Analytics for Hospitals Health-Care Data

MENTOR :

Dr.ARUNACHALAM

TEAM MEMBERS:

DHARINEESH S V

ABISHEK R

AKASH J

GURU PRASANTH J

1. INTRODUCTION

1.1 Project Overview

PROJECT OBJECTIVES

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas to focus: Healthcare Management. While healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital.

This parameter helps hospitals to identify patients of high LOS-risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the

chance of staff/visitor infection. In addition, prior knowledge of LOS can aid in

logistics such as room and bed allocation planning.

What if I have been hired as Data Scientist of Health Man – a not for profit

organization dedicated to manage the functioning of Hospitals in a professional

and optimal manner.

The goal is to accurately predict the Length of Stay for each patient on casebycase basis so that the Hospitals can use this information for optimal resource

allocation and better functioning. The length of stay is divided into 11 different

classes ranging from 0-10 days to more than 100 days**.**

The Overall Objective:

• Know fundamental concepts and can work on IBM Cognos Analytics

• Gain a broad understanding of plotting different visualizations to provide

the suitable solution.

• Able to create meaningful Visualizations and the Dashboard(s).

1.2 Purpose :

• Creating multiple analytical graphs/charts/Visualizations.

• Using the Analytical Visualizations, build the required Dashboard(s).

• Saving and visualizing the final dashboard in the IBM Cognos Analytics.

Build the following visualizations

o Length of Stay for each case of patients.

o Stay by Patient ID using Column Chart

o Severity of illness by Patient-Id using Tree Map

o Age, Department Wise Patient using Table

o Room Availability by Pie Chart

o Dashboard Creation

o Department wise no. of admissions by Waterfall Chart

2.LITERATURE SURVEY

2.1 Existing problem

In order to discuss health data analytics and the role it plays in the health care sector, we must first understand the data that is being collected and analyzed. There is data being collected on the processes and procedures of the business side of health care , but there is also an enormous amount of health data being gathered, stored and analyzed.

Health data is any data relating to the health of an individual patient or collective population. This information is gathered from a series of health information systems (HIS) and other technological tools utilized by health care professionals, insurance companies and government

organizations. We are able to see a holistic view of each individual patient as well as trends tied to location, socioeconomic status, race and predisposition. The information being collected can be broken down into specific datasets that can then be analyzed.

There are a variety of tools and systems used to collect, store, share and analyze health data gathered through various means. These tools include:

1. Electronic Health Records (EHRs)
2. Personal Health Records (PHRs)
3. Electronic Prescription Services (E-prescribing)
4. Patient Portals
5. Master Patient Indexes (MPI)
6. Health-Related Smart Phone Apps and more

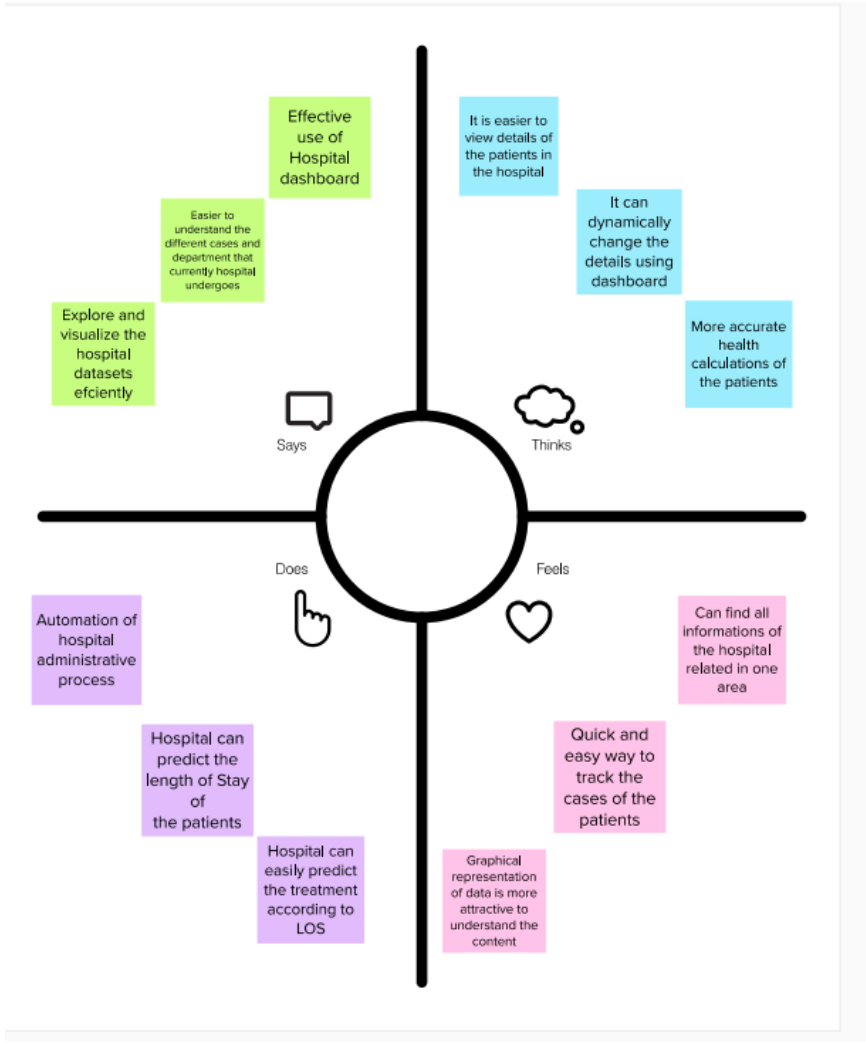
2.2 References

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4341817/
- Raghupathi W. Data Mining in Health Care. In: Kudyba S, editor. *Healthcare Informatics: Improving Efficiency and Productivity.* 2010. pp. 211–223. [Google Scholar]
- 2. Burghard C. *Big Data and Analytics Key to Accountable Care Success.* 2012. [Google Scholar]
- 3. Dembosky A. *"Data Prescription for Better Healthcare." Financial Times, December 12, 2012, p. 19.* 2012.

