Big Data Architecture and Governance

Individual Project | Brigham and Women's Hospital

Navaneeta Naik

Northeastern University
College of Engineering

Brigham and Women's Hospital

- Industry: Healthcare
- History/Background:
  - Brigham and Women's Hospital was established with the 1980 merger of three Harvard-affiliated hospitals
  - In the early 1990s, BWH pioneered Computerized Physician Order Entry (CPOE) to prevent medication errors, which is now a nationally-accepted safety practice. BWH has received patient safety awards for its electronic Medication Administration Record (eMAR) and barcoding system
  - In the 2019 U.S. News & World Report hospital rankings, BWH was ranked second in Massachusetts and thirteenth nationally
  - Their Research Institute is one of the most powerful biomedical research institutes in the world and the third largest recipient of National Institutes of Health (NIH) funding

SWOT Analysis

Strengths
- Biomedical Research NIH funding
- Use of Machine Learning and AI in CPOD
- Competitive & Skilled Staff
- High Annual Revenue $7.6B
- Computerized Physician Order Entry (CPOE)
- Efficient Supply Chain

Weaknesses
- COVID-19 increased vacancy rates
- Unpaid Internship
- Staffing Management problems

Opportunities
- Improvement of Leadership
- Innovations in research
- International Collaborations

Threats
- Fierce Competitors – Dana Farber, Mass General
- Staff loss due to better employment opportunities
- Alternative Medicine

Project 1: BioMed

Issue Description

- Everyday, BWH witnesses a lot of patients with various medical backgrounds and health conditions.
- The data from these patients can help in biomedical research and drug discovery
- Using this huge amount of data can be very useful in various research studies and developing precision therapies and assessing risks of therapies and medications

Impact Statement : BioMed

- Issue Name: BioMed
- The purpose of this research study is to effectively assess and improve patient care and research at Brigham and Women’s Hospital

Project Mandate: BioMed

→ Page 1/3

1. PURPOSE

The purpose of this project is to provide analytics based on the data gathered from patients to be used for biomedical research and develop precision therapies at BWH Brigham Research Institute (BRI)

2. SPONSOR

The sponsor for this project will be Patient Centered Comparative Effectiveness Research Center's leadership team that can access patient data, previous research data. The data analytics team will be responsible for processing this data and providing insights to various departments and the CIO will oversee these processes

3. BACKGROUND

This project is a part of the larger program at BRI that is the Patient Centered Comparative Effectiveness Research Center where they mainly focus on improving patient care and health reform. Using this project, data can be collected & processed to form meaningful reports and insights to improve patient care and research

4. PROJECT OBJECTIVES

This project will be able to process and use the data generated and harness it to conduct analysis and gain insights on the patient data to use it in research and providing this data in a meaningful way to practitioners to allow them to understand patient needs and precision therapies for various groups to cater to the needs.

Project Mandate: BioMed

→ Page 2/3

The cost in getting this data is not very high as the data will be collected from EHR, Biomedical research historical data, surgery records, patient records at BWH. The cost will however be outweighed by the profits and the forecasts that the researchers and practitioners will gain by the reports and insights provided by the project.

5. SCOPE

The major dependency of this project is the data that is generated by BRI data, patient records and EHR data . Based on the analytics provided by this data, better decisions can be made in terms of the therapies, research at BRI

6. CONSTRAINTS

Since this project is mainly based on data collection, analysis and deriving meaningful insights. One of the constraints associated with the project would be to get the required amount of resources such as the analytics team to perform the analytics. Also, the time required to distinguish the useful data from the collected (historical) data, can cause significant obstruction to the free flow of this project.

7. INTERFACES

The current and historical data collected from BWH and patients will be a major interface for this project. Data analysis, Business Intelligence will also be a major part in this project.

Project Mandate: BioMed

→Page 3/3

8. QUALITY EXPECTATIONS

- The quality expectation would be that the data collected, and the insights provided by this data will enable in better allocation of resources for research and improvement of the patient care
- The key requirements for the project is rigorous data collection and analysis
- The data collection should conform to or reference national and international standards and protocols, if they exist

9. OUTLINE BUSINESS CASE

- The business justification for doing this project is to improve research at BWH Brigham Research Institute Patient Centered Comparative Effectiveness Research Center
- The insights derived from this project will improve precision therapies and patient care at Brigham and Women's Hospital

