

Industrial Internship Report on Health Care Data Management

Prepared by

Padala Satya Sai Likhitha

Executive Summary

This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT).

This internship was focused on a Health Care Data Management provided by UCT. We had to finish the project including the report in 6 weeks' time.

My project was the healthcare system faced challenges with managing an ever-increasing volume of patient data, complex medical records, and the need for quick access to critical information. The existing on-premises infrastructure struggled to handle the scale and security demands, leading to inefficiencies, delayed decision-making, and potential risks to patient data.

This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

I was grateful for the opportunity to participate in the USC/UCT Healthcare Data Management Program. This program is highly competitive, and I was honored to be selected. The program provided me with a valuable learning experience, and it has helped me to launch my career in healthcare data management.

TABLE OF CONTENTS

1	Preface	3
2	Introduction	4
2.1	About UniConverge Technologies Pvt Ltd	4
2.2	About upskill Campus.....	8
2.3	Objective	9
2.4	Reference	Error! Bookmark not defined.
2.5	Glossary.....	Error! Bookmark not defined.
3	Problem Statement.....	9
4	Existing and Proposed solution	10
5	Proposed Design/ Model	13
5.1	High Level Diagram (if applicable)	Error! Bookmark not defined.
5.2	Low Level Diagram (if applicable).....	Error! Bookmark not defined.
5.3	Interfaces (if applicable).....	Error! Bookmark not defined.
6	Performance Test	14
6.1	Test Plan/ Test Cases	15
6.2	Test Procedure.....	16
6.3	Performance Outcome.....	16
7	My learnings.....	17
8	Future work scope	17

1 Preface

Summary of the whole 6 weeks' work.

About need of relevant Internship in career development.

Brief about Your project/problem statement.

Opportunity given by USC/UCT.

How Program was planned



Your Learnings and overall experience.

Thank to all (with names), who have helped you directly or indirectly.

Your message to your juniors and peers.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various **Cutting Edge Technologies** e.g. **Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoraWAN), Java Full Stack, Python, Front end** etc.



