56000 Innovation Website. (2021). Retrieved from https://www.iso56000.com/

Johri, V., Singh, V., & Srivastava, I. (2014). Predicting Child's Health using Big Data Analytics. *International Journal of Advanced Research in Computer Science*, *5*(7), 193-198.

Kangin, D., Angelov, P., Iglesias, J. A., & Sanchis, A. (2015). Evolving Classifier TEDAClass for Big Data. *Procedia Computer Science*, *53*, 9-18. doi:10.1016/j.procs.2015.07.274

Kaur, K., & Rani, R. (2015). Managing data in healthcare information systems: many models, one solution. *Computer*(3), 52.

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit. [electronic resource] the definitive guide to dimensional modeling*: Indianapolis, Ind. : Wiley, c2013.

3rd ed.

Klettke, M., Awolin, H., Störl, U., Müller, D., & Scherzinger, S. (2017). *Uncovering the evolution history of data lakes.* Paper presented at the Big Data (Big Data), 2017 IEEE International Conference on.

Klotz, L., Horman, M., Bi, H. H., Bechtel, J. J. I. J. o. P., & Management, P. (2008). The impact of process mapping on transparency.

Kondylakis, H., Koumakis, L., Tsiknakis, M., & Marias, K. (2018). *Implementing a data management infrastructure for big healthcare data.* Paper presented at the Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on.

Kuiler, E. W. (2014). From Big Data to Knowledge: An Ontological Approach to Big Data Analytics. *Review of Policy Research, 31*(4), 311-318. doi:10.1111/ropr.12077

Lecklider, T. (2015). More data than you can imagine. *EE: Evaluation Engineering, 54*(5), 20-21.

Leung, L. J. J. o. f. m., & care, p. (2015). Validity, reliability, and generalizability in qualitative research. *4*(3), 324.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*: sage.

Lopes, J., Guimarães, T., & Santos, M. F. J. P. C. S. (2020). Predictive and prescriptive analytics in healthcare: a survey. *170*, 1029-1034.

Madison, K. (2014). HEALTH REGULATORS AS DATA STEWARDS. *North Carolina Law Review, 92*(5), 1605-1636.

Maier, M. W. (1996). *Developments in system architecting.* Paper presented at the Proceedings of ICECCS'96: 2nd IEEE International Conference on Engineering of Complex Computer Systems (held jointly with 6th CSESAW and 4th IEEE RTAW).

Marten, S., Riccardo, S., & Klaus, M. (2012). A framework for business analytics in performance management. *International Journal of Productivity and Performance Management*(1), 110. doi:10.1108/17410401311285327

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: the management revolution. *Harvard business review, 90*(10), 60-68.

McCurdy, R. K. (2001). *A Comparison of the Audit and Accreditation Tools Used By The Health Care Financing Administration, The Texas Department of Insurance, and The National Committee on Quality Assurance: The Cost of Multi-Agency Oversight on Medicare+ Choice Plans in Texas*. Retrieved from

Mello, R., Leite, L. R., & Martins, R. A. (2014). *Is big data the next big thing in performance measurement systems?* Paper presented at the IIE Annual Conference. Proceedings.

Merelli, I., Pérez-Sánchez, H., Gesing, S., & D'Agostino, D. (2014). Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives. *Biomed Research International, 2014*, 134023-134023. doi:10.1155/2014/134023

Mezghani, E., Exposito, E., Drira, K., Da Silveira, M., & Pruski, C. (2015). A Semantic Big Data Platform for Integrating Heterogeneous Wearable Data in Healthcare. *J Med Syst, 39*(12), 185. doi:10.1007/s10916-015-0344-x

Microsoft Website - Azure Pricing calculator. (2021). Retrieved from https://azure.microsoft.com/en-us/pricing/calculator/

Minger, S. (2014). Inspiring innovation. *Chemistry & Industry*, *78*(7), 38.

Moore, K. D., Eyestone, K., & Coddington, D. C. (2013). The big deal about big data. *Healthcare Financial Management: Journal Of The Healthcare Financial Management Association, 67*(8), 60.