Vision Diagram: BioMed

DATA SOURCE

Patients Health Records

Hospital Data Records

Research Data

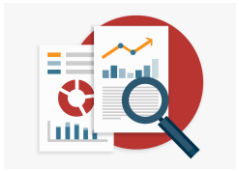
DATA INGESTION

Data Integration

python talend tableau hadoop

Data Analysis

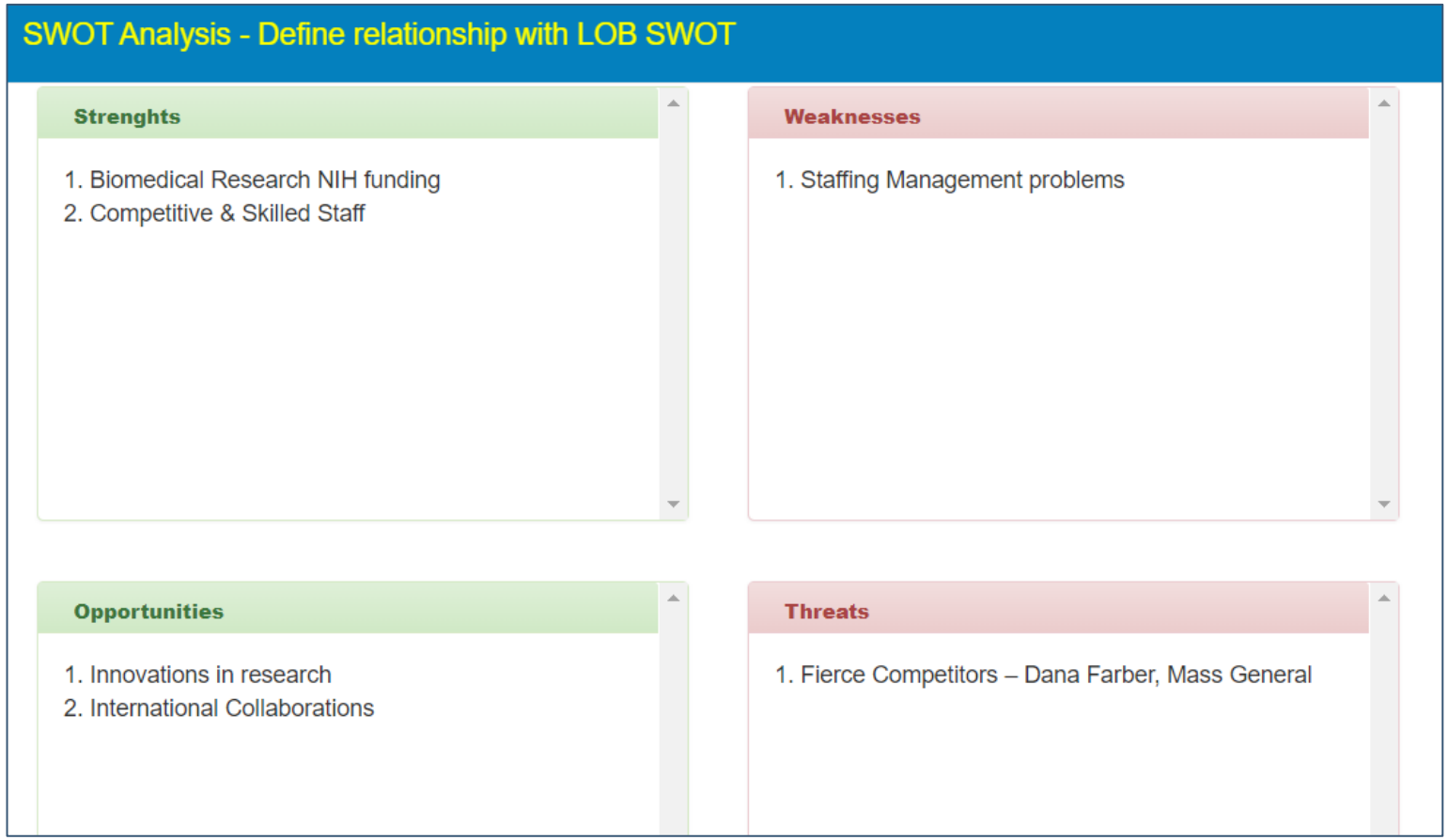
DESTINATION



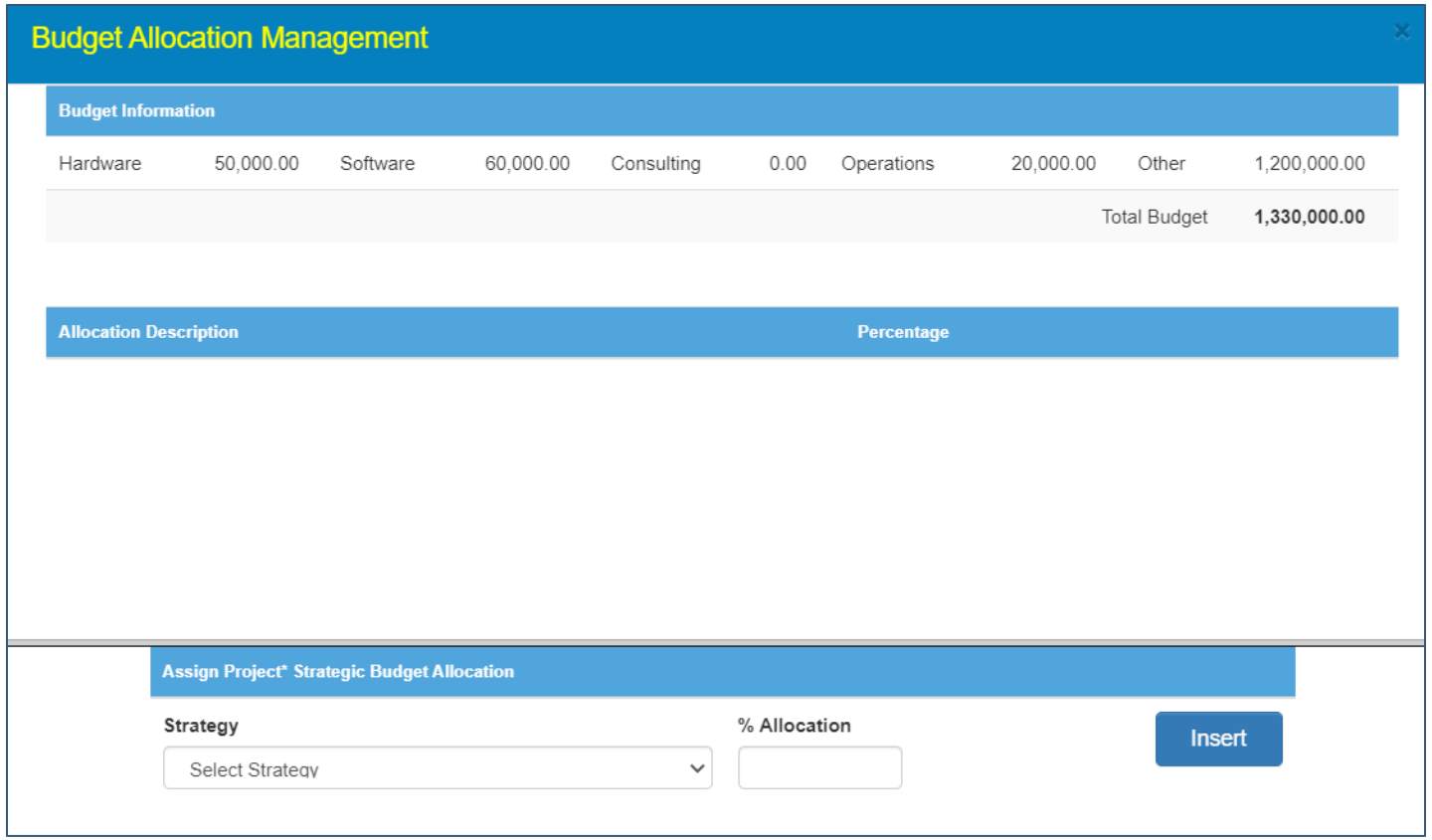
Visualizations

Reports sent to BRI for drug discovery and precision medicine research

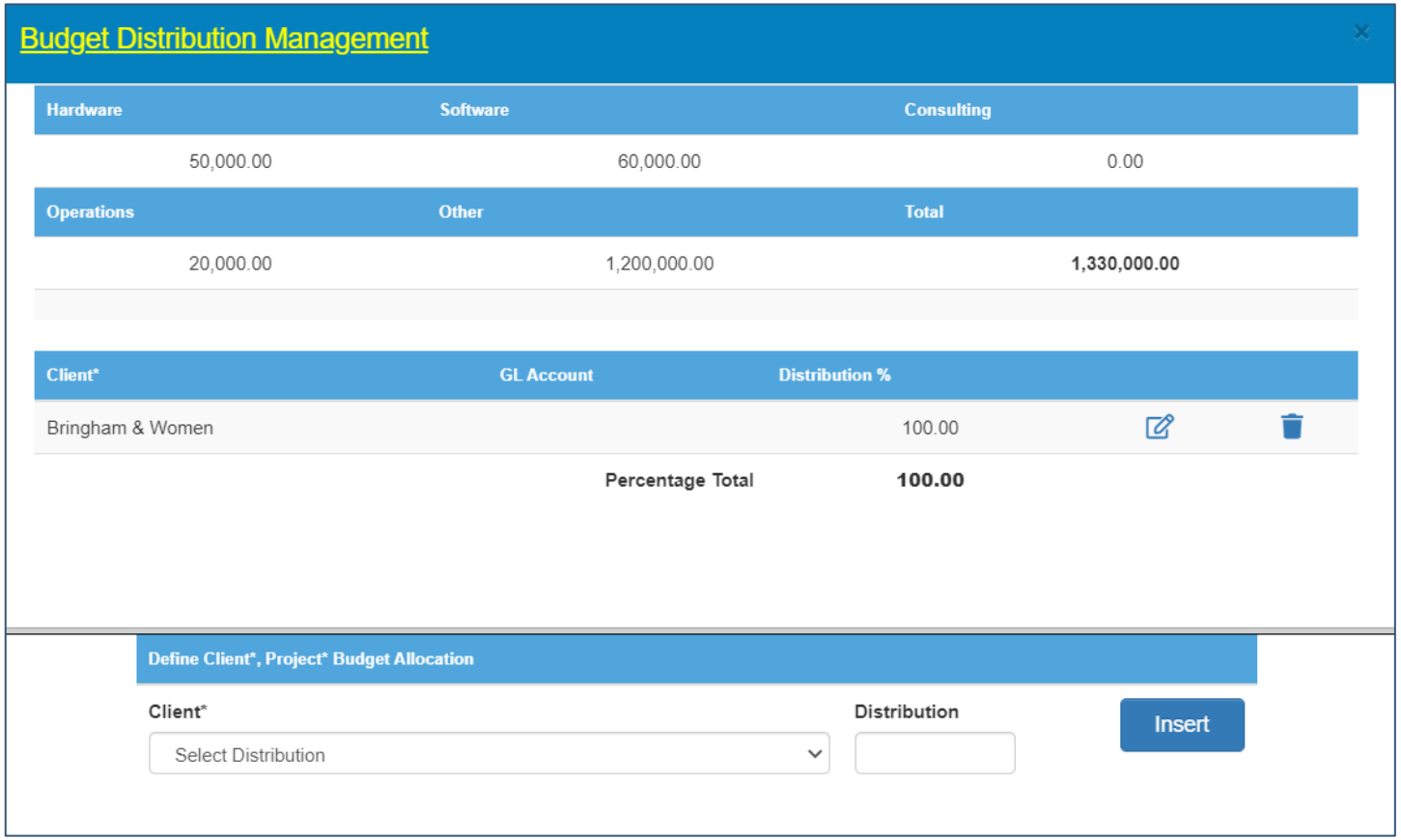
SWOT Analysis BioMed



Budget Allocation: BioMed



Distribution: BioMed



ROI: BioMed

Return On Investment Information

Project* Estimated Cost saving information **Record is Updated!**

Year	Qtr	FTE	FTE Cost	Other Cost	Note	Ops Cost	Note	Updated By	Date
2022	1	7	700,000.00	1,000.00	Reduce 7 employees	0.00		Navaneeta, Naik	26 Apr 2021
2022	3	2	180,000.00	0.00	Reduce 2 employees	0.00		Navaneeta, Naik	26 Apr 2021
		Total	9	880,000.00	1,000.00	0.00			

Resource Planning: BioMed

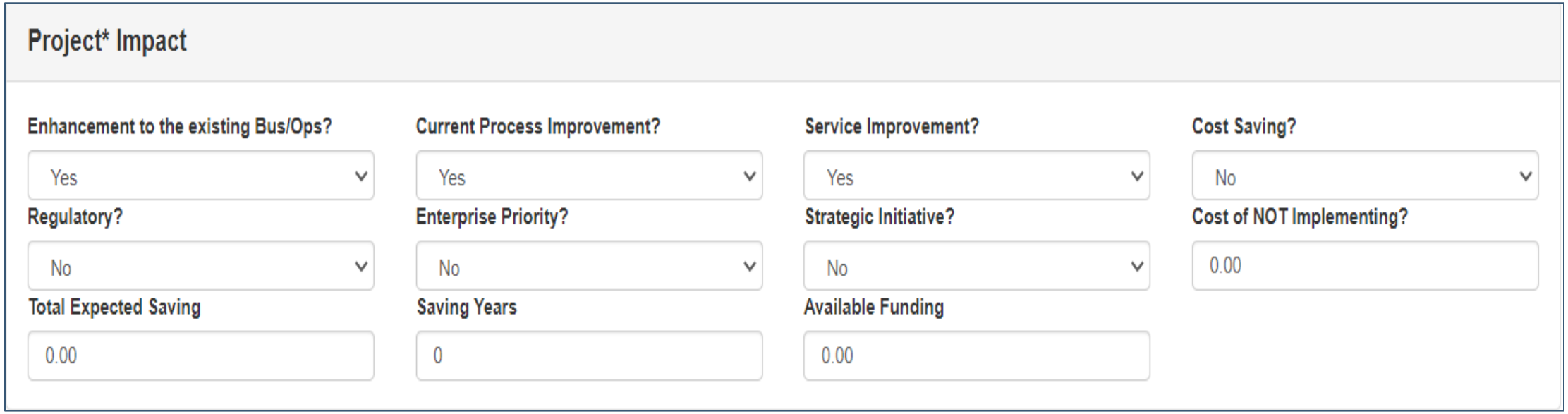
Resource Planning -- BioMed Start from(2021)

Category/Name	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Business Architects	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00
Data Analyst	2.00	2.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	4.00	4.00	4.00
Data Engineer	2.00	2.00	5.00	5.00	5.00	5.00	4.00	4.00	4.00	4.00	4.00	3.00
Data visualizers	1.00	0.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00
Database Admin	2.00	2.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
Datacenter Technician	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Network Architect	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Project Manager	1.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00

Risks & Issues: BioMed

		Type	Exposure	Date	Assign Date	Expected Date	Description	Date Closed	Resource	Updated	By
		Issue	3	03/10/2021	03/10/2021	03/10/2021	Cost Management is an issue associated with this project as it will require right talent and tools f+		Navaneeta, Naik	11/03/2021	Navaneeta, Naik
		Risk	3	03/05/2021	03/12/2021	03/10/2021	There may be a lot of inaccurate and missing data which needs to be identified and filtered +		Navaneeta, Naik	11/03/2021	Navaneeta, Naik