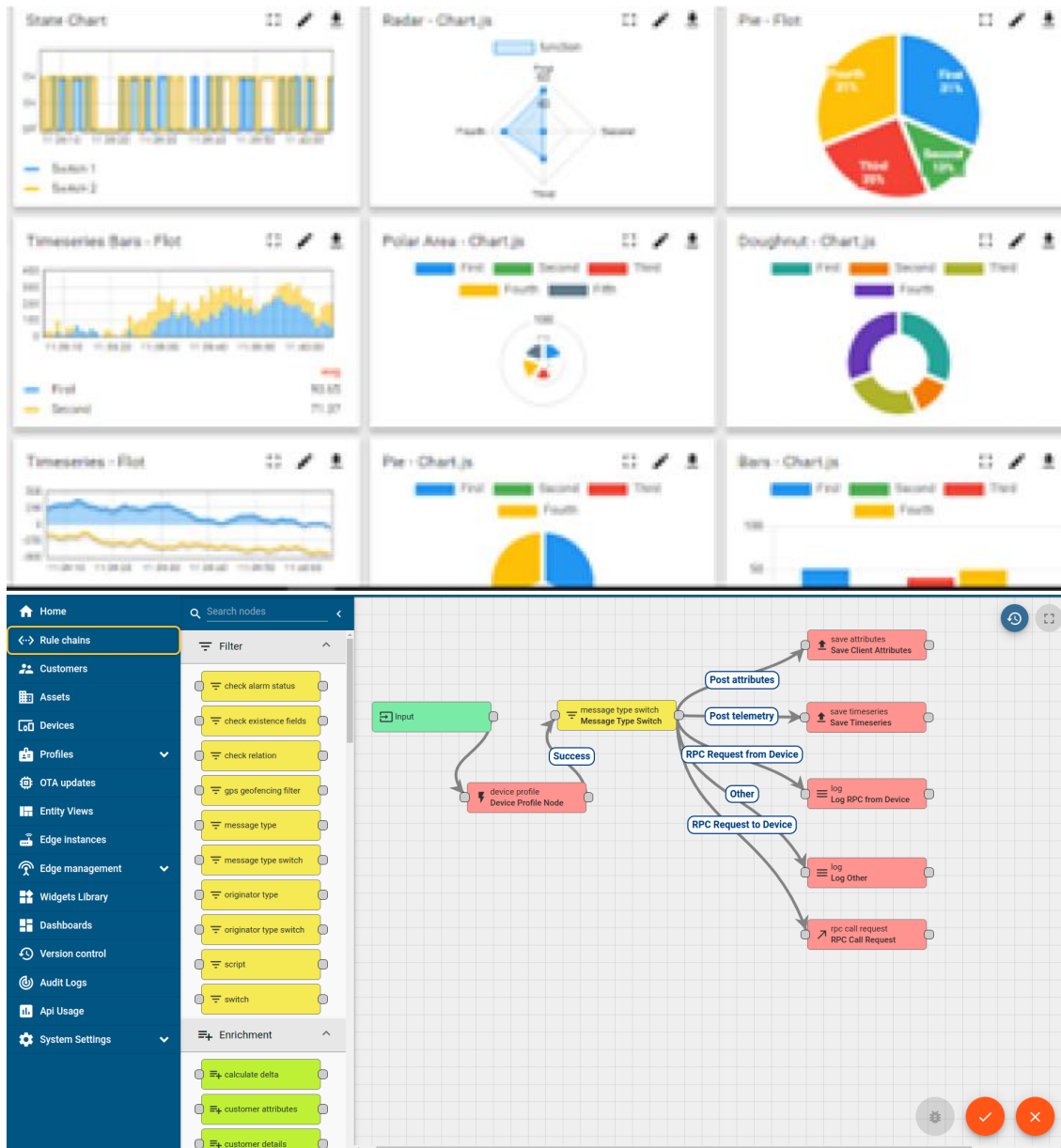
i. UCT IoT Platform ()

UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable “insight” for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
- It supports both cloud and on-premises deployments.

It has features to

- Build Your own dashboard
- Analytics and Reporting
- Alert and Notification
- Integration with third party application(Power BI, SAP, ERP)
- Rule Engine



FACTORY WATCH

ii. Smart Factory Platform ()

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleash the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they want to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.



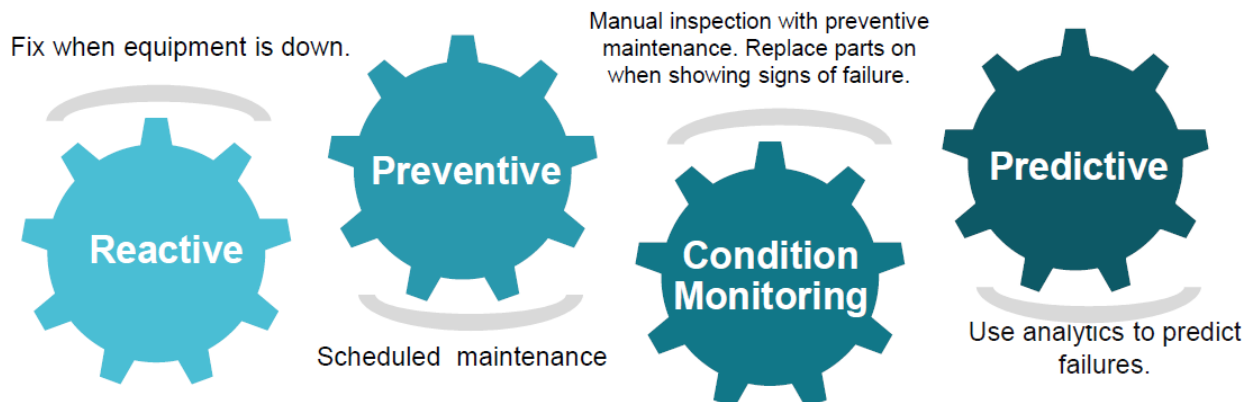


iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN technology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

