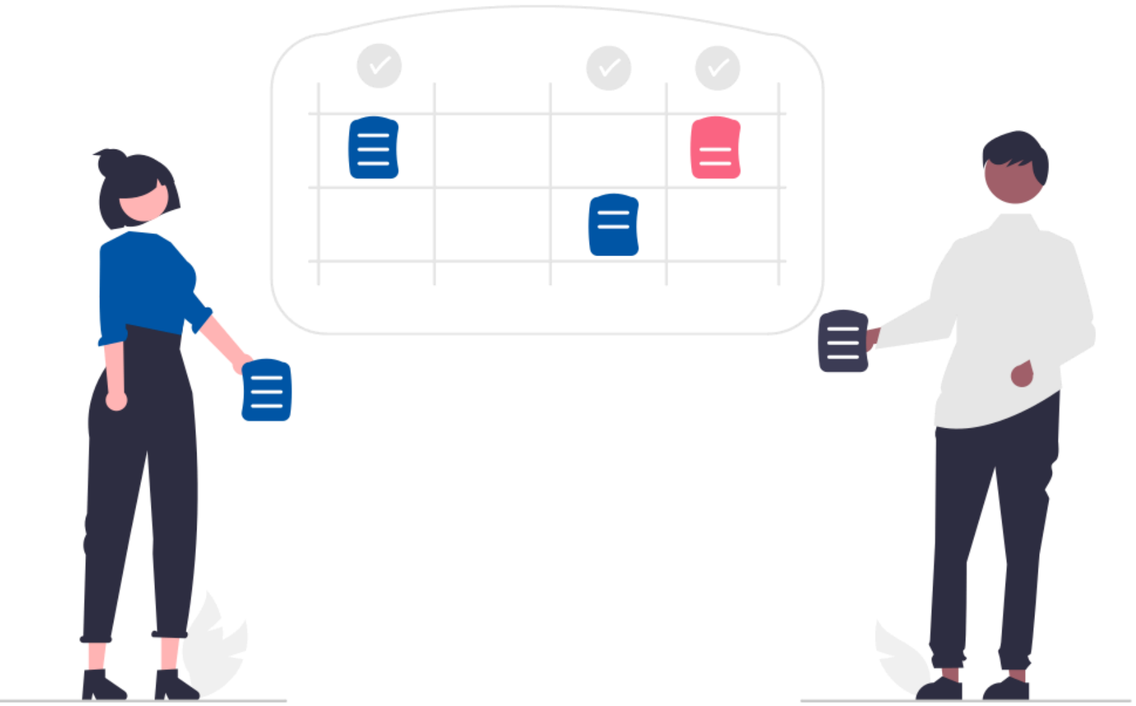
PBL5 - PEC5
ADVANCED DATABASE SYSTEMS
Review

DEPARTMENT OF COMPUTER ENGINEERING
PCET'S PIMPRI CHINCHWAD COLLEGE OF ENGINEERING

CloudDocs: Cloud-Based Document Management System with AWS S3 Integration

Presented By:
Atharva Lende(121B1B144)
Sakshi Kulkarni (122B2B289)
Janhavi Mali (122B2B292)
Vinayak Shete (122B2B303)

Problem Definition



Organizations rely heavily on documents and files to manage their operations, collaboration, and communication. These documents must be stored securely, accessed quickly, and managed efficiently

Introduction

- Traditional file storage solutions like local servers or basic cloud storage may not always offer the scalability, accessibility, and fault tolerance necessary to handle growing document volumes.
- Additionally, file access and management across multiple devices and geographic locations become challenging, especially when considering security, availability, and ease of use.
- The problem is that many existing document management solutions are either costly or inefficient in scaling with an organization's growth.
- The project aims to solve various issues by leveraging AWS S3 for document storage, which provides high durability and availability.

Literature Survey

Paper Name	Methodology	Strengths	Weaknesses
Cloud Storage: The Future of Data Backup and Recovery (Zhang et al., 2010)	The paper discusses the architecture and mechanisms of cloud storage, focusing on AWS S3's scalability, durability, and redundancy. It examines the cost-effectiveness of cloud solutions compared to traditional systems.	- Cost-effective alternative to traditional storage - High durability and reliability - Data replication across multiple availability zones for enhanced data integrity and loss prevention	- Dependence on internet connectivity - Potential security concerns related to data breaches - Limited control over data management in the cloud

Literature Survey

Paper Name	Methodology	Strengths	Weaknesses
Cloud Computing Security Issues and Challenges (Ali et al., 2015)	The study addresses data security and access management in cloud storage, focusing on secure data transmission, encryption, and role-based access control (RBAC) to protect sensitive information.	- Highlights importance of encryption and access control - Emphasizes RBAC for enhanced security - Addresses various security challenges in cloud environments	- May not cover all potential security threats - Implementation of encryption and RBAC can be complex and resource-intensive