		Risk	4	03/05/2021	03/05/2021	03/05/2021	The data that will be used for this project will be collected from multiple sources and storage of t+		Navaneeta, Naik	11/03/2021	Navaneeta, Naik
		Risk	4	03/20/2021	03/20/2021	03/20/2021	The main basis of this project depends mainly on good analytics. If the data analytics performed is +		Navaneeta, Naik	11/03/2021	Navaneeta, Naik

Impact: BioMed



Project 2: Cross Departmental EHR

Issue Description

- Once a patient is discharged after receiving the necessary medication, there is no way for the doctor to keep a tab on his health
- After a few months, the same patient may come into a different department for a different issue
- There is no data that will inform the doctor about the patient's previous records of medications even within the same hospital
- For this, healthcare providers or practitioners should have access to a huge amount of patient health data and be able to share it securely within different departments

Impact Statement: Cross Departmental EHR

- Issue Name: Cross Departmental EHR
- The aim of this project is to improve patient care and quality of services for patients at Brigham and Women’s Hospital.

Project Mandate: Cross Departmental EHR

→ Page 1/3

1. PURPOSE

The purpose of this project is to improve patient care and quality of services at Brigham and Women's Hospital by giving an access to all medical and medication records of a patient across all departments.

2. SPONSOR

The sponsor for this project will be executives from the enterprise data team and asset management team. They will have access to EHR and data from various departments across BWH.

3. BACKGROUND

Healthcare providers or practitioners should have access to a huge amount of patient previous health data and be able to share it securely within different departments. The project will be a stand-alone activity to improve patient care.

4. PROJECT OBJECTIVES

The practitioner having access to all previous medical history of a patient across all departments over the years, will help determine patient health conditions and right track of medication and therapies. This will improve patient care at BWH and avoids unnecessary use of resources which will otherwise not be useful.