Mullins, C. S. (2018). Data Lake Versus Data Warehouse: Understanding the Differences, 33.

Nelson, E. C., Homa, K., Mastanduno, M. P., Fisher, E. S., Batalden, P. B., Malcolm, E. F., . . . Safety, P. (2005). Publicly reporting comprehensive quality and cost data: A health care system's transparency initiative. *31*(10), 573-584.

OneFlorida i2b2 Data Guide. (2018). In U. o. Florida (Ed.).

Ouayogodé, M. H., Colla, C. H., & Lewis, V. A. (2017). *Determinants of success in Shared Savings Programs: An analysis of ACO and market characteristics.* Paper presented at the Healthcare.

Pace, D. K., & Sheehan, J. (2002). *Subject matter expert (SME)/peer use in M&S V&V.* Paper presented at the Foundations for the V&V in the 21st Century workshop (Foundations 2002), John Hopkins University.

Parkash, S., & Kaushik, V. K. J. L. (2011). SUPPLIER PERFORMANCE MONITORING AND IMPROVEMENT (SPMI) THROUGH SIPOC ANALYSIS AND PDCA MODEL TO THE ISO 9001 QMS IN SPORTS GOODS MANUFACTURING INDUSTRY. *7*(4).

Parmenter, D. (2010). *Key performance indicators. [electronic resource] : developing, implementing, and using winning KPIs*: Hoboken, N.J. : John Wiley & Sons, c2010. 2nd ed.

Patrick, J. R. (2015). *The effect of cardionet home telemonitoring for congestive heart failure patients: An observational research study.* (75), ProQuest Information & Learning, Retrieved from https://login.ezproxy.net.ucf.edu/login?auth=shibb&url=https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2015-99080-235&site=eds-live&scope=site Available from EBSCOhost psyh database.

Phillips, J. J., Buzachero, V., Phillips, P., & Phillips, Z. (2013). *Measuring ROI in Healthcare: Tools and Techniques to Measure the Impact and ROI in Healthcare Improvement Projects and Programs: Tools and Techniques to Measure the Impact and ROI in Healthcare Improvement Projects and Programs*: McGraw Hill Professional.

Pipe, P. (1997). The Data Mart: A New Approach to Data Warehousing. *International Review of Law, Computers & Technology, 11*(2), 251-262. doi:10.1080/13600869755695

Re: Presentation Strategy and Innovation (2021).

Quwaider, M., & Jararweh, Y. (2015). A cloud supported model for efficient community health awareness. *Pervasive and Mobile Computing*. doi:10.1016/j.pmcj.2015.07.012

Rodriguez, F., Scheinker, D., & Harrington, R. A. J. C. r. (2018). Promise and perils of big data and artificial intelligence in clinical medicine and biomedical research. *123*(12), 1282-1284.

Saria, S. (2014). A $3 Trillion Challenge to Computational Scientists: Transforming Healthcare Delivery. *IEEE Intelligent Systems, 29*(4), 82-87. doi:10.1109/MIS.2014.58

Sarkar, A., Mukhopadhyay, A. R., & Ghosh, S. K. J. I. J. o. L. S. S. (2011). Selection of critical processes for “process improvement”.

Shah, S. (2016). Big Data In Practice: How 45 successful companies used big data analytics to deliver extraordinary results. *Actuary*, 33.

Shenkman, E., Hurt, M., Hogan, W., Carrasquillo, O., Smith, S., Brickman, A., & Nelson, D. J. A. M. (2018). OneFlorida Clinical Research Consortium: linking a clinical and translational science institute with a community-based distributive medical education model. *93*(3), 451.

Shneiderman, B., & Plaisant, C. (2015). Sharpening analytic focus to cope with big data volume and variety. *IEEE Computer Graphics and Applications*(3), 10.

Sony, M., Antony, J., & Douglas, J. A. J. T. T. J. (2020). Essential ingredients for the implementation of Quality 4.0: A narrative review of literature and future directions for research.

