# BUAN 6335.501:ORGANIZATION FOR BUSINESS ANALYTICS PLATFORMS

## CASE STUDY:

# DATA MANAGEMENT SOLUTION FOR UNIVERSITY


**GROUP - 11**

ASTHA BHATIYA – AXB210151

DEVIKA PAWAR - DRP210005

HARITHA JAMPANI - HXJ220031

RUTUJA PABALE - RNP210003

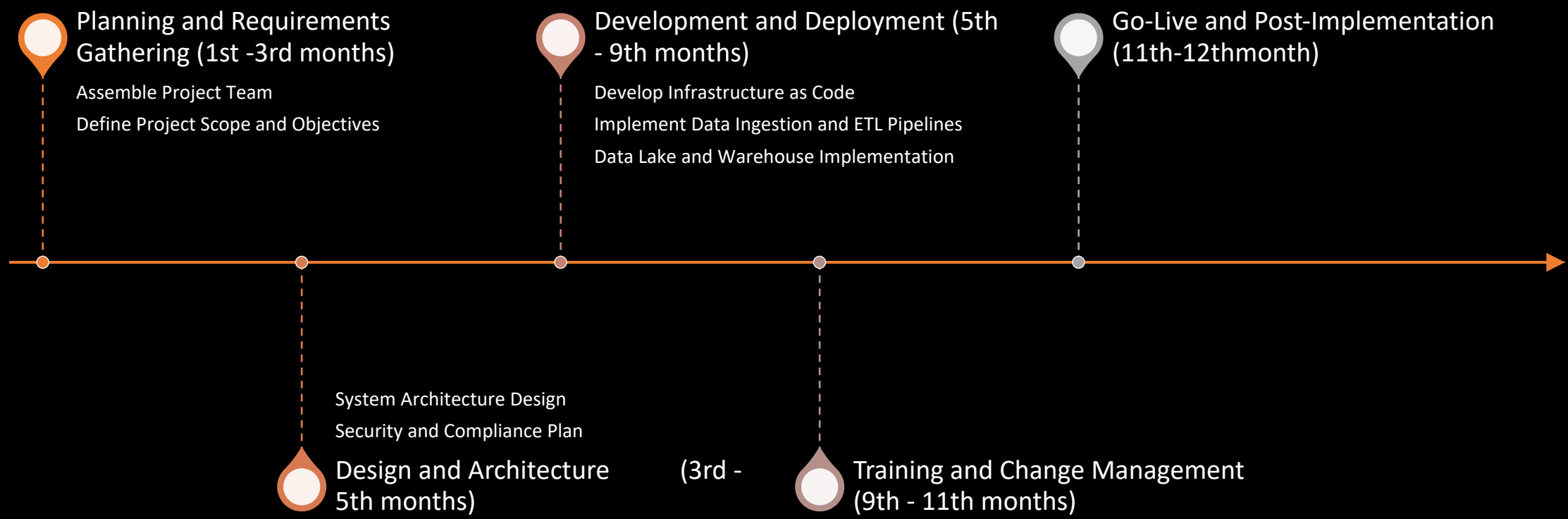
SHIVAANI VIJAYAKUMAR - SXV210084

NITHIN REDDY - NXP210094

# PROBLEM STATEMENT
# Challenges in the university's current data architecture

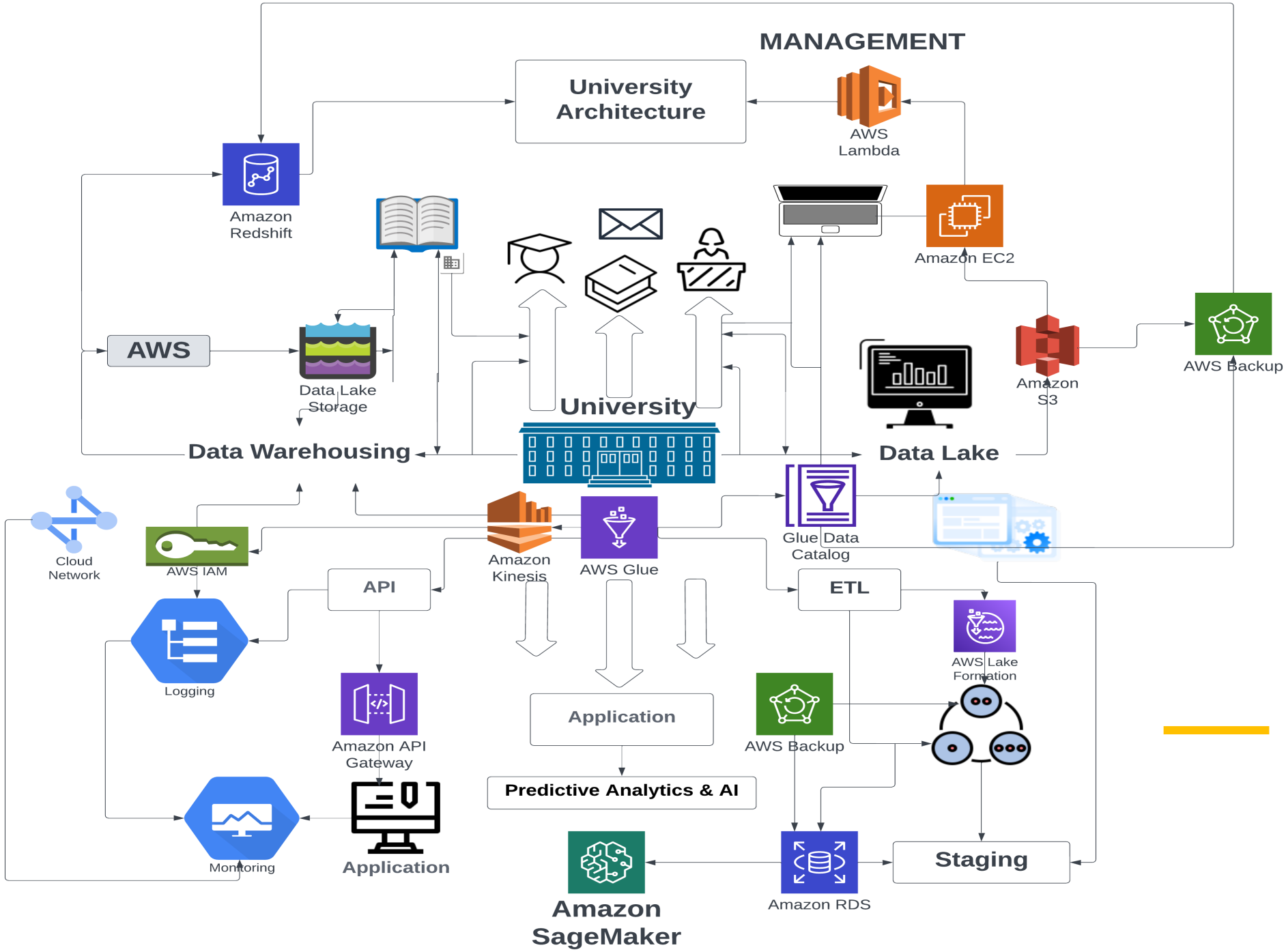
| Challenge | Description |
| --- | --- |
| Outdated Data Architecture | The university's current data architecture is outdated and consists of a mix of cloud and on-premise data centers. This leads to challenges in data management and analytics. |
| Lack of Unified Data Architecture | The university lacks a unified data architecture, resulting in data silos and difficulties in accessing and integrating data from different sources. |
| Inefficient Data Management | The university faces difficulties in managing and organizing its data, leading to inefficiencies in data processing and analysis. |
| Limited Predictive Analytics | Due to the outdated data architecture, the university has limited capabilities in leveraging predictive analytics to make data-driven decisions. |
| Data Security Risks | The mix of cloud and on-premise data centers poses data security risks, including potential breaches and unauthorized access to sensitive information. |
| Lack of Data Standardization | The university lacks standardized data formats and structures, making it challenging to ensure consistency and accuracy in data analysis and reporting. |

# MILESTONE OF THE UNIVERSITY DATA MANAGEMENT



# DATA SOURCES FOR CLOUD-BASED ARCHITECTURE

- Grades, enrollment, and attendance data are stored in the Student Information System (SIS).
- Performance, activities, and interactions in virtual courses.
- Campus administration, human resources, and financial operations.
- Data on student engagement and resource usage in digital learning are available from online educational platforms.
- Use statistics, research publications, and scholarly resources are all part of library and academic research.
- Data from outside educational software and tools is referred to as external educational tools.
- Data on alumni and fundraising, including contributions, events, and involvement of alumni.
- Views from staff and student questionnaires and assessments.
- Industry data, educational research, and demographic trends.

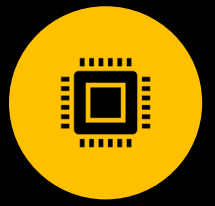



# DATA ARCHITECTURAL STRATEGY : OUR APPROACH

# DATA INTEGRATION

COMBINING VARIOUS DATA SETS FROM DIFFERENT DEPARTMENTS WITHIN THE UNIVERSITY BY USING AWS S3.

