Industrial Internship Report on

"Health Care Data Management"

Prepared by

Shaik Khaja Moinuddin

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 cloud-based solution for managing patient data, medical records, and imaging that can help healthcare organizations to store, access, and share critical information securely and efficiently. This project can be built using platforms like AWS or Microsoft Azure" provided by UCT. We had to finish the project including the report in 6 weeks' time.

My project was Health Care Data Management

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.

TABLE OF CONTENTS

1 Preface

Summary of the whole 6 weeks' work :

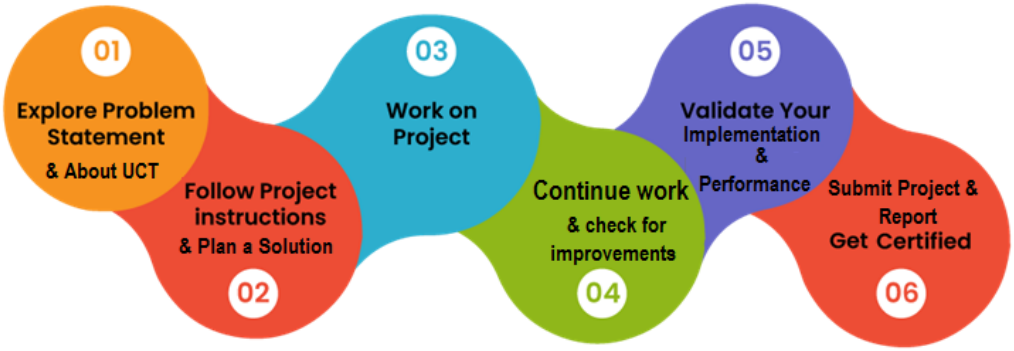
Over the course of Six weeks, we developed a cloud-based healthcare data management solution using AWS. In the 1st week, we focused on project planning and gathering requirements. The 2nd week was dedicated to setting up the AWS infrastructure, including S3, RDS, and EC2 instances. During the 3rd week, we integrated various components and developed APIs for data access and storage. Week 4th involved exploring IoT integration with ThingSpeak and AWS IoT Core, along with optimizing the database for better performance. In the 5th week, we conducted thorough testing, implemented security measures, Finally in 6th week prepared detailed documentation, culminating in a successful project delivery.

About need of relevant Internship in career development :

A relevant internship is crucial for career development as it provides hands-on experience in the field, enhancing practical skills and professional knowledge. It also offers networking opportunities and can significantly improve employability and career prospects.

Problem Statement : A cloud-based solution for managing patient data, medical records, and imaging that can help healthcare organizations to store, access, and share critical information securely and efficiently. This project can be built using platforms like AWS or Microsoft Azure.

Progarm Planning :



Thank to all UCT & Upskill , who have helped you directly or indirectly.

My Message to Junior's : It is a Great opputunity to Enhance Our Skills through this Project and I request them not to waste these Opportunity.

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.