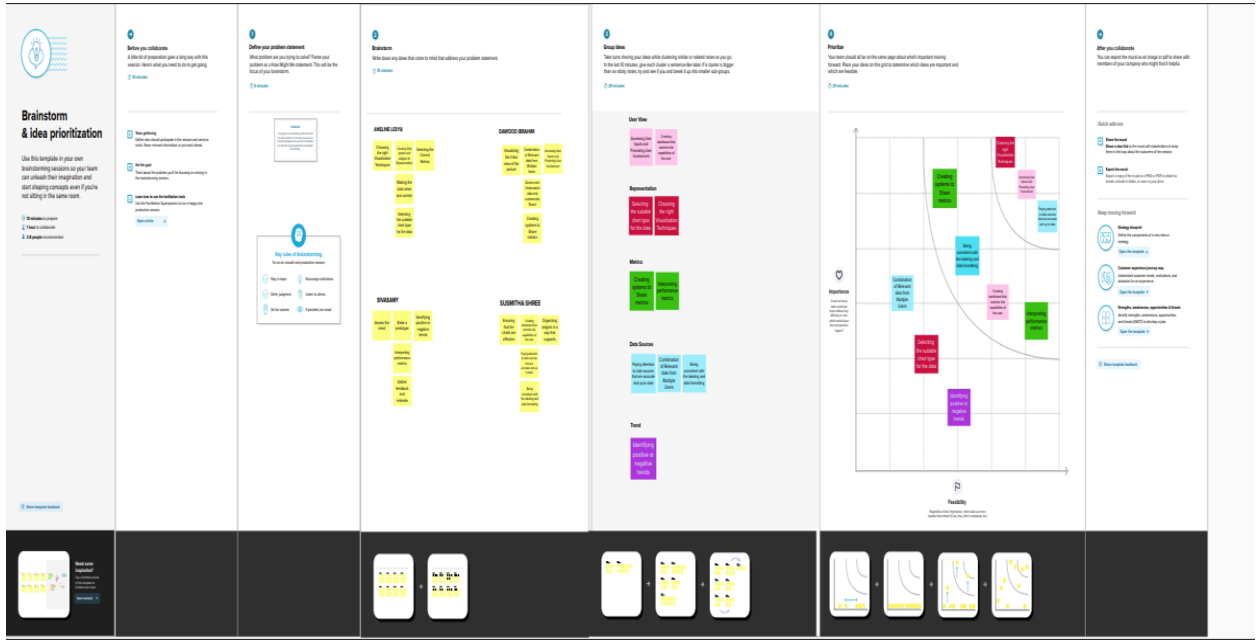
2.3 Problem Statement Definition:

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas to focus: Healthcare Management. While healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital. This parameter helps hospitals to identify patients of high LOS-risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning .Suppose you have been hired as Data Scientist of Health Man – a not for profit organization dedicated to manage the functioning of Hospitals in a professional and optimal manner**.**

3. IDEATION & PROPOSED SOLUTION


Effective use of Hospital dashboard
Easier to understand the different cases and department that currently hospital undergoes
Explore and visualize the hospital datasets efciently
Says
It is easier to view details of the patients in the hospital
It can dynamically change the details using dashboard
More accurate health calculations of the patients
Thinks
Does
Automation of hospital administrative process
Hospital can predict the length of Stay of the patients
Hospital can easily predict the treatment according to LOS
Feels
Can find all informations of the hospital related in one area
Quick and easy way to track the cases of the patients
Graphical representation of data is more attractive to understand the content

3.2 Ideation & Brainstorming



3.3 Proposed Solution

Proposed Solution Template:

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To Predict the Length of the stay for each patient on case by case basis so that the hospital can use the information for optimal resource allocation and better functioning.
2.	Idea / Solution description	To create a dashboard for monitoring the length of stay of patients and also include the necessary data which were provided by the hospital to create an effective virtual dashboard using Cognos analytics.
3.	Novelty / Uniqueness	By Using Data Visualization techniques, the dashboard supports clinicians and hospital managers in viewing and exploring data on processes and outcomes of care in an interactive manner.
4.	Social Impact / Customer Satisfaction	The hospitals can use this dashboard to view their daily hospital records and they can update the given inputs and get the expected output effectively.
5.	Business Model (Revenue Model)	While using this dashboard the hospitals can easily get regular updates on the patients and this was widely applicable in all departments of the hospitals. The Hospital staff can easily login into the dashboard and view the risk rate of the patients according to the length of stay in the hospital and can give proper treatment according to that.
6.	Scalability of the Solution	The scalability of this project was a. The hospitals can view the length of stay of the patient case by case basis b. The patient id, department name, other hospital-related details, etc. c. It also helped to visualize an interactive dashboard efficiently.

3.4 Problem Solution fit

Many different types and forms of data are combined when analytics techniques are applied in healthcare settings, medical research, and public health departments. Health-related data sources include electronic health records (EHRs), genomics and post-genomics, bioinformatics, medical imaging, sensor informatics, medical informatics, and health informatics. The analysis entails many discrete steps:

1. Integrate heterogeneous data types.
2. Ensure the quality of the data upon reception and throughout the analysis.
3. Create data models.
4. Interpret the results of the analysis.
5. Validate the analysis results.

The benefits of data analytics in healthcare are realized in seven distinct areas:

- Support clinical treatment decisions from physicians and other health professionals.
- Improve the accuracy and speed of identifying patients at highest risk of disease.
- Provide greater detail in the EHRs of individual patients.
- Make the provision of healthcare more efficient, which reduces costs.
- Promote preventive measures by giving patients greater insight into their health and treatment goals.
- Integrate data from consumer fitness devices and other patient-provided sources of health data.
- Deliver real-time alerts to healthcare providers by analyzing health data at the collection point.

4.REQUIREMENT ANALYSIS

4.1 Functional requirement

There are a lot of software requirements specifications included in the functional requirements of the Hospital Management System, which contains various processes, namely Registration, Check out, Report Generation, and Database.
Registration Process of SRS (Software Requirements Specification)

● Adding Patients: The Hospital Management enables the staff at the front desk to include new patients in the system.

● Assigning an ID to the patients: The HMS enables the staff at the front desk to provide a unique ID for each patient and then add them to the record sheet of the patient. The patients can utilize the ID throughout their hospital stay.

4.2 Non- Functional requirement

There are a lot of software requirements specifications included in the non-functional requirements of the Hospital Management System, which contains various processes, namely Security, Performance, Maintainability, and Reliability.

Security:

● Patient Identification: The system needs the patient to recognize herself or himself using the phone.

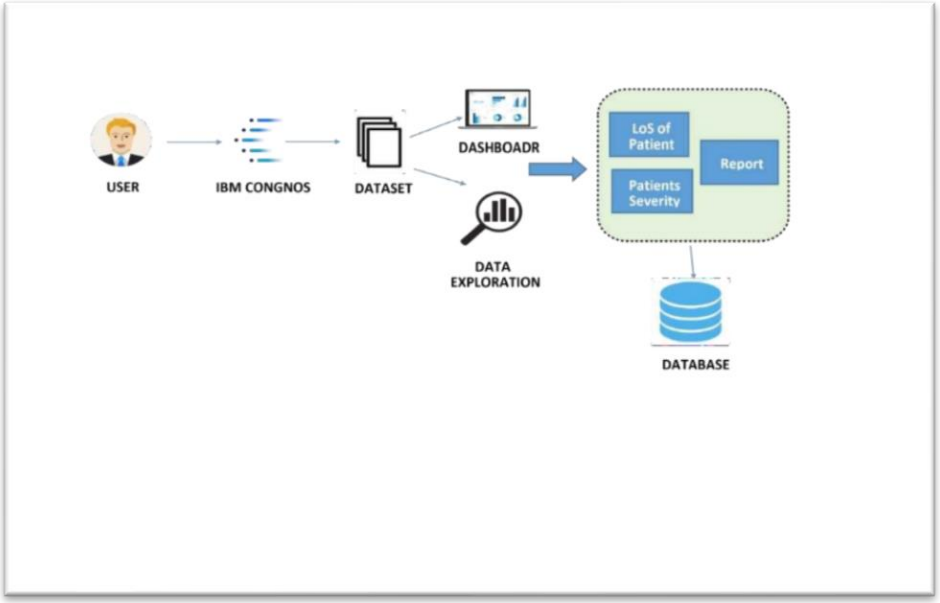
● Logon ID: Any users who make use of the system need to hold a Logon ID and password.