UTILIZING AWS GLUE TO ASSURE DATA CONSISTENCY DURING THE EXTRACT, TRANSFORM, AND LOAD PROCESSES.



USING AMAZON KINESIS FOR REAL-TIME DATA FEEDS, IMPROVING THE CAPACITY FOR QUICK DATA ANALYSIS.

TRANSFERRING CURRENT DATA TO THE CLOUD USING EFFECTIVE AND SAFE TECHNIQUES.

ENSURING SMOOTH DATA TRANSFER BETWEEN VARIOUS PLATFORMS, SUCH AS ADMINISTRATIVE PLATFORMS, SIS, AND LMS.

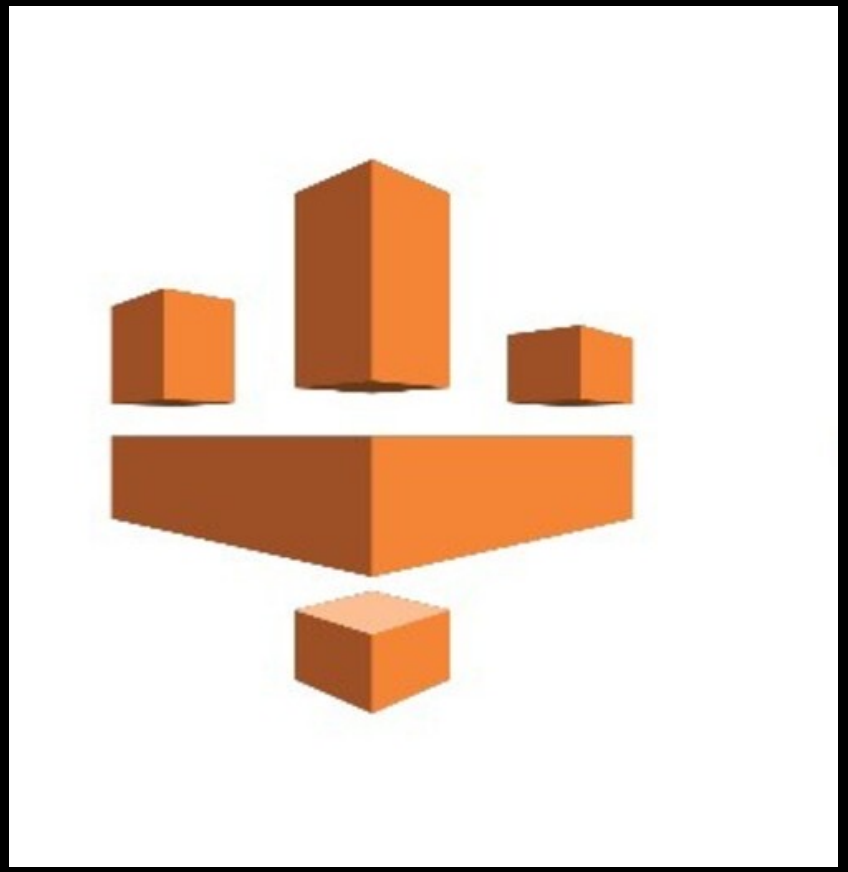
RETAINING HIGH DATA QUALITY BY USING TOOLS FOR VALIDATION, DEDUPLICATION, AND CLEANING.

RESPECTING DATA SECURITY GUIDELINES AND DIRECTIVES WHILE INTEGRATING SYSTEMS.

CREATING A SYSTEM THAT CAN ADJUST TO EVOLVING UNIVERSITY REQUIREMENTS AND INCREASING DATA VOLUMES.

# DATA MANAGEMENT USING AWS GLUE

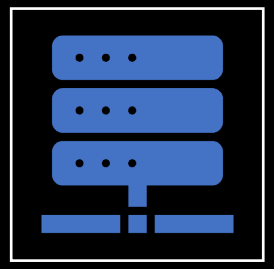
- Integrating and preparing data by using AWS Glue as the primary ETL (Extract, Transform, Load) service.
- Serverless, scalable architecture of Glue simplifies the processes involved in data extraction and transformation.
- Ensuring data quality through the uniform cleaning, deduplication, and transformation of data from multiple sources.
- Establishing a seamless connection with AWS Redshift to load transformed data and AWS S3 for data extraction.
- The capacity of AWS Glue to identify data schemas and generate ETL scripts automatically.
- Utilizing AWS services to process data in almost real-time so that timely analytics are supported.
- Building effective and scalable data pipelines to meet the expanding data requirements of the university.
- Ensuring that data governance guidelines and privacy standards are followed by every ETL process.