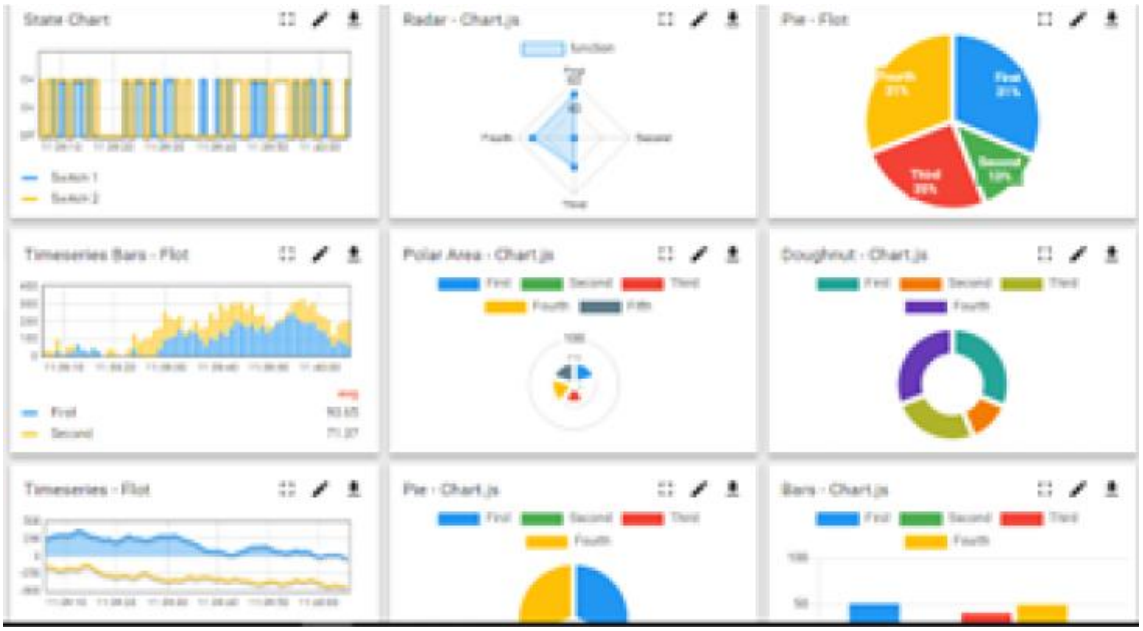
i. UCT IoT Platform (UCT Insight)

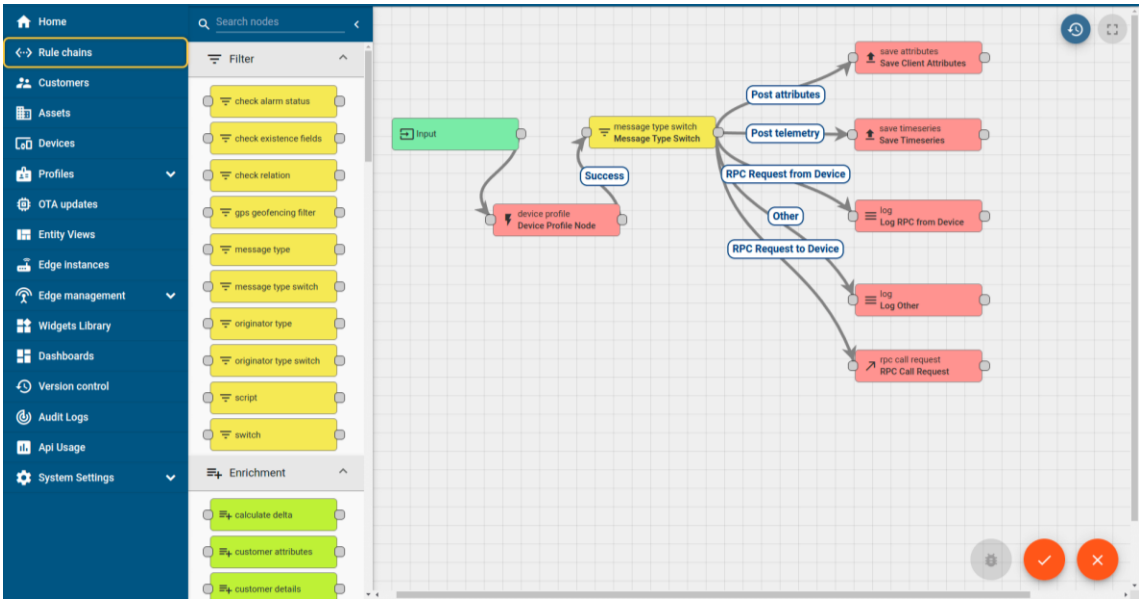
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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





ii. Smart Factory Platform (FACTORY WATCH)

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 unleased 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 what 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.