● Modifications: Any modifications like insert, delete, update, etc. for the database can be synchronized quickly and executed only by the ward administrator.

● Front Desk Staff Rights: The staff at the front desk can view any data in the Hospital Management system, and add new patients record to the HMS but they don't have any rights to alter any data in it.

5. PROJECT DESIGN

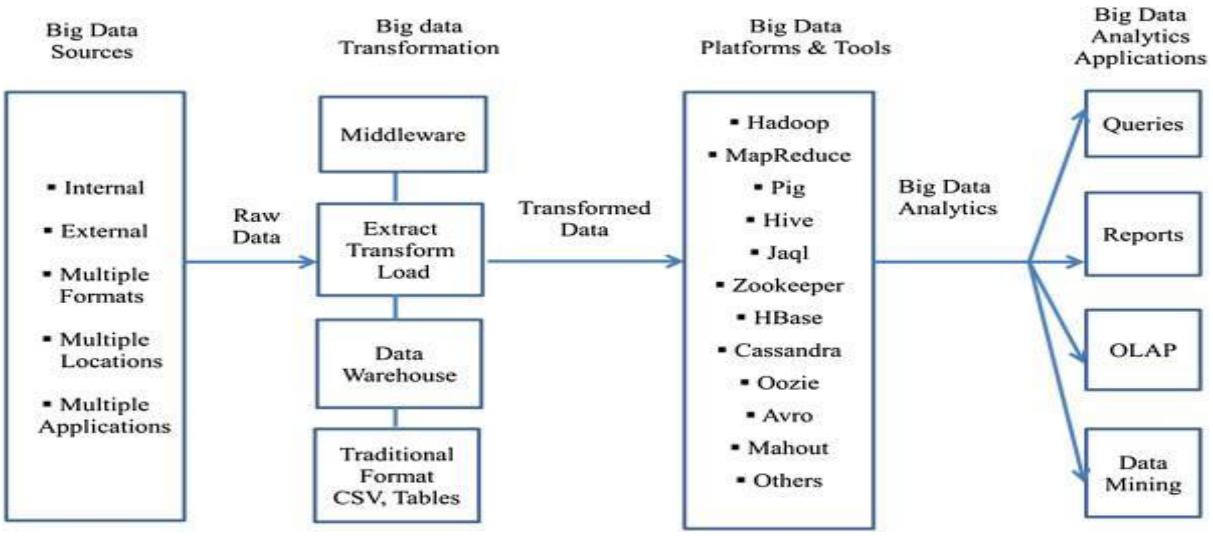
5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture

The conceptual framework for a big data analytics project in healthcare is similar to that of a traditional health informatics or analytics project. The key difference lies in how processing is executed. In a regular health analytics project, the analysis can be performed with a business intelligence tool installed on a stand-alone system, such as a desktop or laptop. Because big data is by definition large, processing is broken down and executed across multiple nodes. The concept of distributed processing has existed for decades. What is relatively new is its use in analyzing very large data sets as healthcare providers start to tap into their large data repositories to gain insight for making better-informed health-related decisions. Furthermore, open source platforms such as Hadoop/MapReduce, available on the cloud, have encouraged the application of big data analytics in healthcare.

While the algorithms and models are similar, the user interfaces of traditional analytics tools and those used for big data are entirely different; traditional health analytics tools have become very user friendly and transparent. Big data analytics tools, on the other hand, are extremely complex, programming intensive, and require the application of a variety of skills. They have emerged in an ad hoc fashion mostly as open-source development tools and platforms, and therefore they lack the support and user-friendliness that vendor-driven proprietary tools possess. As Figure 1 indicates, the complexity begins with the data itself.



5.3 User Stories

User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the dashboard by entering my email, and password, and confirming my password.	I can access my account in the dashboard	High	Sprint-1
		USN-2	As a user, I will receive a confirmation email once I have registered for the dashboard	I can receive a confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the dashboard through Social Media	I can register & access the dashboard with Social Media Login	Low	Sprint-2
		USN-4	As a user, I can register for the dashboard through Gmail	I can register and access dashboard with Gmail	Medium	Sprint-2
	Login	USN-5	As a user, I can log into the application by entering email & password	I can login to the account in my email login.	High	Sprint-2
	Dashboard	USN-6	As a user ,I can use my account in my dashboard for uploading dataset.	I can login to the account for uploading dataset.	Medium	Sprint-3
Customer (Web user)	Website	USN-7	As a user ,I can use my dashboard in website	I can login into the dashboard by visiting website.	Medium	Sprint-3
Customer Care Executive		USN-8	As a user ,I can contact Customer care Executive for my login.	I can contact customer executive for my login.	High	Sprint-4
Administrator		USN-9	As a user ,I can contact administrator for my queries.	I can contact administrator for solving my queries.	High	Sprint-4
Exploration	Dashboard	USN-10	As a user, I can prepare data by using Exploration Techniques.	I can prepare data by using Exploration Techniques.	High	Sprint-3
Presentation	Dashboard	USN-11	As a user, I can Present data in my dashboard.	I can present data by using my account in dashboard.	High	Sprint-4
Visualization	Dashboard	USN-12	As a user, I can Prepare Data by using Visualization Techniques.	I can prepare data by using Visualization Techniques.	High	Sprint-3

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	3
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	3
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2	Low	3
Sprint-2		USN-4	As a user, I can register for the application through Gmail	2	Medium	3
Sprint-2	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	3
Sprint-3	Dashboard	USN-6	As a user ,I can use my account in my dashboard for uploading dataset.	5	Medium	4
Sprint-3	Website	USN-7	As a user ,I can use my dashboard in website	3	Medium	4
Sprint-4	Dashboard	USN-8	As a user ,I can contact Customer care Executive for my login.	5	High	4

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-4	Dashboard	USN-9	As a user ,I can contact administrator for myqueries.	5	High	4
Sprint-3	Dashboard	USN-10	As a user, I can prepare data by using Exploration Techniques.	5	High	4
Sprint-4	Dashboard	USN-11	As a user, I can Present data in my dashboard.	5	High	4
Sprint-3	Dashboard	USN-12	As a user, I can Prepare Data by using Visualization Techniques.	5	High	4