Project Mandate: Cross Departmental EHR

→ Page 2/3

The cost will be for getting a secure platform to share this data between departments, but that will be worth it because it will be helpful for practitioners and patients

5. SCOPE

The major deliverables of the Project is a secure way to share the EHR across the departments. The major dependencies would be that current and historical data will have to consolidated to generate a standardized data manual template so that it can be used across the departments

6. CONSTRAINTS

The constraints would be the availability of data from all departments, consolidation, data cleaning and processing

7. INTERFACES

All the patients' data and historical data across all the departments is needed for this project. This can be a tedious and time-consuming process; hence integration tools and big data analysis can be used, and cloud services will be needed to store this huge data

Project Mandate: Cross Departmental EHR

→Page 3/3

8. QUALITY EXPECTATIONS

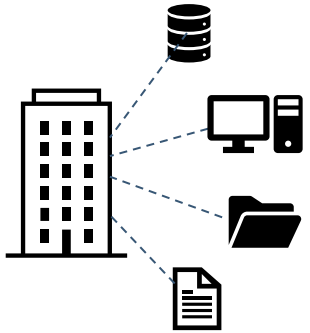
- The quality expectation would be that the data shared across the departments will help in the improvement of the patient care at Brigham and Women's Hospital
- The key requirements for the project is continuous data collection and updating, storage and ensuring appropriate maintenance of this data so it can be used by all the departments
- The data collection and updating should conform to or reference national and international standards and protocols, if they exist and made sure that only authorized personnel get access to this secure data

9. OUTLINE BUSINESS CASE

- The strategic need for the project is to improve how practitioners provide diagnosis, therapies and medication only based on symptoms without access to previous records of the patient
- This data will help practitioners get access to secure patient data that can be used to improve patient care diagnosis and medication
- This will also improve the reputation of the hospital a notch above where it stands now