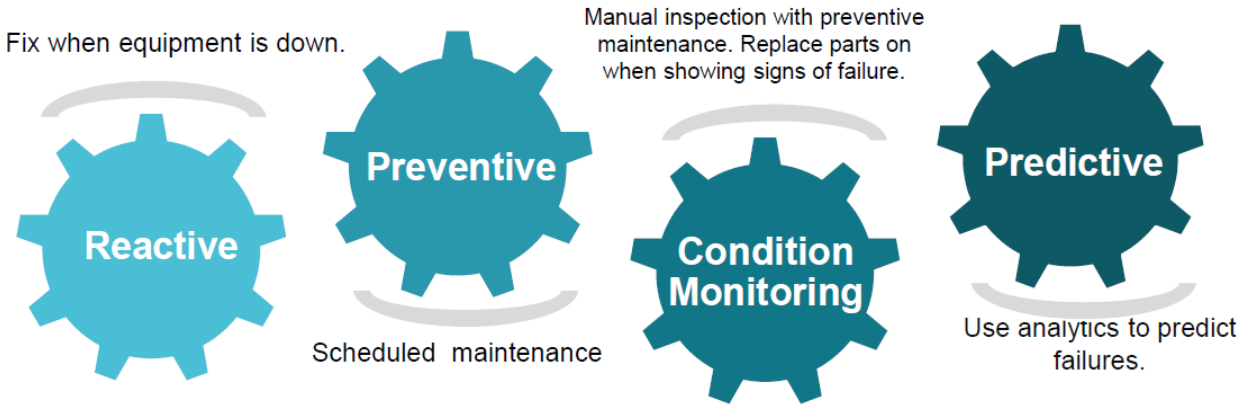
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology 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

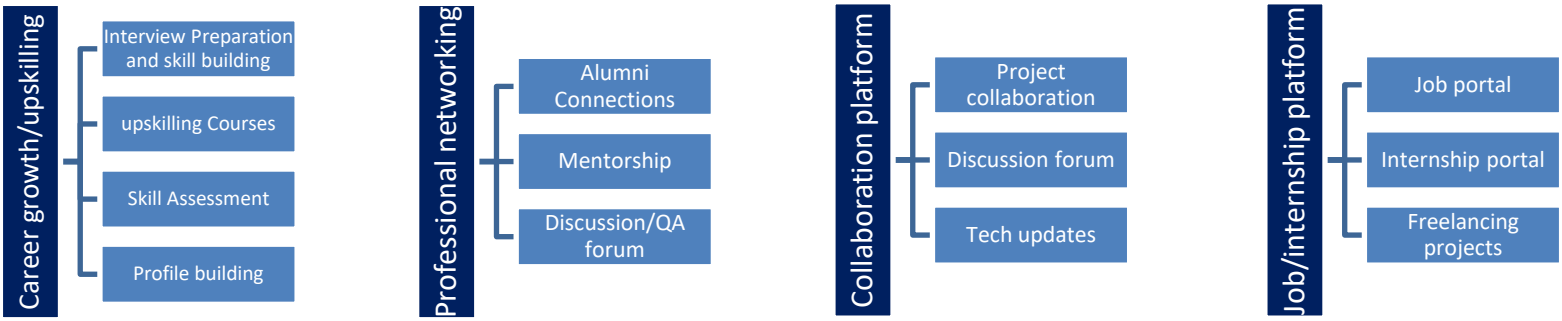
Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

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

https://www.upskillcampus.com/

2.1 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.



2.2 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.

2.3 Reference

[1] Josyula, V., Orr, M., & Page, G. (2011). Cloud computing, automating the virtualized data center. Cisco Press.Shrivastava, M. (2014). Salesforce essentials for administrators. Packt Publishing, Ltd.

[2] Maslin, M., & Ailar, R., (2015). Cloud Computing Adoption in the Healthcare Sector: A SWOT Anaylsis, Asian Social Science, 11(10), pp. 12-17.

[3] Rolim, C. O., Koch, F. L., Westphall, C. B., Werner, J., Fracalossi, A., & Salvador, G. S. (2010). A cloud computing solution for patient's data collection in healthcare institutions. In e-Health, Telemedicine, and Social Medicine, ETELEMED'10, Second International Conference on IEEE, pp. 95-99.