Project Tracker, Velocity & Burndown Chart: (4 Marks)

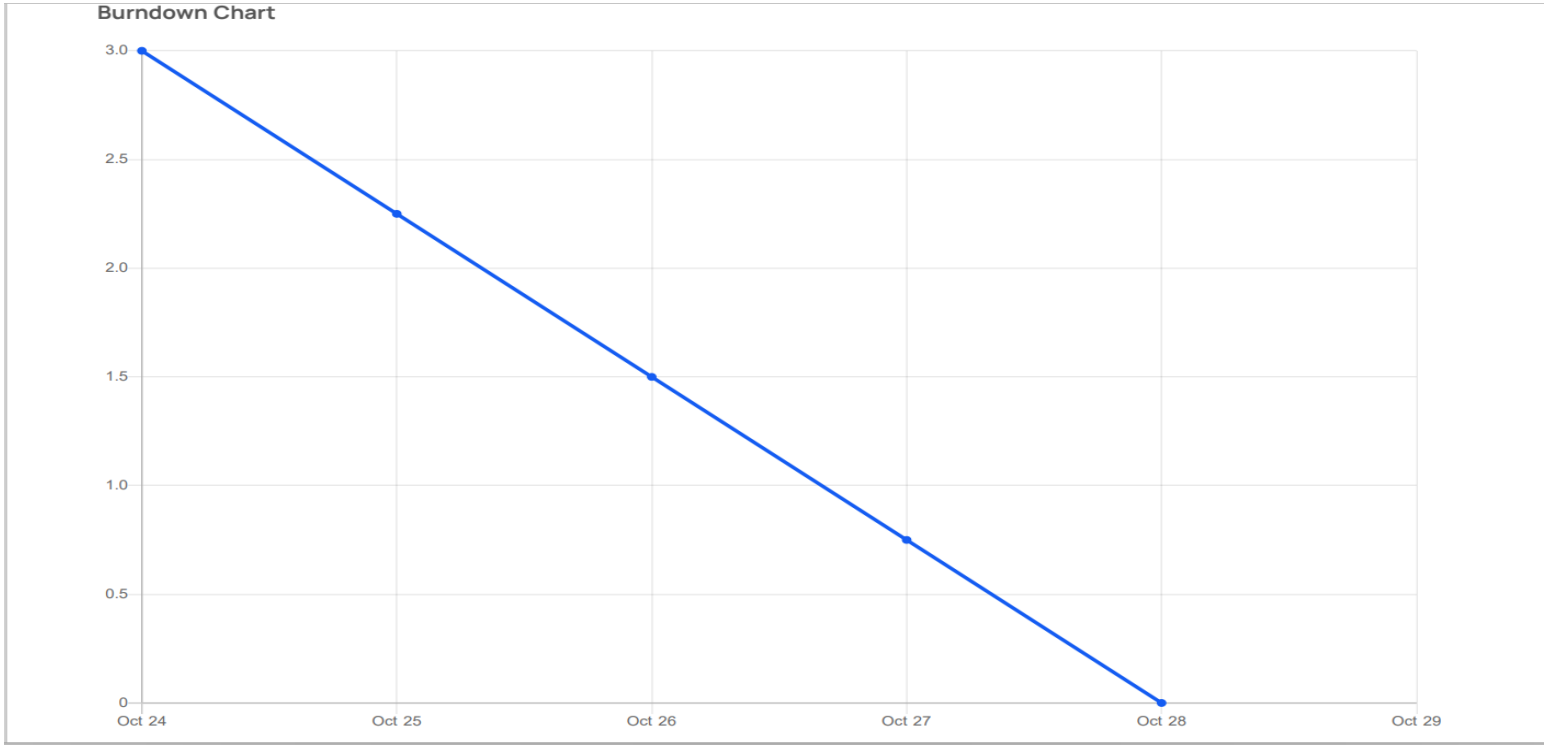
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	3	6 Days	24 Oct 2022	29 Oct 2022	3	29 Oct 2022
Sprint-2	5	6 Days	31 Oct 2022	05 Nov 2022	5	05 Nov 2022
Sprint-3	18	6 Days	07 Nov 2022	12 Nov 2022	18	12 Nov 2022
Sprint-4	15	6 Days	14 Nov 2022	19 Nov 2022	15	19 Nov 2022

Velocity:

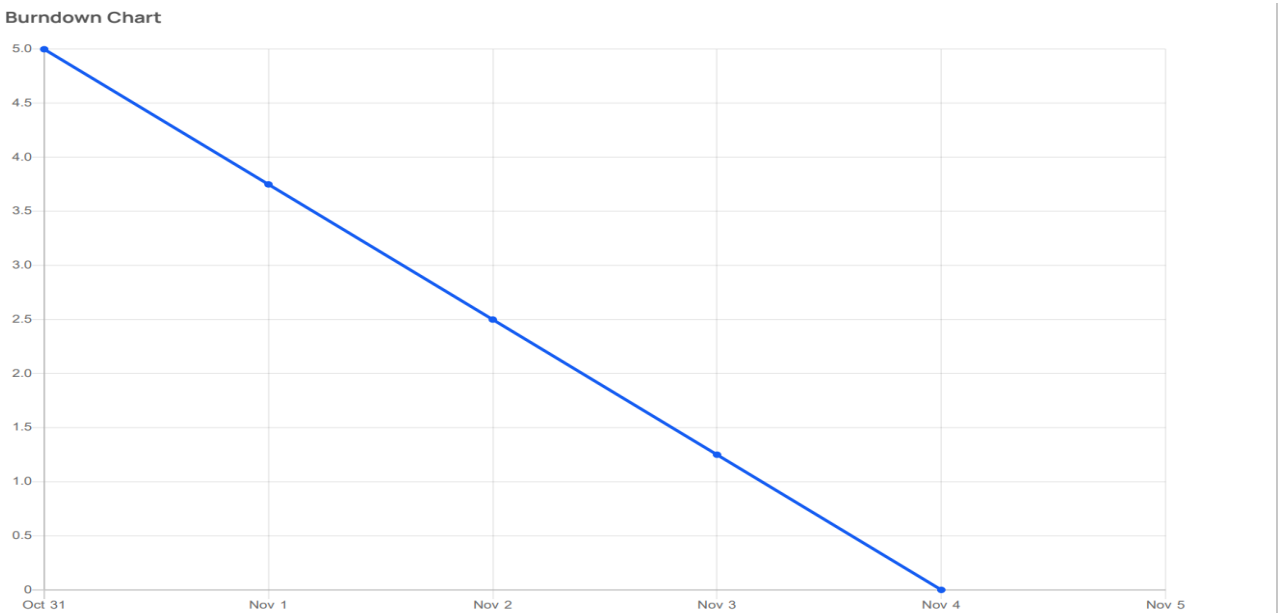
Sprints	Sprint Duration	Velocity	Actual Velocity
Sprint-1	6	3	2
Sprint-2	6	5	1.2
Sprint-3	6	18	0.35
Sprint-4	6	15	0.4