Literature Survey

Paper Name	Methodology	Strengths	Weaknesses
Cloud-Based Document Management System and Security Framework (Singh et al., 2012)	This research discusses the deployment of cloud-based Document Management Systems (DMS) and how they improve accessibility and simplify traditional DMS architectures, using AWS IAM for access control.	- Improves accessibility of documents - Reduces complexity of DMS architecture - Leverages AWS for scalability and flexibility in document management	- Possible concerns about data security and compliance in the cloud - Reliance on internet access for document management

Literature Survey

Paper Name	Methodology	Strengths	Weaknesses
A Comparative Analysis of Document Management Systems (Gupta et al., 2016)	The paper evaluates various cloud-based DMS solutions, comparing features such as cost efficiency, scalability, and ease of use, emphasizing the advantages of cloud platforms over traditional systems.	- Highlights cost efficiency and scalability - Improves efficiency of document retrieval and sharing - Comprehensive comparison of various DMS solutions	- May not provide in-depth analysis of specific systems - Results can vary based on individual organizational needs

Literature Survey

Paper Name	Methodology	Strengths	Weaknesses
Architecting for the Cloud: AWS Best Practices (Varia et al., 2013)	The book provides an overview of building scalable and fault-tolerant applications using AWS services, discussing S3, EC2, NGINX for load balancing, auto-scaling, and the pay-as-you-go model.	- Comprehensive insights into AWS best practices - Emphasizes cost-effective scalability - Highlights advantages of AWS services for application development	- Focuses primarily on AWS, which may limit applicability to other cloud platforms - May lack in-depth technical implementation details

Literature Survey

Paper Name	Methodology	Strengths	Weaknesses
Disaster Recovery in Cloud Computing: A Survey (Jain et al., 2014)	This survey examines built-in disaster recovery mechanisms offered by cloud providers, focusing on AWS S3's multi-region replication for data backup during region-wide failures.	- Highlights important disaster recovery strategies - Discusses AWS's multi-region capabilities - Emphasizes the need for high availability in cloud systems	- Limited focus on implementation details - May not address all disaster recovery scenarios

Motivation

- **Accessibility and Convenience:**

Enables anytime, anywhere access to documents, supporting remote work and distributed teams.

- **Scalability for Growing Data Needs:**

Automatically scales with organizational growth, eliminating the need for costly hardware upgrades.

- **Enhanced Security and Data Integrity:**

Utilizes AWS's robust encryption and access controls, ensuring sensitive data remains secure and protected.

- **Cost Efficiency:**

Adopts a pay-as-you-go model, reducing costs related to maintaining local servers and physical storage.

- **Simplifying Document Management:**

Provides a user-friendly interface for non-technical users, streamlining document upload, download, and management.

- **Disaster Recovery and Business Continuity:**

Leverages AWS's disaster recovery capabilities to ensure document availability during system failures or disasters.