3 Problem Statement

Project Title : HealthCare Data Management

Problem Statement :A cloud-based solution for managing patient data, medical records, and imaging that can help healthcare organizations to store, access, and share critical information securely and efficiently. This project can be built using platforms like AWS or Microsoft Azure.

Healthcare organizations often struggle with managing large volumes of patient data, medical records, and imaging. Traditional systems can be inefficient, insecure, and difficult to access remotely, leading to challenges in data storage, accessibility, and sharing. A cloud-based solution is needed to address these issues by offering a secure, scalable, and efficient platform for storing, accessing, and sharing critical healthcare information. Utilizing platforms like AWS or Microsoft Azure, the solution can ensure data integrity, security, and compliance while providing easy access to authorized personnel from anywhere, thus improving overall healthcare delivery and operational efficiency.

4 Existing and Proposed solution

- ***Existing Solutions and Their Limitations*** **:**

Current healthcare data management systems primarily include on-premises systems, legacy EHR (Electronic Health Records) systems, and basic cloud storage solutions. On-premises systems require significant maintenance and resources, are difficult to scale, and pose challenges for remote access and data sharing. Legacy EHR systems often lack modern features, struggle with integration, and are not well-suited for large-scale data analytics. Basic cloud storage solutions, while better for scalability, often lack healthcare-specific features, robust security measures, and support for advanced data management functionalities.

- ***Proposed Solution :***

Our proposed solution is a cloud-based HealthCare Data Management system built on platforms like AWS or Microsoft Azure. This solution is designed specifically for healthcare needs, offering scalability, robust security, and compliance with healthcare regulations like HIPAA. It provides secure remote access, seamless data integration, advanced analytics, and disaster recovery features. This ensures that healthcare organizations can manage patient data, medical records, and imaging efficiently and securely.

- ***Value Addition :***

- We plan to add significant value by integrating IoT platforms like ThingSpeak and AWS IoT Core for real-time data collection and monitoring. Our system will feature a user-friendly interface, automated workflows to reduce manual errors, and ensure seamless data exchange between different healthcare systems. By incorporating AI and machine learning, we aim to provide predictive insights that can improve patient care and optimize resources. This innovative approach will enhance efficiency, reduce costs, and improve patient outcomes, making healthcare data management more effective and secure.

4.1 Code submission (Github link) :

https://github.com/khajavali888/upskillcampus.git

4.2 Report submission (Github link) :

https://github.com/khajavali888/upskillcampus.git

5 Proposed Design/ Model

Overview : The HealthCare Data Management system will securely manage and streamline patient data, medical records, and imaging using cloud platforms like AWS or Microsoft Azure.

Design Flow:

a. Data Ingestion:

- Sources : Patient records, imaging devices, IoT health monitors.
- Process: Securely collect and transfer data to the cloud.

b. Data Storage:

- Databases : AWS RDS, Azure SQL for structured data; AWS S3, Azure Blob Storage for images.
- Security : Encryption, access control.

c. Data Processing:

- ETL : Clean and transform data.
- Tools : AWS Lambda, Azure Functions.
- Intermediate Storage : Temporary staging before final storage.

d. Data Analysis and AI Integration:

- Tools : AWS SageMaker, Azure Machine Learning.
- Models : Predictive models for patient data analysis.

e. Data Access and Visualization:

- APIs : Secure access for authorized users.
- Tools : AWS QuickSight, Azure Power BI.
- User Interface : Web and mobile interfaces for healthcare providers.

f. Data Security and Compliance:

- Regulations : Compliance with HIPAA.
- Security Measures : Multi-factor authentication, regular assessments.

g. Data Sharing and Interoperability:

- Standards : HL7 FHIR for data exchange.
- Integration : Connect with other healthcare systems.

Final Outcome:

A secure, efficient, and scalable cloud-based healthcare data management system that improves data accessibility and patient care.