Srinivasan, U., & Arunasalam, B. (2013). Leveraging Big Data Analytics to Reduce Healthcare Costs. *IT Professional, 15*(6), 21-28. doi:10.1109/MITP.2013.55

Steiner, R. (1998). *4.1. 3 System Architectures and Evolvability: Definitions and Perspective.* Paper presented at the INCOSE International Symposium.

Sullivan, T. (2017). Epic's rival EHR vendors say they too are making the 'CHR' switch. In: Healthcare IT News.

Sundharakumar, K. B., Dhivya, S., Mohanavalli, S., & Chander, R. V. (2015). Cloud Based Fuzzy Healthcare System. *Procedia Computer Science, 50*, 143-148. doi:10.1016/j.procs.2015.04.076

Terry, N. (2015). Navigating the Incoherence of Big Data Reform Proposals. *Journal of Law, Medicine & Ethics, 43*, 44-47 44p. doi:10.1111/jlme.12214

Teter, M. D. (2014). *Applying subject matter expertise (SME) elicitation techniques to TRAC studies*. Retrieved from

Tobias, K., Marten, S., & Klaus, M. (2011). Integrating business analytics into strategic planning for better performance. *Journal of Business Strategy*(6), 30. doi:10.1108/02756661111180113

ȚOle, A. A. a. t. g. c. (2015). The Importance of Data Warehouses in the Development of Computerized Decision Support Solutions. A Comparison between Data Warehouses and Data Marts. *Database Systems Journal, 6*(4), 18-26.

Tortorella, G. L., Fogliatto, F. S., Mac Cawley Vergara, A., Vassolo, R., Sawhney, R. J. P. P., & Control. (2020). Healthcare 4.0: trends, challenges and research directions. *31*(15), 1245-1260.

Trebble, T. M., Hansi, N., Hydes, T., Smith, M. A., & Baker, M. J. B. (2010). Process mapping the patient journey: an introduction. *341*.

van den Hoven, J. (1998). Data marts: Plan big, build small. *Information Systems Management, 15*(1), 71.

Viceconti, M., Hunter, P., & Hose, R. (2015). Big data, big knowledge: big data for personalized healthcare. *IEEE Journal Of Biomedical And Health Informatics, 19*(4), 1209-1215. doi:10.1109/JBHI.2015.2406883

Walterbusch, M., Martens, B., & Teuteberg, F. J. M. R. R. (2013). Evaluating cloud computing services from a total cost of ownership perspective.

Whittemore, R., Chase, S. K., & Mandle, C. L. J. Q. h. r. (2001). Validity in qualitative research. *11*(4), 522-537.

Wong, H. T., Yin, Q., Murray, K., Slade, D., Zhou, D. H., & Guo, Y. Q. (2015). Big data as a new approach in emergency medicine research. *Journal of Acute Disease, 4*(3), 178.

Wu, D. (2017, 11-14 Dec. 2017). *A big data analytics framework for forecasting rare customer complaints: A use case of predicting MA members' complaints to CMS.* Paper presented at the 2017 IEEE International Conference on Big Data (Big Data).

Wyllie, D., & Davies, J. (2015). Role of data warehousing in healthcare epidemiology. *Journal of Hospital Infection*(4), 267. doi:10.1016/j.jhin.2015.01.005

Yang, C. C., & Veltri, P. (2015). Intelligent healthcare informatics in big data era. *Artificial Intelligence in Medicine*, *65*(2), 75-77. doi:10.1016/j.artmed.2015.08.002

Yang, H., Kundakcioglu, E., Li, J., Wu, T., Mitchell, J. R., Hara, A. K., . . . Weaver, K. E. (2014). Healthcare Intelligence: Turning Data into Knowledge. *IEEE Intelligent Systems, 29*(3), 54-68. doi:10.1109/MIS.2014.45

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5): sage.

Young, J. M. (2014). An epidemiology of big data.

Yusoff, M. S. B. J. R. (2019). ABC of content validation and content validity index calculation. *11*(2), 49-54.