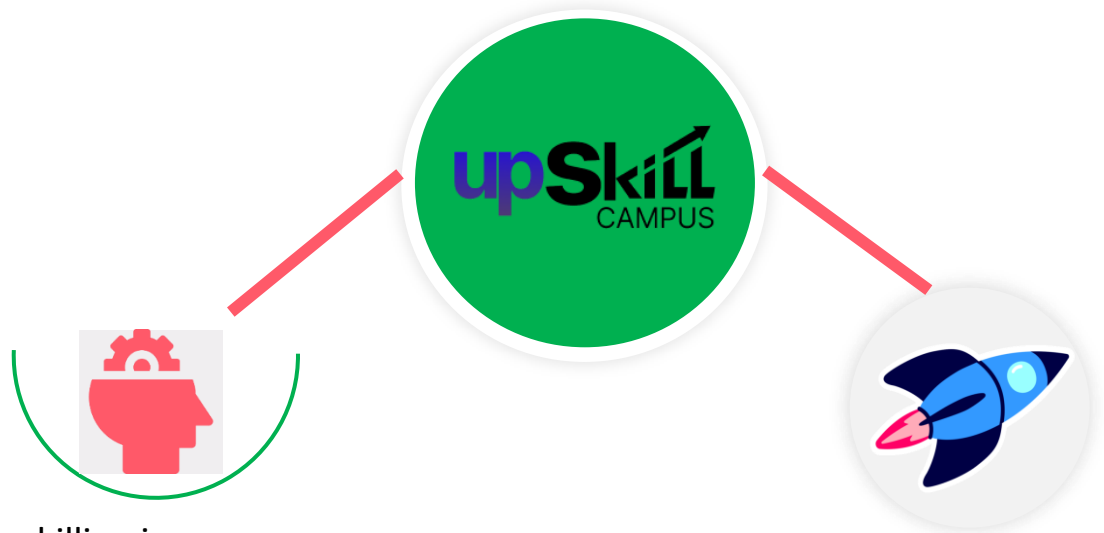
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

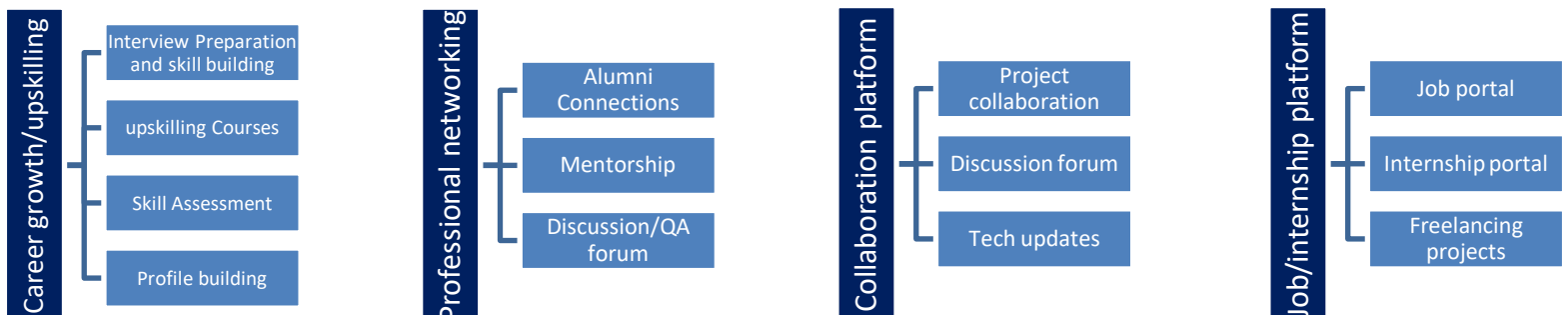
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry

upSkill Campus aiming to upskill 1 million learners in next 5 year

<https://www.upskillcampus.com/>



2.3 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

2.4 Objectives of this Internship program

The objective for this internship program was to

- get practical experience of working in the industry.
- to solve real world problems.
- to have improved job prospects.
- to have Improved understanding of our field and its applications.
- to have Personal growth like better communication and problem solving.