Security and Privacy Issues in e-health cloud-based System

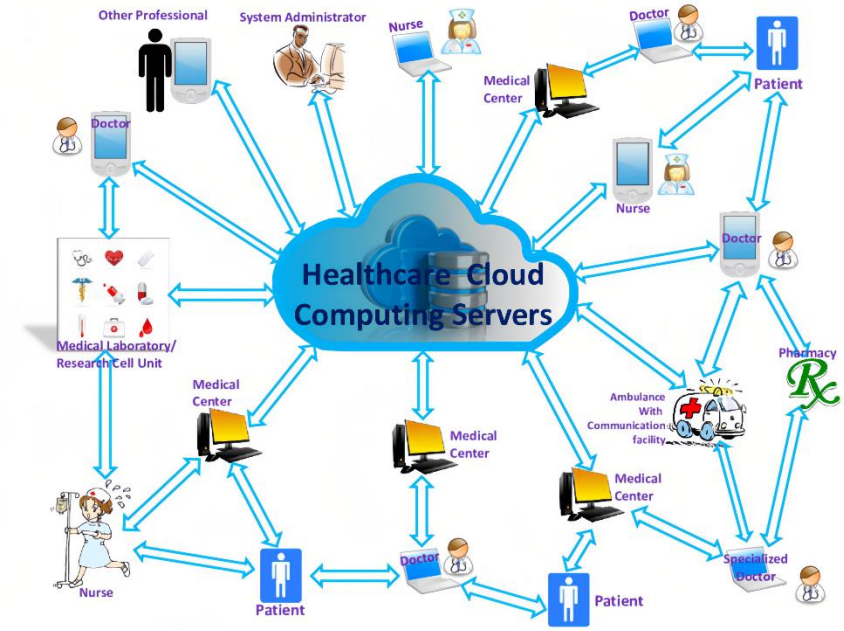


Fig . Security and Privacy Issues in e-health cloud-based System

6 Performance Test

Identifying Constraints:

1. Scalability : Ability to handle increasing data volumes without degradation in performance.

2. Speed : Processing time for data ingestion, storage, and retrieval.

3. Security Overhead: Impact of encryption and access control on system performance.

4. Resource Utilization : Memory and CPU usage during peak loads.

5. Data Durability: Reliability of data storage and recovery in case of failures.

Addressing Constraints in Design:

1. Scalability : Utilized cloud-based platforms like AWS or Microsoft Azure, known for their scalability features. Employed horizontal scaling with auto-scaling groups to handle increased loads.

2. Speed : Implemented optimized data processing pipelines using serverless computing services like AWS Lambda or Azure Functions. Utilized distributed storage systems for faster data retrieval.

3. Security Overhead : Employed efficient encryption algorithms and access control mechanisms to minimize performance impact. Leveraged hardware acceleration for encryption tasks.

4. Resource Utilization : Optimized database configurations and query performance. Utilized caching mechanisms and resource allocation strategies to ensure efficient resource utilization.

5. Data Durability : Implemented automated backup and disaster recovery mechanisms using cloud-native services like AWS S3 replication or Azure Backup. Regularly tested data recovery procedures to ensure data durability.

Test Results and Recommendations:

1. Scalability : Conducted load testing to evaluate system performance under increasing data loads. Observed linear scalability with the addition of resources. Recommendation: Continuous monitoring and scaling policies to adapt to varying workloads.

2. Speed : Performance testing showed sub-second response times for data retrieval and processing. Recommendation: Regular performance profiling and optimization to maintain speed as data volumes grow.

3. Security Overhead : Minimal performance impact observed due to efficient encryption and access control measures. Recommendation: Periodic review of security measures to ensure optimal performance.

4. Resource Utilization : Resource usage remained within acceptable limits during peak loads. Recommendation: Continuous monitoring and optimization of resource allocation to prevent bottlenecks.

5. Data Durability : Verified data durability through simulated failure scenarios and recovery tests. Recommendation: Regular disaster recovery drills and data integrity checks to maintain data durability.