Velocity – Actual Velocity

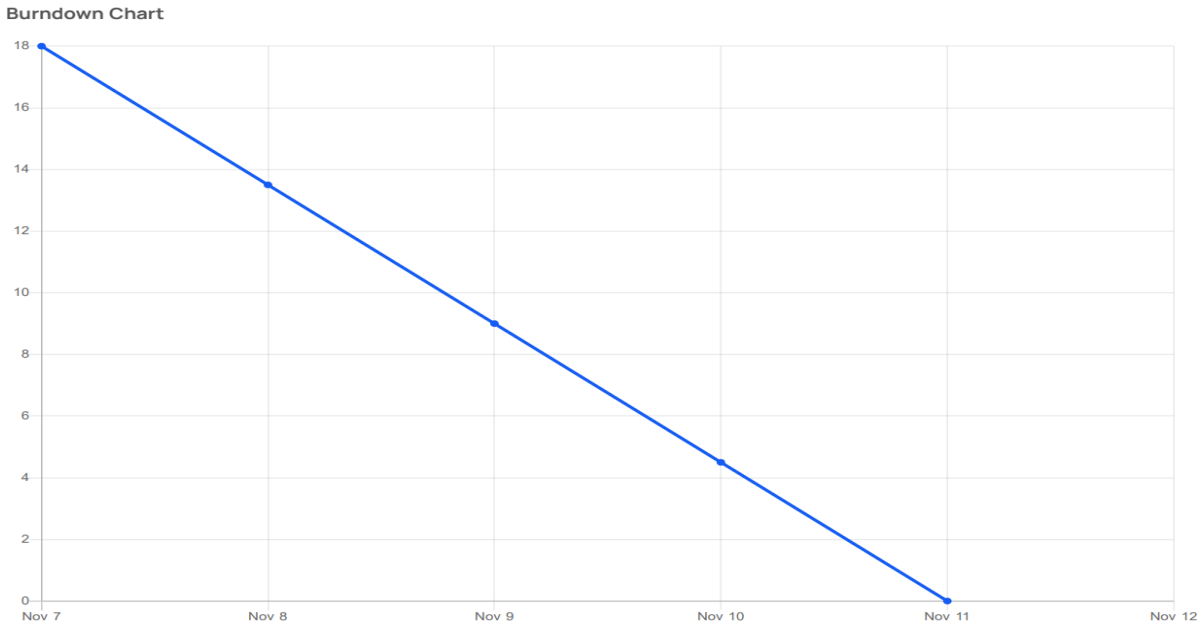
Burndown Chart:



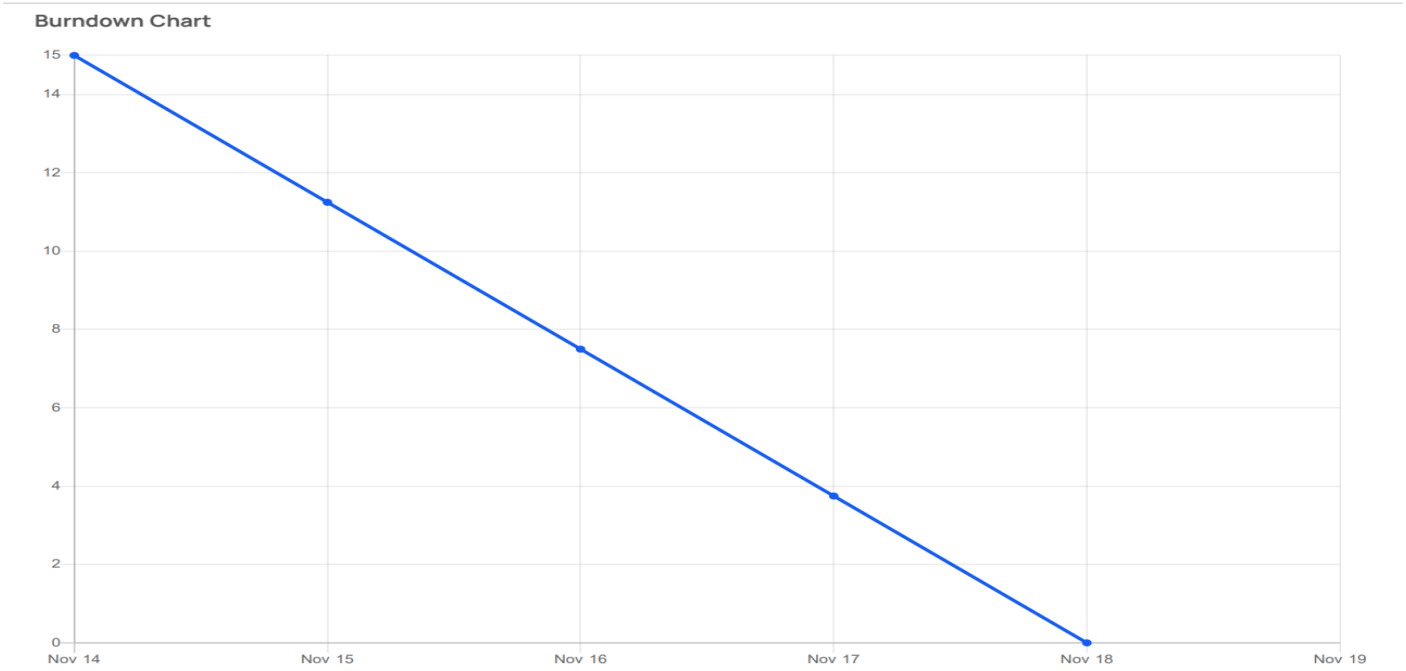
SPRINT -1


Burndown Chart
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
Oct 31
Nov 1
Nov 2
Nov 3
Nov 4
Nov 5