3 Problem Statement

Assigned problem: The healthcare system faced challenges with managing an ever-increasing volume of patient data, complex medical records, and the need for quick access to critical information. The existing on-premises infrastructure struggled to handle the scale and security demands, leading to inefficiencies, delayed decision-making, and potential risks to patient data.

The problem statement provided is about the challenges faced by the healthcare system in managing patient data. The main challenges are:

- The ever-increasing volume of patient data. The amount of data generated by healthcare organizations is growing exponentially, due to the widespread adoption of electronic health records (EHRs) and other digital technologies. This makes it difficult to store, manage, and analyze the data in a timely and efficient manner.
- The complexity of medical records. Medical records are often complex and unstructured, making it difficult to extract the relevant information. This can lead to errors in diagnosis and treatment.
- The need for quick access to critical information. Healthcare providers need to be able to access patient data quickly and easily, in order to make informed decisions about their care. The existing on-premises infrastructure often cannot meet this demand, leading to delays in treatment.

- The security risks of patient data. Patient data is highly sensitive and confidential. If it is not properly secured, it could be stolen or misused. The existing on-premises infrastructure may not be sufficient to protect patient data from cyberattacks.

These challenges can have a negative impact on the quality of care, the efficiency of healthcare organizations, and the security of patient data. To address these challenges, healthcare organizations need to adopt new technologies and approaches to data management. These include:

- Cloud computing: Cloud computing can provide a scalable and secure platform for storing and managing large volumes of data.
- Artificial intelligence (AI): AI can be used to extract insights from complex data sets and to automate tasks such as data analysis and decision-making.
- Data analytics: Data analytics can be used to identify patterns and trends in data, which can be used to improve patient care and decision-making.
- Security measures: Healthcare organizations need to implement strong security measures to protect patient data from cyberattacks.

By adopting these new technologies and approaches, healthcare organizations can overcome the challenges of managing patient data and improve the quality of care for patients.

4 Existing and Proposed solution

Here are some of the existing solutions provided by others to manage healthcare data, and their limitations:

- On-premises data centers: This is the traditional approach to data management, and it is still used by many healthcare organizations. However, on-premises data centers can be expensive to set up and maintain, and they can be difficult to scale. They are also not as secure as cloud-based solutions.
- Data warehouses: Data warehouses are centralized repositories for storing large amounts of data. They can be used to store and analyze data from a variety of sources, such as electronic health records (EHRs), clinical trials, and claims data. However, data warehouses can be expensive to build and maintain, and they can be difficult to use.
- Data lakes: Data lakes are a type of data warehouse that is designed to store all types of data, regardless of its format or structure. This makes it easier to store and analyze unstructured data, such as text and images. However, data lakes can be difficult to manage and secure.
- Cloud-based data management solutions: Cloud-based data management solutions are hosted on remote servers and accessed over the internet. This makes them more scalable and cost-effective than on-premises data centers. They are also more secure, as the data is encrypted and

protected by firewalls. However, cloud-based solutions can be complex to set up and manage, and they may not be suitable for all organizations.

Each of these solutions has its own limitations, and the best solution for a particular organization will depend on its specific needs and requirements.

Here are some additional limitations of existing solutions:

- Lack of interoperability: Different healthcare organizations often use different data formats and systems, which makes it difficult to share data between them.
- Security risks: Patient data is highly sensitive and confidential, and it is important to take steps to protect it from cyberattacks.
- Compliance challenges: Healthcare organizations are subject to a variety of regulations governing the privacy and security of patient data. These regulations can make it difficult to manage data effectively.