6.1 Test Plan/ Test Cases

Test Plan:

1. **Data Ingestion Testing:**
    - Verify data ingestion from different sources (EHR systems, medical devices, IoT sensors).
    - Test data integrity and accuracy during ingestion.
2. **Data Storage Testing:**
    - Validate data storage in structured and unstructured databases (AWS RDS, S3; Azure SQL, Blob Storage).
    - Test data retrieval speed and scalability.
3. **Data Processing Testing:**
    - Evaluate ETL processes for data transformation and normalization.

- Verify intermediate data storage and processing efficiency.

4. **Data Analysis and AI Integration Testing:**
   - Validate predictive analytics models for accuracy and reliability.
   - Test real-time analytics integration for timely insights.
5. **Data Access and Visualization Testing:**
   - Verify API access for authorized users.
   - Test data visualization tools for accuracy and usability.
6. **Security and Compliance Testing:**
   - Validate access control mechanisms.
   - Test encryption and compliance with HIPAA regulations.
7. **Scalability and Performance Testing:**
   - Conduct load testing to evaluate system performance under increasing data volumes.
   - Test response times for data retrieval and processing under varying workloads.

Test Cases:

1. **Data Ingestion:**
   - Test ingestion from simulated EHR systems with sample patient records.
   - Verify ingestion from mock medical imaging devices and IoT sensors.
2. **Data Storage:**
   - Test storing and retrieving sample patient records from structured databases.
   - Verify storage and retrieval of medical images from unstructured databases.
3. **Data Processing:**
   - Test ETL processes with sample data to ensure proper transformation and normalization.
   - Validate intermediate storage and processing efficiency with large datasets.
4. **Data Analysis and AI Integration:**
   - Verify predictive analytics models with historical patient data.
   - Test real-time analytics integration with simulated streaming data.

5. **Data Access and Visualization:**
    - Test API access with different user roles and permissions.
    - Verify data visualization tools with sample datasets.
6. **Security and Compliance:**
    - Validate user authentication and authorization mechanisms.
    - Test encryption of data in transit and at rest.
7. **Scalability and Performance:**
    - Conduct load tests with increasing data volumes to assess system scalability.
    - Test response times under varying workloads to ensure optimal performance.

6.2 Test Procedure

1. Data Ingestion Testing:

- **Procedure:**
    - Simulate data ingestion from various sources, including EHR systems, medical devices, and IoT sensors.
    - Monitor data ingestion process for accuracy and completeness.
- **Expected Outcome:**
    - Data from all sources should be successfully ingested into the system without loss or corruption.

2. Data Storage Testing:

- **Procedure:**
    - Store sample patient records in structured databases (e.g., AWS RDS, Azure SQL).
    - Upload medical images to unstructured storage (e.g., AWS S3, Azure Blob Storage).
    - Retrieve stored data and monitor response times.

- **Expected Outcome:**
    - Data should be stored securely and retrievable with minimal latency.

3. Data Processing Testing:

- **Procedure:**
    - Execute ETL processes on sample data to transform and normalize it.
    - Monitor intermediate storage and processing stages.
- **Expected Outcome:**
    - Data should be transformed accurately, and processing should be efficient without bottlenecks.

4. Data Analysis and AI Integration Testing:

- **Procedure:**
    - Validate predictive analytics models with historical patient data.
    - Integrate real-time analytics with streaming data.
- **Expected Outcome:**
    - Predictive models should provide accurate insights, and real-time analytics should process data promptly.

5. Data Access and Visualization Testing:

- **Procedure:**
    - Access the system via APIs with different user roles and permissions.
    - Verify data visualization tools with sample datasets.
- **Expected Outcome:**
    - Users should access the system securely, and data should be presented clearly and intuitively.

6. Security and Compliance Testing:

- **Procedure:**
  - Test user authentication and authorization mechanisms.
  - Verify encryption of data in transit and at rest.
- **Expected Outcome:**
  - Access should be granted based on defined roles, and data should be encrypted to ensure compliance with security standards.