- **Support for Remote and Distributed Teams:**

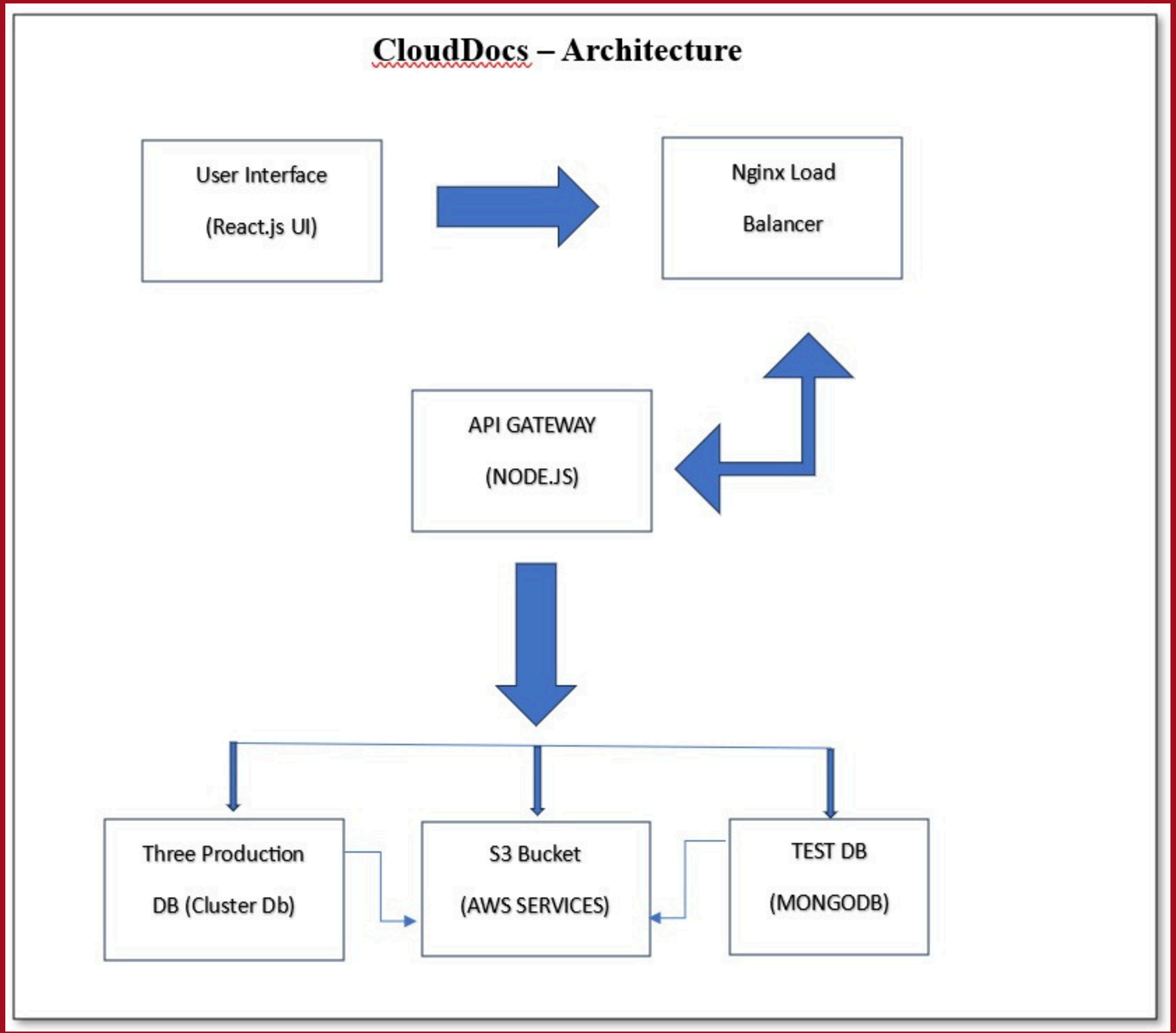
Facilitates centralized storage with decentralized access, enabling seamless collaboration across locations.

Scope

This project primarily focuses on building a scalable, secure, and reliable document management system using AWS S3 for storage. The key features of the project are:

1. **Uploading and Managing Documents: Users will be able to upload, list, download, and delete documents using a web-based UI.**
2. **Cloud-Based Infrastructure: Leveraging AWS services such as S3, EC2, and IAM for secure and scalable storage and processing.**
3. **Database Scalability: Using a MongoDB cluster for document metadata management, ensuring data availability across different nodes in case of failure.**
4. **Load Balancing and Frontend Caching: NGINX will be used to handle load balancing and cache frontend requests, improving performance and ensuring fault tolerance.**
5. **Security: Implementing IAM roles and access policies to control who can upload, view, or delete documents.**
6. **Testing and Monitoring: The system will include continuous monitoring and periodic testing to ensure reliability.**

System Architecture



Database Design

1) S3 Bucket

- **The S3 Bucket stores the actual document files such as PDFs, Word documents, and images. Each file is given a unique key and can be accessed via its S3 URL.**
- **Structure: Files can be organized in folders within the bucket if needed, though each file is identified by a unique key even if the same name is used.**
- **Advantages: S3 automatically manages data redundancy and durability, ensuring that files are never lost.**

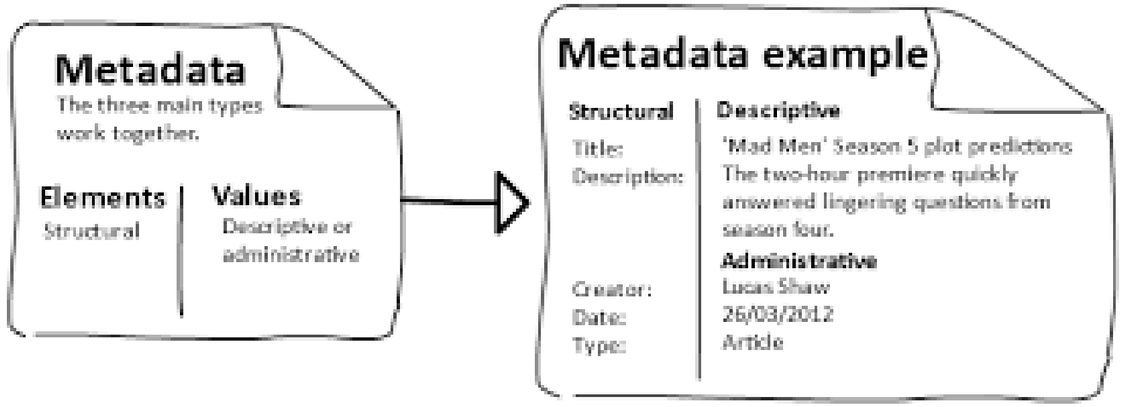
2)Metadata: MongoDB Stores Metadata

- **MongoDB stores metadata about the documents uploaded to S3. This includes details like:**