Despite these limitations, there are a number of innovative solutions emerging that are designed to address the challenges of healthcare data management. These solutions are using artificial intelligence (AI), machine learning, and other technologies to improve the efficiency, security, and compliance of data management.

My proposed solution is:

The healthcare system opted to adopt cloud computing to address these challenges comprehensively. A hybrid cloud approach was chosen to cater to sensitive patient data, while leveraging public cloud services for scalability and cost-effectiveness.

1. Electronic Health Records (EHR) on Private Cloud:

Sensitive patient data, such as Electronic Health Records, were stored in a private cloud environment hosted within the healthcare system's own data centers. This ensured compliance with industry regulations like HIPAA while providing granular control over data access and encryption.

2. Scalable Data Processing on Public Cloud:

To handle data-intensive tasks like medical imaging analysis and genomic sequencing, the healthcare system utilized public cloud resources. With on-demand scalability, the system could process large datasets efficiently without the need for expensive on-premises infrastructure upgrades.

3. Telemedicine and Remote Monitoring:

Leveraging the cloud, the healthcare system implemented a telemedicine platform and remote patient monitoring solutions. Healthcare providers could conduct virtual consultations, and patients could monitor vital signs from home. This enhanced patient access to healthcare services and reduced the need for in-person visits, especially for rural or remote patients.

4. Data Analytics and AI-Driven Insights:

By integrating cloud-based data analytics and Artificial Intelligence (AI) capabilities, the healthcare system could derive valuable insights from vast amounts of patient data. AI algorithms assisted in early diagnosis, personalized treatment plans, and predictive analysis, improving patient outcomes and overall healthcare quality.

5. Disaster Recovery and Data Backup:

Cloud computing provided an efficient disaster recovery solution for critical patient data. Regular backups of data stored in private and public clouds ensured business continuity in the face of unforeseen events.

4.1 Code submission (Github link):

<https://github.com/Likhithapadala79/upskillcampus/blob/main/HealthCareDataManagement.py>

4.2 Report submission (Github link):

https://github.com/Likhithapadala79/upskillcampus/blob/main/HealthCareDataManagement_Likhitha_USC_UCT.pdf

5 Proposed Design/ Model

here are some more details about the design flow of my proposed solution:

- ❖ **Data collection:** The first step is to collect data from a variety of sources, such as electronic health records (EHRs), clinical trials, and claims data. The data should be collected in a consistent and structured manner, so that it can be easily stored and analyzed. This can be done through a variety of methods, such as:
 1. **Direct patient data collection:** This involves collecting data directly from patients, such as through surveys or interviews.
 2. **Indirect patient data collection:** This involves collecting data from other sources, such as EHRs or clinical trials.
 3. **Observational data collection:** This involves collecting data by observing patients or their environment.
 4. **Data storage:** The data should be stored in a secure and reliable location. Cloud-based data storage is a good option, as it is scalable and cost-effective. The data should be stored in a way that makes it easy to access and analyze. This can be done by using a data warehouse or data lake
- ❖ **Data cleaning:** The data should be cleaned to remove errors and inconsistencies. This is important to ensure that the data is accurate and reliable. Data cleaning can be a complex and time-consuming process, but it is essential to ensure the quality of the data.
- ❖ **Data analysis:** The data should be analyzed to identify patterns and trends. This can be done using a variety of statistical and machine learning techniques. The specific techniques that are used will depend on the specific data and the goals of the analysis.
- ❖ **Data visualization:** The results of the data analysis should be visualized in a way that is easy to understand. This can help to communicate the findings to stakeholders and make informed decisions. Data visualization can be done using a variety of tools, such as charts, graphs, and maps.
- ❖ **Data sharing:** The data should be shared with other healthcare organizations to improve the quality of care and research. This can be done through data sharing agreements and by developing common data standards. Data sharing can be facilitated by using a data exchange platform.
- ❖ **Data governance:** Data governance practices should be implemented to ensure that the data is managed in a consistent and compliant manner. This includes establishing policies and procedures for data collection, storage, access, and use. Data governance can be implemented using a data governance framework.