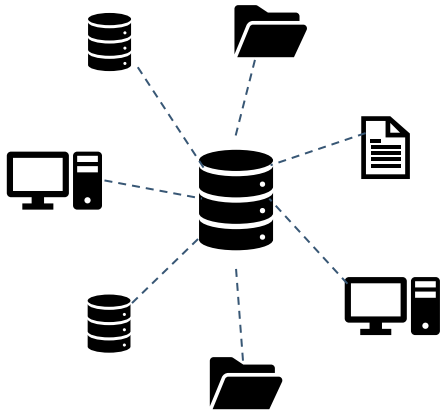
Vision Diagram: Cross departmental EHR

DATA SOURCE



Electronic Health Records from all departments

DATA INGESTION

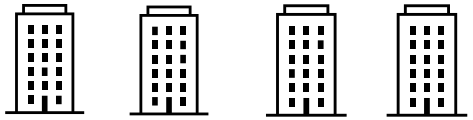


Masterdata

Access given to authorized people

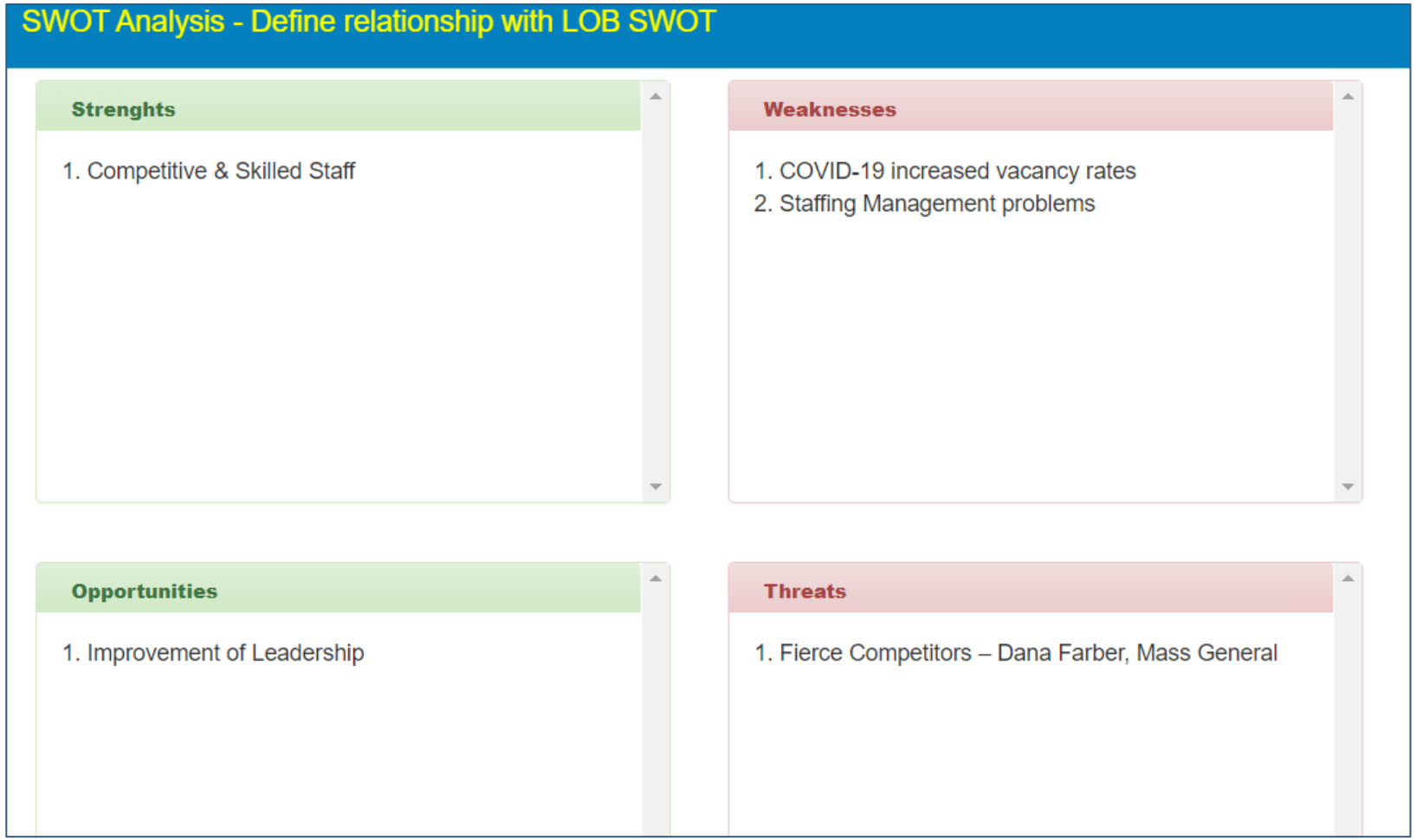
DESTINATION

Masterdata

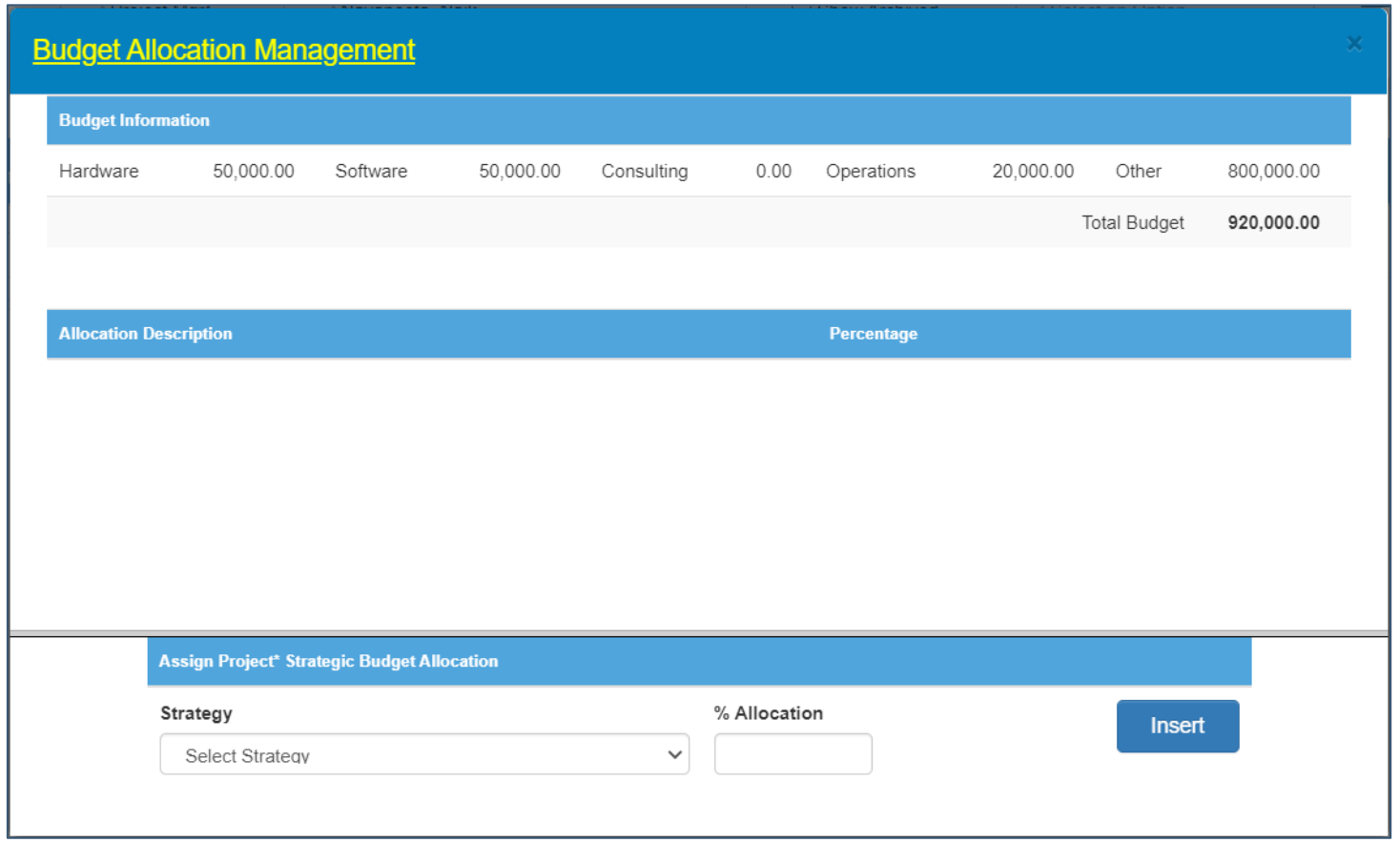


Data accessed securely by all departments' authorized personnel

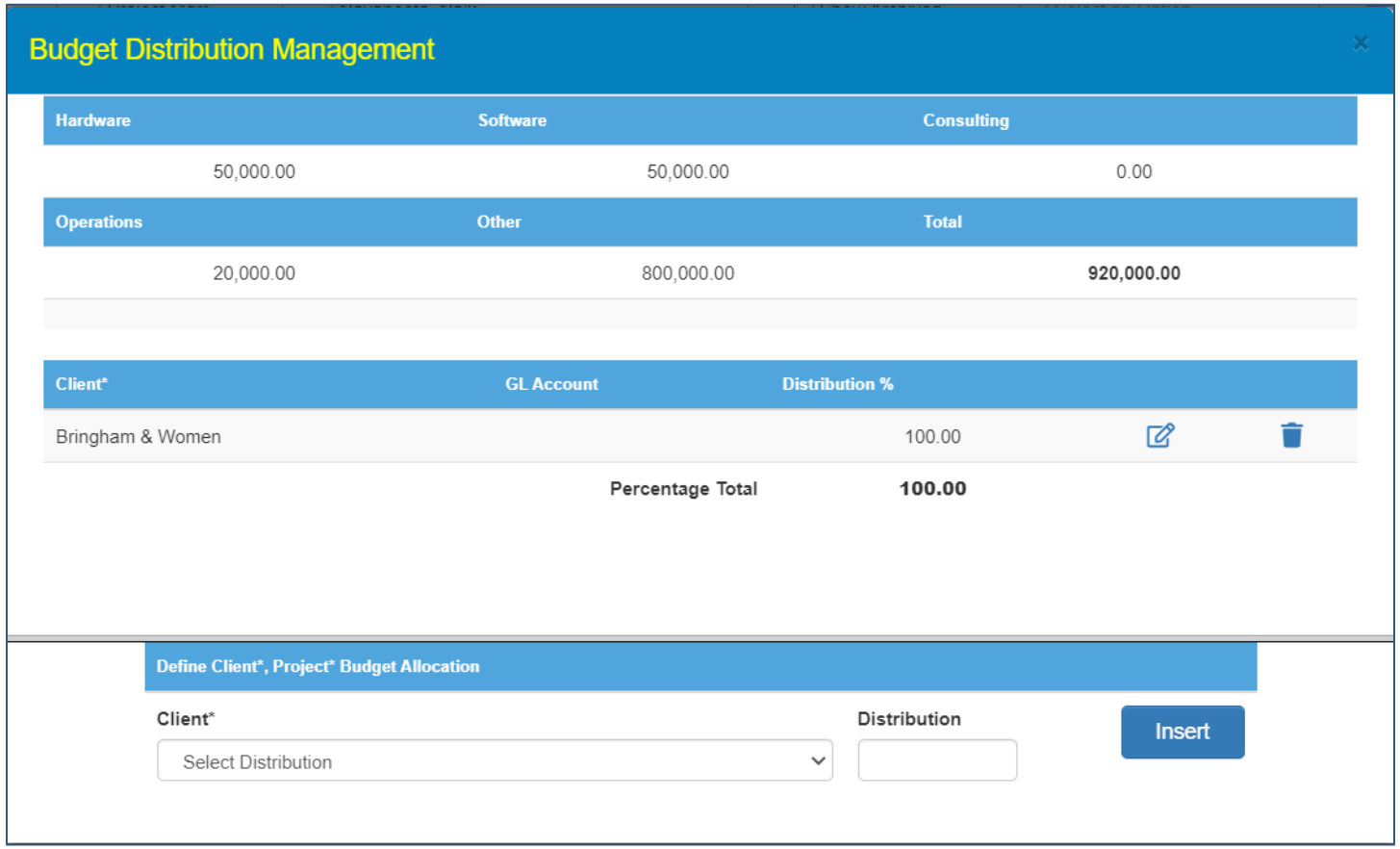
SWOT Analysis: Cross departmental EHR



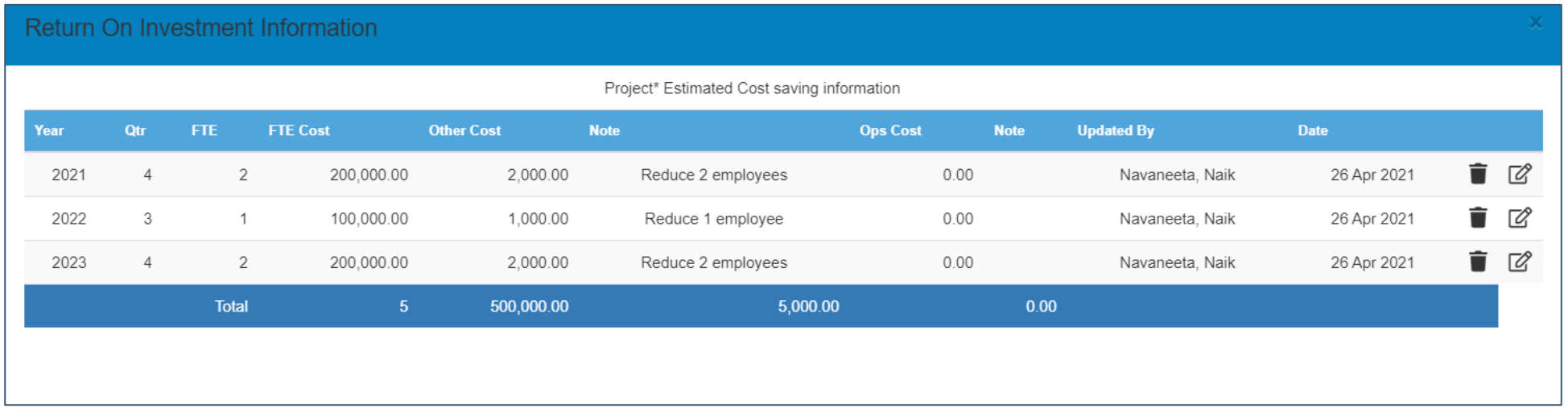
Budget Allocation: Cross departmental EHR



Distribution: Cross departmental EHR



ROI: Cross departmental EHR



Return On Investment Information

Project* Estimated Cost saving information

Year	Qtr	FTE	FTE Cost	Other Cost	Note	Ops Cost	Note	Updated By	Date
2021	4	2	200,000.00	2,000.00	Reduce 2 employees	0.00		Navaneeta, Naik	26 Apr 2021
2022	3	1	100,000.00	1,000.00	Reduce 1 employee	0.00		Navaneeta, Naik	26 Apr 2021
2023	4	2	200,000.00	2,000.00	Reduce 2 employees	0.00		Navaneeta, Naik	26 Apr 2021
		Total	5	500,000.00	5,000.00		0.00		

Resource Planning: Cross departmental EHR

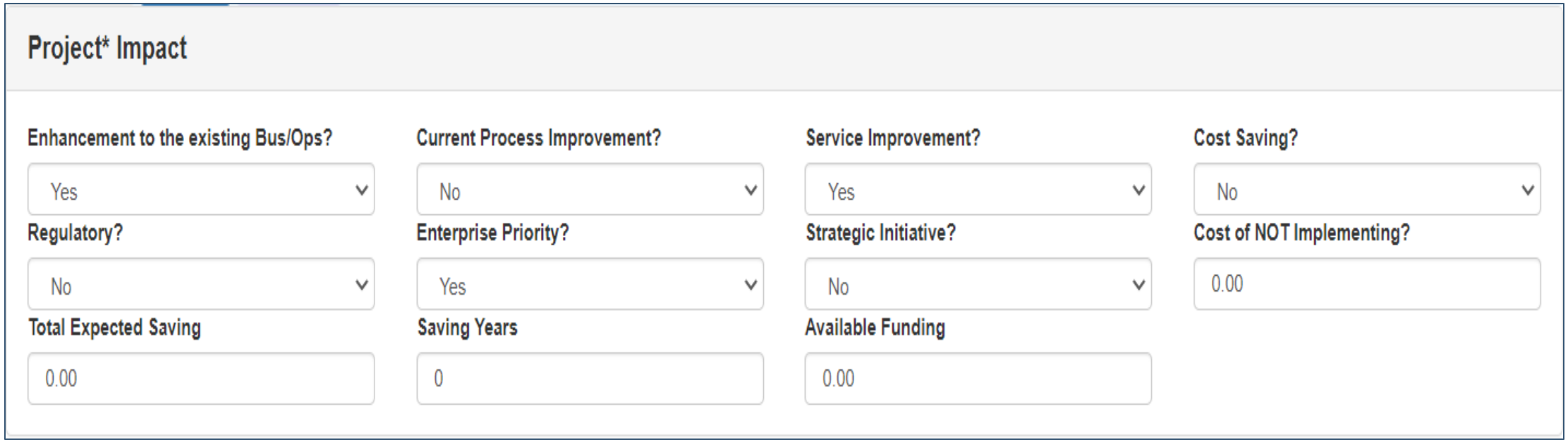
Resource Planning -- Cross Departmental EHR Start from(2021)

Category/Name	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Business Architects	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
Data Engineer	3.00	3.00	6.00	6.00	6.00	6.00	6.00	6.00	5.00	5.00	5.00	4.00
Database Admin	2.00	2.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
Network Architect	2.00	2.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
Project Manager	2.00	2.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
Systems Architect	1.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00

Risks & Issues: Cross departmental EHR

		Type	Exposure	Date	Assign Date	Expected Date	Description	Date Closed	Resource	Updated	By
		Issue	2	03/05/2021	03/05/2021	03/05/2021	Cost Management is an issue associated with this project as it will require right talent and tools f+		Navaneeta, Naik	11/03/2021	Navaneeta, Naik
		Issue	6	03/05/2021	03/05/2021	03/05/2021	Getting access to EHR will not be easy due to HIPAA regulations and protection laws+		Navaneeta, Naik	11/03/2021	Navaneeta, Naik