SPRINT- 2

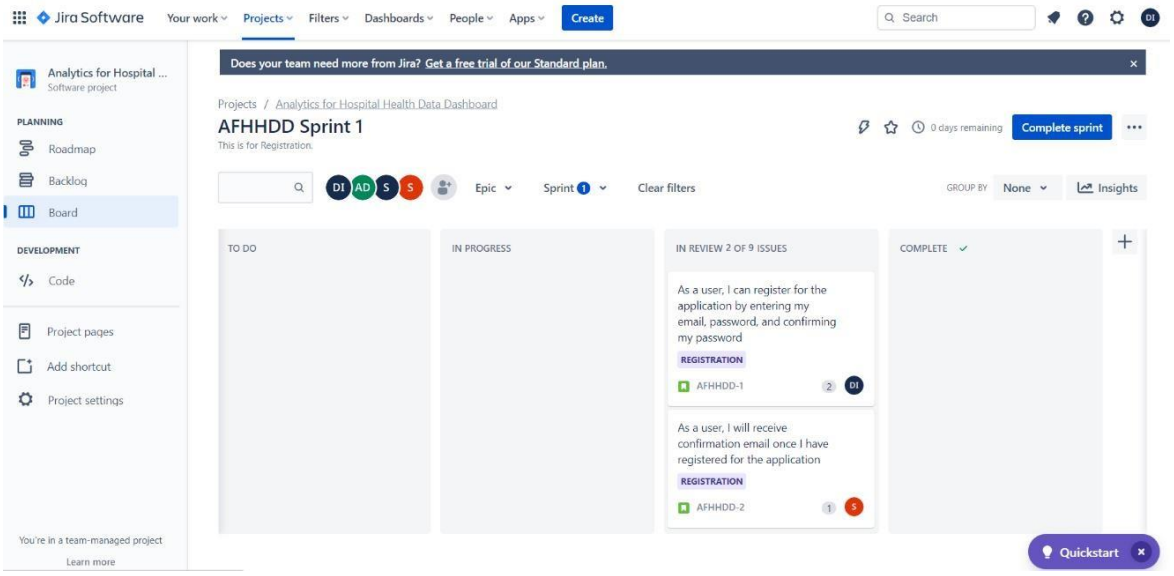

Burndown Chart
18
16
14
12
10
8
6
4
2
0
Nov 7
Nov 8
Nov 9
Nov 10
Nov 11
Nov 12

SPRINT- 3

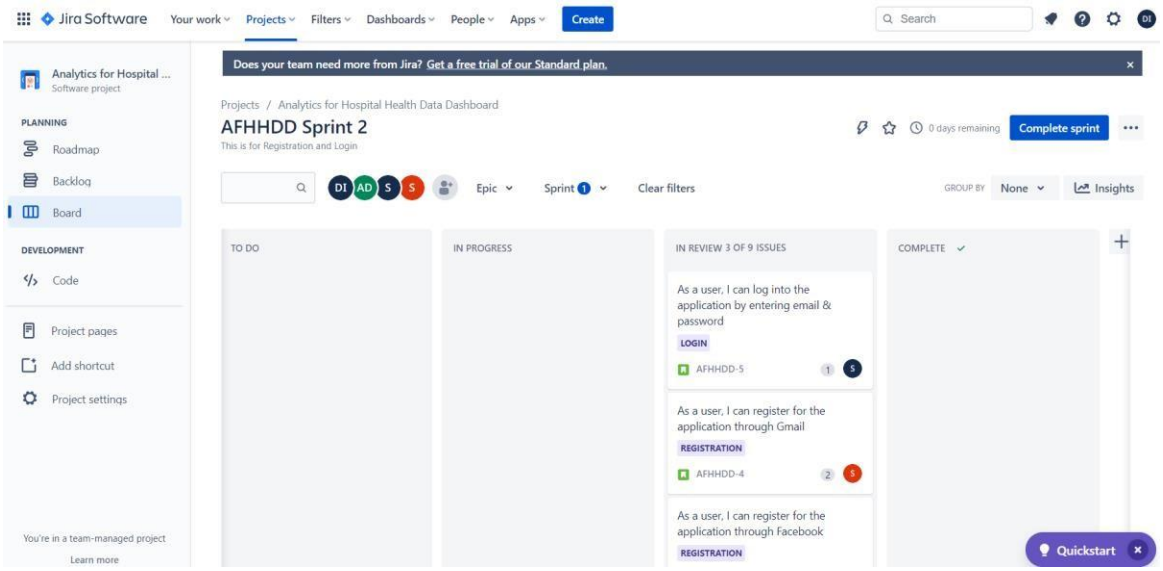

Burndown Chart
15
14
12
10
8
6
4
2
0
Nov 14
Nov 15
Nov 16
Nov 17
Nov 18
Nov 19

Reports from JIRA

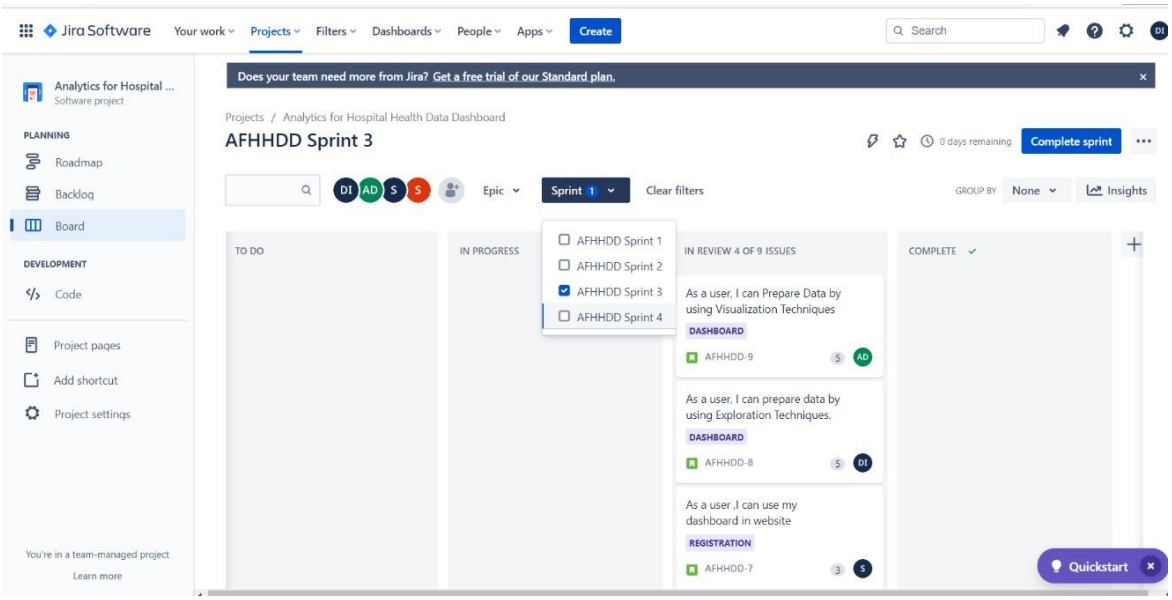
In Review



Sprint 1

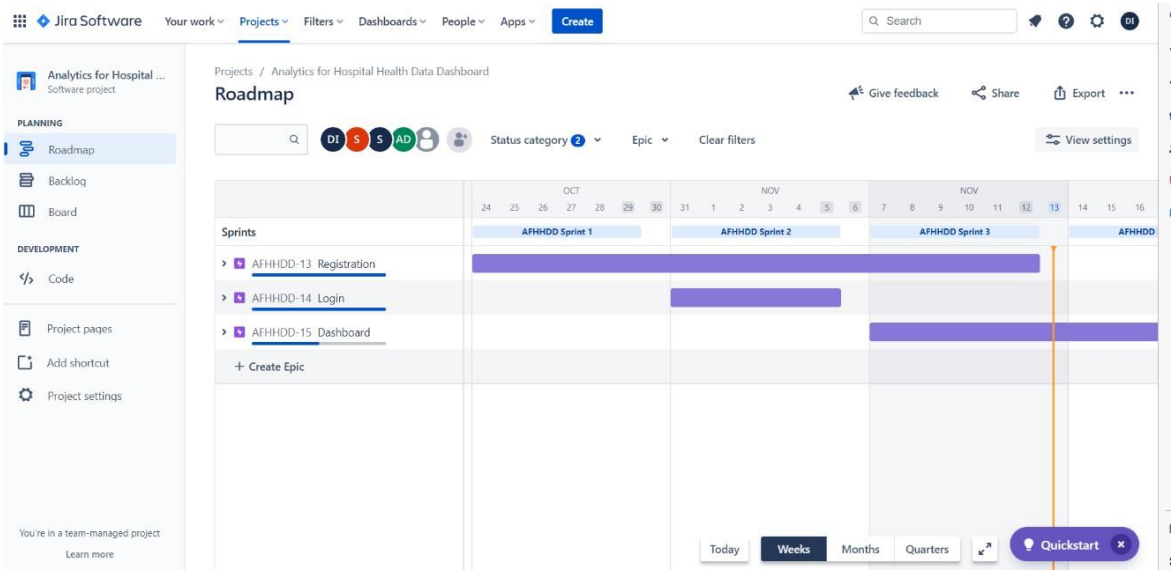


Sprint 2



Sprint 3

On Progress



RoadMap

7. CODING & SOLUTIONING

CODE LINK:

https://colab.research.google.com/drive/183bKFA31hWRiGwsTDJxd7rv0Xyiw9Tww?usp=sharing

DATA SET LINK :

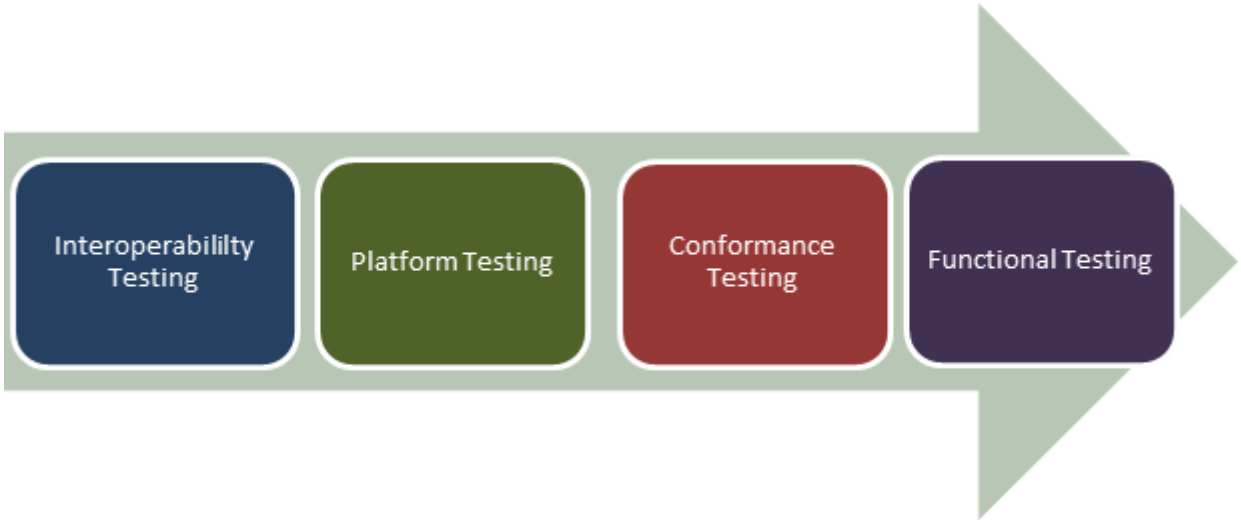
https://drive.google.com/file/d/1slC0MhsJHeuODVkhIXrdNOX_aBDySSBh/view

8. TESTING

Before preparing test scenarios certain requirement of the system should be considered. For example, health-care providers (Doctors/Hospitals) provide care 24/7, so the patient check-in software needs to be available at all times. Also, it needs to communicate with insurance companies to validate policy information, send claims and receive remittances. Here, the architecture should define the different components of the system, the protocol to communicate with insurance companies, and how to deploy the system so that it complies 24/7.

As a tester, you need to ensure that the healthcare software system meets the desired load/performance benchmark.

Other Testing Types for Healthcare Application



- **Functional Testing**: Testing healthcare application against functional capabilities
- **Conformance Testing**: Conformance test Healthcare security requisites and industry frameworks
- **Platform Testing**: Testing of applications on Mobile platform and applications testing for cross-browser compatibility
- **Interoperability Testing**: Testing conformance to interoperability standards (Eg; DICOM, HL7, CCD/CDA)

Testing Challenges in Healthcare Application

Testing challenges in testing healthcare application are no different than other web application testing.

- Requires expertise in testing, and usually, it is high in cost