This is just a proposed design flow, and the specific implementation will vary depending on the specific needs of the organization. However, this flow provides a good starting point for designing a healthcare data management system.

Here are some additional considerations for designing a healthcare data management system:

- ❖ The size and complexity of the data: The size and complexity of the data will affect the choice of data storage and analysis techniques.
- ❖ The privacy and security requirements: The data should be stored and analyzed in a way that protects the privacy and security of patient data.
- ❖ The cost: The cost of the system should be considered, as well as the benefits that it will provide.
- ❖ The scalability: The system should be scalable to accommodate the growth of the data.
- ❖ The usability: The system should be easy to use by healthcare professionals.

6 Performance Test

Here are some of the performance constraints that need to be considered in the design of a healthcare data management system:

- Memory: The system should be able to store and process large amounts of data. This means that the system should have enough memory to store the data and enough processing power to analyze the data.
- MIPS (Million Instructions Per Second): The system should be able to process data quickly. This means that the system should have a high MIPS rating.
- Accuracy: The system should be able to produce accurate results. This means that the system should be designed to minimize errors.
- Durability: The system should be able to store data reliably. This means that the system should be designed to protect data from corruption and loss.
- Power consumption: The system should consume power efficiently. This is important to reduce the operating costs of the system.

These are just some of the performance constraints that need to be considered in the design of a healthcare data management system. The specific constraints that need to be considered will vary depending on the specific needs of the organization.

Here are some ways to take care of these constraints in the design:

- Use cloud-based storage: Cloud-based storage is a scalable and cost-effective way to store large amounts of data.
- Use efficient data compression algorithms: Data compression algorithms can be used to reduce the amount of data that needs to be stored and processed.
- Use parallel processing: Parallel processing can be used to speed up the processing of data.
- Use fault-tolerant design: Fault-tolerant design can be used to protect data from corruption and loss.
- Use energy-efficient hardware: Energy-efficient hardware can be used to reduce the power consumption of the system.

In addition to the constraints mentioned above, there are other factors that can impact the performance of a healthcare data management system, such as the network bandwidth, the type of data being processed, and the number of users accessing the system. It is important to carefully consider all of these factors when designing a system to ensure that it meets the performance requirements.

6.1 Test Plan/ Test Cases

Here are some test cases that are used to test the performance of a healthcare data management system:

- Test case 1: The system should be able to store and process a large amount of data. This can be tested by loading the system with a large amount of data and measuring the time it takes to store and process the data.
- Test case 2: The system should be able to process data quickly. This can be tested by running a benchmark test on the system and measuring the time it takes to complete the benchmark.
- Test case 3: The system should be able to produce accurate results. This can be tested by running a test on the system with known inputs and comparing the results to the expected results.
- Test case 4: The system should be able to store data reliably. This can be tested by corrupting some of the data and then verifying that the system can still retrieve the data correctly.
- Test case 5: The system should consume power efficiently. This can be tested by measuring the power consumption of the system under different loads.

These are just some of the test cases that can be used to test the performance of a healthcare data management system. The specific test cases that need to be used will vary depending on the specific needs of the organization.

In addition to the test cases mentioned above, it is also important to test the system for security vulnerabilities. This can be done by penetration testing. Penetration testing is a process of simulating an attack on the system to identify any potential vulnerabilities.

Here are some additional considerations for testing the performance of a healthcare data management system:

- The test environment: The test environment should be representative of the production environment. This means that the test environment should have the same hardware, software, and data as the production environment.
- The test data: The test data should be representative of the real-world data that will be used by the system. This means that the test data should be large enough and complex enough to test the system under realistic conditions.
- The test methodology: The test methodology should be well-defined and repeatable. This means that the tests should be conducted in a consistent manner so that the results can be reliably compared.