		Risk	3	03/10/2021	03/10/2021	03/10/2021	Storage of the huge amount of EHR data so that it is accessible across departments will be an issue +		Navaneeta, Naik	11/03/2021	Navaneeta, Naik
		Risk	5	03/05/2021	03/05/2021	03/05/2021	Data breaches are the most common risks associated with healthcare data+		Navaneeta, Naik	11/03/2021	Navaneeta, Naik
		Risk	7	03/05/2021	03/10/2021	03/10/2021	EHR data is most sensitive data and controlled access to this huge consolidated data is one of the o+		Navaneeta, Naik	11/03/2021	Navaneeta, Naik

Impact: Cross Departmental EHR



Project Selection

Factors	BioMed	Cross Departmental EHR
Budget	$1,330,000	$920,000
Return of Investment	$880,000	$500,000
Impact	Improve Research at BRI	Improve Patient Care at BWH
Benefits	Drug Discovery & Precision Therapy studies	Helps physicians make better decisions and developing treatments

Project Selection

- After considering the analytic and business prospects for the two projects, I would like to go ahead with the Cross Departmental EHR project.
- Has a measurable and attainable goal
- Less Investment
- The success of this project is crucial for the betterment of patient care services at Brigham and Women’s Hospital
- This project will help practitioners and patients at large by giving them a history of all patients securely

FUNCTIONAL & NON-FUNCTIONAL REQUIREMENTS

Functional Requirements

- **Data:** The EHR data is available in many forms of structured and unstructured data, this will need to be integrated
- **Robust system:** Since this project involves large amount of data, the system should be robust
- **Reporting:** This data can be made available in the form of reports & dashboards to researchers and scientists for precision therapy & drug discoveries.
- **Authentication:** Proper authentication needs to be placed as this is very sensitive patient information and only authorized people should get access to this information. Also, all the users must be monitored for suspicious activities
- **Cybersecurity:** Data needs to be protected against data breaches, cyber threats
- **HIPAA** compliance and laws regarding medical records need to be adhered