# AMAZON KINESIS FIREHOSE FOR REAL-TIME DATA STREAMING

- Captures and automatically loads streaming data into AWS services for analysis and storage.
- Seamlessly integrates with Amazon S3, Redshift, and Elasticsearch Service for data storage and analytics.
- Offers an easy-to-use solution for streaming large amounts of data without managing infrastructure.
- Supports transforming incoming streaming data using AWS Lambda, enhancing data quality and format.
- Automatically scales to match the throughput of incoming data, ensuring efficient handling of data peaks.
- Provides reliable data transfer with automatic retries, data encryption, and data loss prevention mechanisms.
- Facilitates near-real-time analytics, enabling quick insights and decision-making processes for the university.

# STREAMLINING UNIVERSITY DATA MANAGEMENT WITH AWS SERVICES

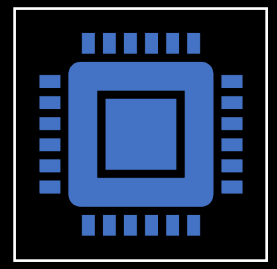


**Amazon S3 for Data Storage**:

S3 is our central data lake repository, offering secure, durable, and scalable storage. It's ideal for a variety of use cases including websites, mobile applications, backup and restore, archives, and big data analytics, making it a versatile foundation for university operations.

**Amazon EC2 for Scalable Operations**:

EC2 provides resizable compute capacity in the cloud. It's used to host virtual servers and manage a variety of workloads, offering flexibility with different instance types. EC2 supports our university's dynamic needs with its on-demand resource availability, ensuring we can scale operations efficiently.

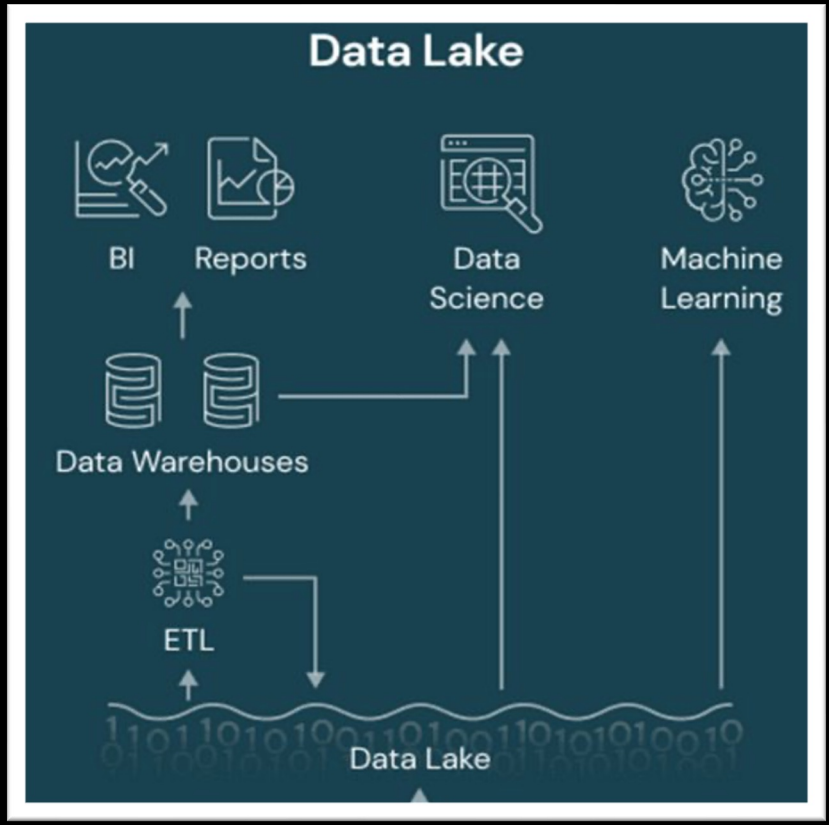


**Amazon Lambda for Backend Processing**:

Lambda enables us to run code in response to triggers from AWS services like S3 and EC2 without managing servers, streamlining our data workflows. It integrates with S3 for data storage and Kinesis for data streaming, facilitating real-time data processing and advanced analytics. Lambda automatically scales with the workload, efficiently handling activity spikes during peak university operations.