- The test results: The test results should be carefully analyzed to identify any potential problems with the system. The results should also be used to make necessary adjustments to the design or configuration of the system.

6.2 Test Procedure

Here is a possible test procedure for testing the performance of a healthcare data management system:

- Define the test cases: The first step is to define the test cases that will be used. This can be done by considering the performance requirements of the system and the specific features that need to be tested.
- Prepare the test environment: The next step is to prepare the test environment. This includes setting up the hardware and software, loading the test data, and configuring the system.
- Run the tests: The third step is to run the tests. This can be done manually or by using a test automation tool.
- Analyze the results: The fourth step is to analyze the results of the tests. This includes identifying any potential problems with the system and making necessary adjustments to the design or configuration of the system.
- Repeat the tests: The final step is to repeat the tests to verify that the problems have been resolved.

6.3 Performance Outcome

The performance outcome of a healthcare data management system depends on a number of factors, including the size and complexity of the data, the performance requirements of the system, and the hardware and software used.

In general, a well-designed and implemented healthcare data management system should be able to store and process large amounts of data quickly and efficiently. The system should also be able to produce accurate results and be reliable and secure.

The specific performance outcomes that can be achieved will vary depending on the specific needs of the organization. However, some potential performance outcomes include:

- The ability to store and process terabytes or petabytes of data.
- The ability to process data queries in milliseconds.
- The ability to produce accurate results with a high degree of confidence.
- The ability to operate 24/7 with minimal downtime.
- The ability to protect data from unauthorized access, corruption, and loss.

By carefully considering the performance requirements and by using the right hardware and software, healthcare organizations can achieve the desired performance outcomes for their healthcare data management system.

Here are some additional factors that can impact the performance of a healthcare data management system:

- The type of data: The type of data being processed can have a significant impact on the performance of the system. For example, text data is typically less complex and requires less processing than image or video data.
- The number of users: The number of users accessing the system can also impact performance. A system with a large number of users will need to be able to handle the increased load.
- The network bandwidth: The network bandwidth can also impact performance. A system with a high volume of data transfers will need to have a high network bandwidth.
- The software configuration: The software configuration can also impact performance. A system that is not properly configured can experience performance problems.

7 My learnings

I learned a lot during my internship at USC/UCT. I learned about the different aspects of healthcare data management, including data collection, data cleaning, data analysis, and data visualization. I also learned about the ethical and legal considerations involved in healthcare data management. Overall, I had a great experience at USC/UCT, and I am confident that the skills and knowledge I gained will be valuable in my future career.

Here are some of my learnings:

- The importance of data management in the healthcare industry.
- The challenges of managing large and complex healthcare data.
- The different solutions that can be used to manage healthcare data.
- The importance of performance testing in healthcare data management systems.
- The different factors that can impact the performance of a healthcare data management system.
- The importance of security in healthcare data management systems.

I believe that these learnings will help me in my career growth by giving me a better understanding of the healthcare data management landscape. I can use this knowledge to develop solutions that can help healthcare organizations improve the management of their data.

8 Future work scope

I believe that relevant internships are essential for career development in any field. They provide students with the opportunity to gain practical experience, learn new skills, and network with professionals in their chosen field. In my case, the internship at USC/UCT has given me a valuable introduction to the field of healthcare data management. I have learned about the different challenges and opportunities in this field, and I have developed the skills necessary to succeed in this career.

9 Message to your juniors and peers

To my juniors and peers, I would say that if you are interested in a career in cloud computing, I encourage you to apply for the USC/UCT. This program is a great way to learn about the field and to gain the skills and experience you need to succeed.

10 Thank You message

I would like to thank edunet foundation and upskill campus for providing this internship opportunity. I want to thank kaushlendra singh sisodia And Nitin tyagi, Apurv Adhikari for helping me throughout the internship.