Non-Functional Requirements

Reliability

- Availability: The system should always be up and running as the doctors would access it at any time
- Correctness- The system should assure to provide correct & reliable information to the doctors

Performance

- The system should ensure optimal performance for data acquisition, integration, storage & retrieval
- There should be no data loss
- Able to handle heterogeneous data sources

Usability

- Must have well structured user manuals and informative error messages
- Help facilities and tutorial videos to use the system data
- Must have scheduled downtime and maintenance during the least affected time

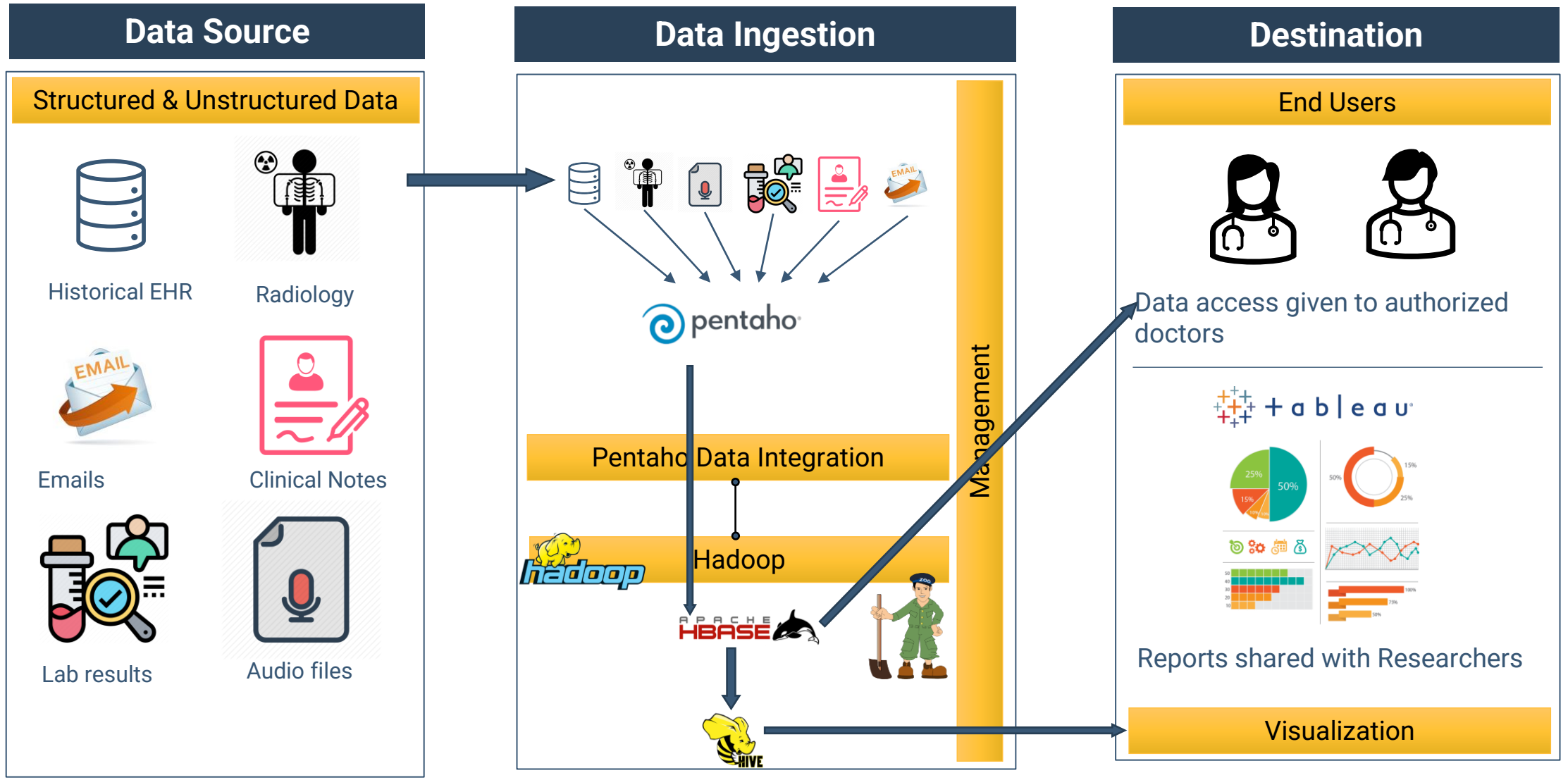
Non-Functional Requirements

Security

- The access permissions for system data may only be changed by the system's data administrator
- Only authorized people should get access to the data
- All the data coming in must be backed up and be held in a secure system

OVERALL STRATEGY & ARCHITECTURE

Vision Diagram



Hadoop

- Hadoop can be integrated with Pentaho
- Hadoop's scalability –Hadooponomics – Ability to scale data storage and processing in a linear fashion. Relies on Horizontal/Linear scaling
- Scalable - Store and distribute very large data sets across hundreds of inexpensive servers that operate in parallel.
- Cost effective - Scale-out architecture that can affordably store all data
- Flexible - Access new data sources and tap into different types of data (both structured and unstructured) to generate value from that data
- Fast - Storage method is based on a distributed file system that basically 'maps' data wherever it is located on a cluster
- Resilient to failure - Data is sent to an individual node, that data is also replicated to other nodes in the cluster, which means that in the event of failure, there is another copy available for use

Pentaho

- Pentaho provides the extract, transform and load functionality that is necessary for the EHR data that is stored in various forms
- Pentaho has about 35 optimized data connectors such as SAP HANA connector, Microsoft Excel, SQL Lite, SQL Server, and Salesforce.com etc. which enables connecting to the data without any addon or third-party connectors.
- The installation is fast and easy.
- Connects to almost any data source as input & output
- The data integration to handle the large flux of the data would be “Pentaho Data Integration”
- Pentaho Kettle does not require the data to be stored in single data base or single source, it can work the same way even if the data is spread across.
- When attempting to process the vast amounts of data collected on a daily basis for reporting, it is critical to have a Data Integration solution that is not only easy to use but easily extendable.

Hbase & Hive

HBase

- A distributed column-oriented database built on top of the Hadoop file system
- It provides random real-time read/write access to data in the Hadoop File System
- HBase is a column-oriented database and the tables in it are sorted by row
- Scalability- HBase can scale limitlessly as load and performance demands increase simply by adding server nodes

Apache Hive

- Data warehouse software project built on top of Apache Hadoop for providing data summarization, query, and analysis.
- SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop.
- Command Line Interface and UI (User Interface) allow an external user to interact with Hive by submitting queries. Provide more user defined interfaces to extend

Tableau

- Business Intelligence tool for visually analyzing the data.
- Create and distribute interactive and shareable dashboards which depict the trends, variations and density of the data in form of graphs and charts.
- Connect to files, relational and Big data sources to acquire and process data.
- Data blending and real time collaboration of huge data sets.
- Visually representing data which was analyzed and will give a fast and easy to use for researchers.