oFile name o Upload date o File size

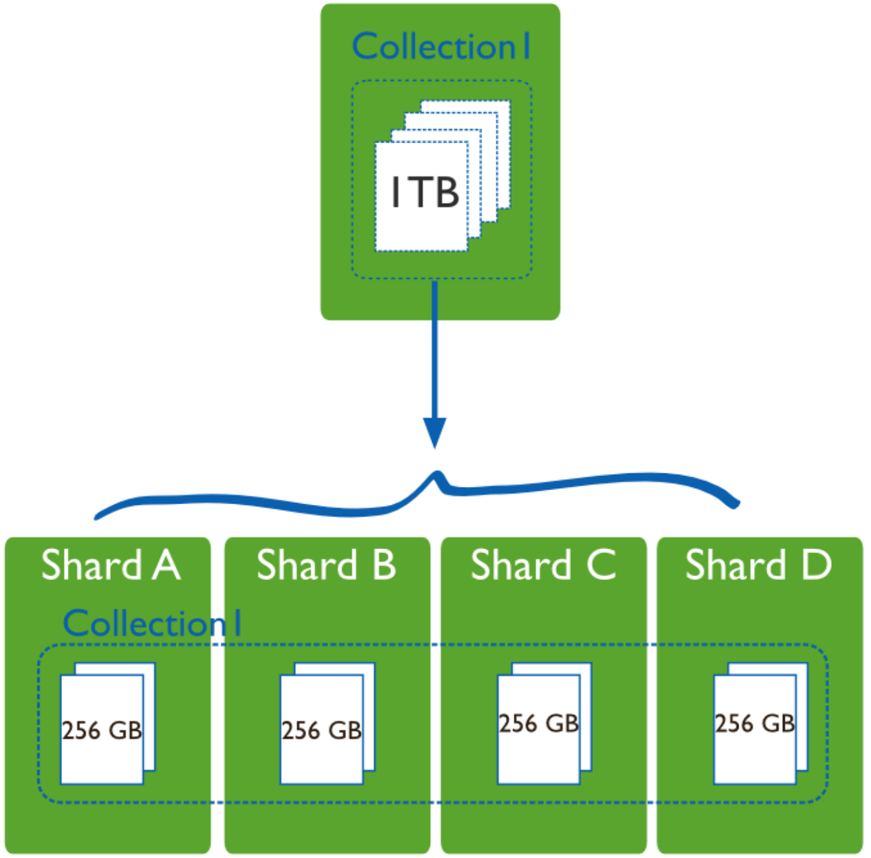
oFile type (PDF, Word, Image, etc.)

- **Why Metadata? Storing metadata in MongoDB allows the system to provide quick access to document information without querying S3 for file details, improving performance.**



3)Clusters: MongoDB Clusters for High Availability

- **Production Clusters: There are three clusters in production that ensure continuous availability of the database. In case of failure of one cluster, the others take over, ensuring no downtime.**
- **Test Database: A separate test database is used for development and testing purposes. This ensures that production data remains unaffected during updates or testing.**



DEMO

Conclusion

The Cloud-Based Document Management System with AWS S3 Integration is designed to be a robust, scalable, and secure solution for organizations needing efficient document management. By leveraging AWS services like S3, EC2, and IAM, along with MongoDB for metadata, the system can provide high availability, fault tolerance, and global accessibility.

References

Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud storage: The future of data backup and recovery. ACM Computing Surveys (CSUR), 42(3), 1-17. DOI: 10.1145/1670679.1670682.

Ali, M., Khan, S. U., & Vasilakos, A. V. (2015). Security in cloud computing: Opportunities and challenges. Information Sciences, 305, 357-383. DOI: 10.1016/j.ins.2015.01.025.

Singh, M., Kalra, S., & Mangat, V. (2012). Cloud-based document management system and security framework. International Journal of Computer Applications, 44(7), 14-19. DOI: 10.5120/6232-8459.

Gupta, P., Gulati, N., & Kumar, A. (2016). A comparative analysis of document management systems: A cloud perspective. Journal of Information Technology & Software Engineering, 6(4), 1-6. DOI: 10.4172/2165-7866.1000188.

THANK YOU!