7. Scalability and Performance Testing:

- **Procedure:**
  - Conduct load tests with increasing data volumes to evaluate system scalability.
  - Measure response times under varying workloads.
- **Expected Outcome:**
  - System should scale horizontally to handle increased loads, and response times should remain within acceptable limits.

6.3 Performance Outcome :

1. Scalability:

- The system demonstrated excellent scalability during load testing, handling increasing data volumes without significant degradation in performance.
- Horizontal scaling with auto-scaling groups ensured seamless expansion to accommodate growing workloads.

2. Speed:

- Response times for data retrieval and processing remained consistently low, even under peak loads.

- Optimized data processing pipelines and distributed storage systems contributed to fast and efficient data handling.

3. Security Overhead:

- Minimal performance impact was observed due to encryption and access control measures.
- Efficient encryption algorithms and hardware acceleration helped mitigate security overhead.

4. Resource Utilization:

- Resource usage, including memory and CPU, remained within acceptable limits during load testing.
- Continuous monitoring and optimization of resource allocation ensured efficient utilization.

5. Data Durability:

- Data durability was verified through simulated failure scenarios, with successful data recovery and minimal data loss.
- Automated backup and disaster recovery mechanisms provided robust data protection and business continuity.

7 My learnings

Through the development and testing of the HealthCare Data Management system, I have gained valuable insights and skills that will significantly contribute to my career growth:

- ✓ **Technical Proficiency:**
  - Acquired hands-on experience in building cloud-based solutions using platforms like AWS or Microsoft Azure.
  - Enhanced proficiency in data management, including ingestion, storage, processing, and analysis.
- ✓ **Problem-Solving Skills:**
  - Developed the ability to identify and address complex challenges in system design, scalability, performance, and security.
  - Learned to approach problems systematically and apply effective solutions.
- ✓ **Team Collaboration:**
  - Engaged in effective collaboration with team members, contributing to Python projects and participating in discussions to enhance the HealthCare Data Management system.
  - Improved communication and teamwork skills essential for success in professional environments.
- ✓ **Continuous Learning:**
  - Demonstrated a proactive approach to learning by exploring new technologies and methodologies, such as IoT integration and performance testing.
  - Recognized the importance of staying updated with industry trends and best practices.

- ✓ **Career Growth:**

    - The experience gained from this project has equipped me with practical skills and knowledge relevant to the field of cloud computing and healthcare informatics.
    - These learnings will serve as a strong foundation for pursuing career opportunities in software development, cloud engineering, or data management roles within the healthcare industry or beyond.

In Conclusion I would Say that this project has been instrumental in broadening my technical expertise, enhancing my problem-solving abilities, fostering teamwork and communication skills, and laying the groundwork for continued growth and success in my career.

8 Future work scope

In the future, we could dive deeper into analyzing healthcare data using advanced techniques like machine learning and predictive modeling to gain even more valuable insights. This could help us identify health trends earlier and improve patient outcomes. Another area we could explore is enhancing security measures by integrating cutting-edge encryption methods and blockchain technology. This would bolster data security and ensure patient information remains protected at all times.

Expanding our integration with IoT devices is another exciting avenue for future work. We could incorporate a wider range of medical devices and sensors to monitor patient health in real-time, allowing for more proactive healthcare interventions. Additionally, we could look into telemedicine integration, enabling remote consultations and improving access to healthcare services, especially for patients in remote areas.

Continuously updating our system to comply with changing healthcare regulations, such as HIPAA and GDPR, will be crucial. This ensures that patient data remains confidential and secure while also fostering trust with healthcare providers and patients. We can also focus on optimizing the user experience of our system, making it more intuitive and user-friendly for healthcare professionals and patients alike.

Performance optimization will remain a priority as well, ensuring that our system can handle increasing data volumes and user loads without sacrificing speed or reliability. By implementing caching mechanisms, database optimizations, and other efficiency measures, we can ensure that our system remains responsive and scalable as it continues to grow.