- Requires interoperability, compliance, regulatory, security, safety testing besides regular testing techniques (Non-Functional, Functional and Integration testing)
- Testing should be done keeping in mind the safety and regulatory standards- as any error can cause a direct effect on patient's life
- Testing team needs to be well aware of the various functionalities, clinical usage, and the environment the software will be used for
- A health-care product should comply with various standards like FDA, ISO, and CMMI before it can be used
- Cross dependency of software- testers need to ensure that any changes in one component or layer should not lead to side effect on the other.

Healthcare device Testing

While health-care device software is not the direct concern of patient, they also require rigorous testing like another software testing. For example, X-ray machines that are controlled by software programs should be tested well because any testing error in software can lead to a serious effect on the patient.

FDA (Food and Drug Administration) has guidelines for mobile and web applications for medical devices. While testing medical devices the proper functional Test Plan along with pass and fail criteria is also the part of FDA guidelines. When a test plan is executed, the results are collected and reported to FDA. This process ensures that the device meets the standard of the regulatory bodies.

Useful tips for Healthcare Testing

While testing software, you can consider some important tips for the testing healthcare system.

- Dates are important and need to be accurate

- While designing test cases consider various parameters like different types of plan, brokers, members, commission, etc.
- Complete knowledge of the domain is required

9. RESULTS

Health Data Analysts take the health care data being collected and use skills such as data acquisition, data management, data analysis and data interpretation to provide actionable insights. The role of big data in health care and the increasing need for improvement in the health care sector has led to a higher demand for qualified health data analysts.