VELERO

Complete project plan

Heatmap	PLC	Order	Type	Milestone/Task Description	%Complete	Est Hours	Est HtC	Assigned To	Start Date	End Date	Status	HoursPosted
Complete	1-Initiation	1	Analysis	Project Initiation	100.00%	30.00	0.00	Navaneeta, Naik	03/05/2021	03/08/2021	Complete	0.00
Complete	1-Initiation	2	Analysis	Architecture Design	100.00%	15.00	0.00	Navaneeta, Naik	03/08/2021	03/10/2021	Complete	0.00
Complete	1-Initiation	2	Analysis	Functional Requirements	100.00%	35.00	0.00	Navaneeta, Naik	03/11/2021	03/15/2021	Complete	0.00
Complete	1-Initiation	3	Analysis	Non-Functional Requirements	100.00%	30.00	0.00	Navaneeta, Naik	03/15/2021	03/18/2021	Complete	0.00
Complete	2-Planning	1	Analysis	Evaluate HIPAA Compliance	100.00%	40.00	0.00	Navaneeta, Naik	03/15/2021	03/22/2021	Complete	0.00
Green	2-Planning	1	Analysis	Project Planning	0.00%	600.00	600.00	Navaneeta, Naik	03/19/2021	06/01/2021	Inprocess	0.00
Green	2-Planning	3	Analysis	Talent Acquisition	0.00%	200.00	200.00	Navaneeta, Naik	06/01/2021	06/30/2021	Open	0.00
Green	3-Execution	1	Analysis	Technical Architecture	0.00%	50.00	50.00	Navaneeta, Naik	07/01/2021	07/07/2021	Open	0.00
Green	3-Execution	1	Analysis	Database & Framework Selection	0.00%	32.00	32.00	Navaneeta, Naik	07/07/2021	07/10/2021	Open	0.00
Green	3-Execution	3	Milestone	Vendor Selection & Order equipment	0.00%	40.00	40.00	Navaneeta, Naik	07/11/2021	07/15/2021	Open	0.00
Green	3-Execution	4	Milestone	Receive Equipment	0.00%	80.00	80.00	Navaneeta, Naik	07/15/2021	07/27/2021	Open	0.00
Green	3-Execution	5	QA	Write Master Test Plan	0.00%	40.00	40.00	Navaneeta, Naik	07/27/2021	08/02/2021	Open	0.00
Green	3-Execution	6	QA	Write Test Cases	0.00%	40.00	40.00	Navaneeta, Naik	08/02/2021	08/07/2021	Open	0.00
Green	3-Execution	6	Not Defined	Setup Test environment	0.00%	24.00	24.00	Navaneeta, Naik	08/09/2021	08/12/2021	Open	0.00

Complete project plan

	Heatmap	PLC	Order	Type	Milestone/Task Description	%Complete	Est Hours	Est HtC	Assigned To	Start Date	End Date	Status	HoursPosted
	Green	3-Execution	6	Development	Data Collection	0.00%	500.00	500.00	Navaneeta, Naik	08/12/2021	10/20/2021	Open	0.00
	Green	3-Execution	7	Development	Transformation & Load Process	0.00%	500.00	500.00	Navaneeta, Naik	08/20/2021	10/25/2021	Open	0.00
	Green	3-Execution	9	QA	System Testing	0.00%	50.00	50.00	Navaneeta, Naik	10/25/2021	11/10/2021	Open	0.00
	Green	3-Execution	10	QA	User Acceptance Testing	0.00%	120.00	120.00	Navaneeta, Naik	11/10/2021	11/30/2021	Open	0.00
	Green	4-Controlling	1	Other	Project Managment & Status Reporting	0.00%	200.00	200.00	Navaneeta, Naik	03/05/2021	12/10/2021	Open	0.00
	Green	5-Closing	1	Not Defined	Training	0.00%	55.00	55.00	Navaneeta, Naik	12/01/2021	12/11/2021	Open	0.00
	Green	5-Closing	2	Systems	Implementation	0.00%	40.00	40.00	Navaneeta, Naik	12/11/2021	12/15/2021	Open	0.00
	Green	5-Closing	3	Systems	Post Implementation Support	0.00%	25.00	25.00	Navaneeta, Naik	12/16/2021	12/20/2021	Open	0.00
	Green	5-Closing	4	Other	Closing	0.00%	8.00	8.00	Navaneeta, Naik	12/20/2021	12/20/2021	Open	0.00

Issues & Risks

		Type	Exposure	Date	Assign Date	Expected Date	Description	Date Closed	Resource	Updated	By
		Issue	3	06/01/2021	06/01/2021	06/01/2021	Employ the right talent to do the tasks which will ensure correct data integration and analysis+		Navaneeta, Naik	30/04/2021	Navaneeta, Naik
		Issue	4	07/01/2021	07/01/2021	07/01/2021	Data loss can be a major issue when handling large data. Ensuring there is no data loss is an aspect+		Navaneeta, Naik	30/04/2021	Navaneeta, Naik
		Issue	6	05/01/2021	05/01/2021	05/01/2021	The system and use of data in this project should adhere to the HIPAA guidelines at all stages of th+		Navaneeta, Naik	30/04/2021	Navaneeta, Naik
		Issue	7	12/01/2021	12/01/2021	12/01/2021	Data collection should ensure every possible data is included. Missing data can be a major issue+		Navaneeta, Naik	30/04/2021	Navaneeta, Naik
		Risk	6	08/01/2021	08/01/2021	08/01/2021	The data is sensitive and any unauthorized access to this data will be a huge risk associated with t+		Navaneeta, Naik	30/04/2021	Navaneeta, Naik
		Risk	7	09/01/2021	04/01/2021	04/01/2021	Data Breach of sensitive EHR data+		Navaneeta, Naik	30/04/2021	Navaneeta, Naik
		Risk	8	09/01/2021	09/01/2021	09/01/2021	Data should be stored in a secure system+		Navaneeta, Naik	30/04/2021	Navaneeta, Naik

Data collection

- The data source is present in the form of databases, text files and various other formats
- This data from various sources are extracted using Pentaho
- Then this data is transformed according to the business rules
- Data from Pentaho is then loaded to the Hbase database
- Hbase provides real-time read/write access to the HDFS (Hadoop File System)
- HBase is used to store this data
- Hive will be used for analytical querying of this data

Security

Hadoop:

- Security features of Hadoop consist of authentication, service level authorization, authentication for Web consoles and data confidentiality
- Data Encryption
- Hadoop has begun using Kerberos authorization support to provide security. Kerberos allows for mutual authentication between client and server

Tableau:

- Trusted Authentication -- a trusted relationship between Tableau Server and one or more web servers
- OAuth for some cloud-based service providers
- Native Authentication managed by Tableau Server Tableau provides automatic login timeouts that can be configured by administrators

Security

Zookeeper:

- ZooKeeper now supports Kerberos security
- Authorization is done via ACLs
- Supports several types of restrictions - Message digest , Hostname, IP address
- Can limit access by function - Read, write, delete, etc

Hbase:

- The security protocol for the client authentication involves Kerberos and SSL
- Database logging and security The data at rest on the under lying file system is transparently encrypted

Management: Apache Zookeeper

- Apache Zookeeper is a service used by a cluster (group of nodes) to coordinate between themselves and maintain shared data with robust synchronization techniques.
- Zookeeper is itself a distributed application providing services for writing a distributed application.
- The performance of can be enhanced by adding more machine to it with minor changes in the configuration without any downtime.
- Zookeeper is very performing management tool and very reliable compared to the other MDM tools
- Benefits- fast, reliable, simple, ordered

THANK YOU