The role of a health data analyst varies based on their position and industry of choice. Regardless of industry, a health data analyst will need to be able to work with, develop and evaluate health information technology (health IT) and other health information systems (HIS).

They may also be expected to:

- Collect or mine data
- Examine current and historical data
- Evaluate raw data
- Build predictive models
- Automate reports

10. ADVANTAGES & DISADVANTAGES

Big data could reduce the recency bias or recency effect bias. Recency bias occurs when the recent events are weigh more heavily than earlier events in order to improve the situation, but it may lead to incorrect decisions . The real time information can also be incorporated into big data. Real time big data has many advantages. For example, any errors or trouble shoot in an organization can be identified immediately and the operational problem can be overcome. This will save time, cost and increase the productivity. The services also can be further improved as the real time provides the latest information on the subject matter. For instance, it will provide the complete information on the patients and at the same time able to administer medical intervention without any delay . In healthcare, big data is also used in predictive analysis which is to identify and address the medical issues before it becoming an unmanageable problem. Healthcare professionals are able to reduce the risk and overcome the issue with the information derived from the big data . Apart from that, big data is also able to help identify frauds in healthcare especially on insurance claims. Fraudulent, inconsistency and false claims can be flagged. This will facilitate insurance companies to prevent losses . Big data can also benefit healthcare through data management, electronic medical records and data analysis. The big data will help to find and identify the right population or target group. Big data consists of diverse group of population and certain group can be identified for risk assessment and screenings. The existence of big data will also allow development or modification of a program or intervention to target the health problem It will enables clinical trials to be initiated immediately. Big data will give a clearer picture on the type of population as well as their medical problem. The pattern of the distribution or disease information will allow quick development of intervention program as well as targeting the affected group as early as possible . Data growth of pharmaceutical industries were derived from patients, caregivers, retailers and Research and Development (R&D). Big data could facilitate the pharmaceutical companies to identify new potential and effective drugs and deliver it to the users more quickly . Issues with Big Data There is a huge challenge in big data in terms of data protection, collection and sharing of health data and data usage . Big data analytics with the use of sophisticated technologies has the potential to transform the data repositories and make informed decisions. Issues such as privacy, security, standards and governance to be addressed . Information such as nanoparticulate therapy on cancer

treatment could be also be incorporated in big data to provide an overview and best treatment for cancer especially when nanotechnology is important in drug delivery in cancer treatment.

11 & 12 Conclusions and future prospects

Nowadays, various biomedical and healthcare tools such as genomics, mobile biometric sensors, and smartphone apps generate a big amount of data. Therefore, it is mandatory for us to know about and assess that can be achieved using this data. For example, the analysis of such data can provide further insights in terms of procedural, technical, medical and other types of improvements in healthcare. After a review of these healthcare procedures, it appears that the full potential of patient-specific medical specialty or personalized medicine is under way. The collective big data analysis of EHRs, EMRs and other medical data is continuously helping build a better prognostic framework. The companies providing service for healthcare analytics and clinical transformation are indeed contributing towards better and effective outcome. Common goals of these companies include reducing cost of analytics, developing effective Clinical Decision Support (CDS) systems, providing platforms for better treatment strategies, and identifying and preventing fraud associated with big data. Though, almost all of them face challenges on federal issues like how private data is handled, shared and kept safe. The combined pool of data from healthcare organizations and biomedical researchers have resulted in a better outlook, determination, and treatment of various diseases. This has also helped in building a better and healthier personalized healthcare framework. Modern healthcare fraternity has realized the potential of big data and therefore, have implemented big data analytics in healthcare and clinical practices. Supercomputers to quantum computers are helping in extracting meaningful information from big data in dramatically reduced time periods. With high hopes of extracting new and actionable knowledge that can improve the present status of healthcare services, researchers are plunging into biomedical big data despite the infrastructure challenges. Clinical trials, analysis of pharmacy and insurance claims together, discovery of biomarkers is a part of a novel and creative way to analyze healthcare big data.

Big data analytics leverage the gap within structured and unstructured data sources. The shift to an integrated data environment is a well-known hurdle to overcome. Interesting enough, the principle of big data heavily relies on the idea of the more the information, the more insights one can gain from this information and can make predictions for future events. It is rightfully projected by various reliable consulting firms and health care companies that the big data healthcare market is poised to grow at an exponential rate. However, in a short span we have witnessed a spectrum of analytics currently in use that have shown significant impacts on the decision making and performance of healthcare industry. The exponential growth of medical data from various domains has forced computational experts to design innovative strategies to analyze and interpret such enormous amount of data within a given timeframe. The integration of computational systems for signal processing from both research and practicing medical professionals has witnessed growth. Thus, developing a detailed model of a human body by combining physiological data and "-omics" techniques can be the next big target. This unique idea can enhance our knowledge of disease conditions and possibly help in the development of novel diagnostic tools. The continuous rise in available genomic data including inherent hidden errors from experiment and analytical practices need further attention. However, there are opportunities in each step of this extensive process to introduce systemic improvements within the healthcare research.

High volume of medical data collected across heterogeneous platforms has put a challenge to data scientists for careful integration and implementation. It is therefore suggested that revolution in healthcare is further needed to group together bioinformatics, health informatics and analytics to promote personalized and more effective treatments. Furthermore, new strategies and technologies should be developed to understand the nature (structured, semi-structured, unstructured), complexity (dimensions and attributes) and volume of the data to derive meaningful information. The greatest asset of big data lies in its limitless possibilities. The birth and integration of big data within the past few years has brought substantial advancements in the health care sector ranging from medical data management to drug discovery programs for complex human diseases including cancer and neurodegenerative disorders. To quote a simple example supporting the stated idea, since the late 2000′s the healthcare market has witnessed advancements in the EHR system in the context of data collection, management and usability. We believe that big data will add-on and bolster the existing pipeline of healthcare advances instead of replacing skilled manpower, subject knowledge experts and intellectuals, a notion argued by many. One can clearly see the transitions of health care market from a wider volume base to personalized or individual specific

domain. Therefore, it is essential for technologists and professionals to understand this evolving situation. In the coming year it can be projected that big data analytics will march towards a predictive system. This would mean prediction of futuristic outcomes in an individual's health state based on current or existing data (such as EHR-based and Omics-based). Similarly, it can also be presumed that structured information obtained from a certain geography might lead to generation of population health information. Taken together, big data will facilitate healthcare by introducing prediction of epidemics (in relation to population health), providing early warnings of disease conditions, and helping in the discovery of novel biomarkers and intelligent therapeutic intervention strategies for an improved quality of life.

13. APPENDIX

CODE LINK:

https://colab.research.google.com/drive/183bKFA31hWRiGwsTDJxd7rv0Xyiw9Tww?usp=sharing

GITHUB LINK:

https://github.com/IBM-EPBL/IBM-Project-16577-1659617920