# DATA LAKE: CORE OF UNIVERSITY'S DATA STRATEGY

- Functions as a vast reservoir to store all data types - structured (like CSV files), semi-structured, and unstructured (such as video/audio files), regardless of scale.
- Designed to store massive volumes of data efficiently, focusing on minimizing costs without compromising data integrity.
- Prioritizes stability and security, ensuring data protection and compliance with educational data standards.
- Adapts to the university's growing data needs, offering a scalable and flexible management solution for diverse data types.
- Facilitates a range of uses from enhancing administrative processes to powering advanced academic research.
- Acts as a centralized pool for all raw data extracts, providing a unified view of source data for comprehensive analysis.
- Integrates with AWS Lake Formation to streamline the setting up and securing of the data lake, automating tedious tasks like data cataloging and enhancing data governance.



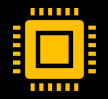
# DATA WAREHOUSING USING AWS RED SHIFT

- Utilizing Amazon Redshift, which provides high-performance analytics, as the main data warehouse.
- Combining various data sets from administrative, SIS, and LMS systems into a single repository to enable coordinated analysis.
- Redshift's speedy scalability to manage complicated queries and varied data volumes.
- Redshift receives data from the AWS S3 data lake with ease, guaranteeing data accessibility for analytics.
- Encrypting data and adhering to standards for educational data security.
- Facilitating the use of machine learning, predictive modeling, and complex data analysis.
- Real-time analytics to assist dynamic decision-making processes being facilitated.
- Storing massive amounts of data with Redshift's affordable storage options.

# STREAMLINING DATABASE MANAGEMENT WITH AMAZON RDS

- Amazon RDS simplifies database setup, operation, and scaling, offering a hassle-free database management experience.
- Compatible with popular databases like MySQL, PostgreSQL, Oracle, and SQL Server, providing flexibility in database choice.
- Ensures data safety with automated backups and applies patches and updates to the database software automatically.
- Easily scales database's compute and storage resources to meet the demands of the university's growing data needs.
- Offers Multi-AZ deployments for enhanced availability and built-in automated failover from primary to replicas.
- Seamlessly integrates with AWS services like Lambda for event-driven processing and EC2 for additional computing needs.

# API MANAGEMENT IN UNIVERSITY DATA ARCHITECTURE



### Central Hub

Amazon API Gateway serves as the gateway to the university's backend systems, enabling secure interactions between applications and services.

### Routing & Integration

It directs API requests to services like AWS Lambda for computation, Amazon EC2 for application hosting, or directly to Amazon S3 and Redshift for data retrieval.

### Workflow Orchestration

Manages complex workflows, triggering data processing tasks, and coordinating real-time data streams.

### Security

API Gateway secures API endpoints with user authentication, authorization controls, and traffic management.

### Monitoring

Works with AWS CloudWatch and IAM for detailed API usage tracking and secure access control.

### Unified Interface

Simplifies access for various university applications, ensuring smooth data sharing across departments.

### Outcome

A streamlined, secure, and scalable API ecosystem supporting the university's operational demands.

# SECURITY AND DATA GOVERNANCE IN CLOUD MIGRATION

- Implement Amazon IAM to ensure only authorized access to sensitive AWS resources.
- Utilize AWS KMS for robust encryption of data, both at rest and in transit.
- Adhere to FERPA, GDPR, and other educational data standards using AWS compliance tools.
- Regularly assess security and monitor compliance in real-time with AWS CloudTrail and AWS Config.
- Develop a disaster recovery plan with frequent data backups via AWS for data integrity and availability.
- Implement fine-grained data access policies with AWS Lake Formation for enhanced data governance.
- Ensure secure data sharing with adherence to governance policies across the university's departments and users.
- Conduct training and awareness programs for enhancing security and compliance culture among staff and students.


# COST OPTIMIZATION STRATEGIES

Make periodic modifications according to existing requirements; for resource optimization, use AWS Trusted Advisor.

For consistent workloads, use Reserved Instances; for less expensive, interruptible tasks, use Spot Instances.

Amazon S3 lifecycle policies for data management and effective network utilization.

Use Elastic Load Balancing and AWS Auto Scaling to optimize resource utilization.

Optimize AWS Glue and Lambda executions for serverless and affordable ETL tasks.

Utilize AWS consolidated billing and monitor expenses with AWS Cost Explorer.

Utilize the affordable SageMaker instances and assess the effectiveness of ML models.

Periodic audits to identify savings opportunities and update to more cost-effective technologies.

# PREDICTIVE ANALYTICS AND AI WITH AWS SAGEMAKER

- Develop, train, and deploy machine learning models using AWS SageMaker.
- Analyze university data on SageMaker for insights into student performance and operational efficiency.
- Utilize AWS Lex for natural language processing and Rekognition for image/video analysis.
- Implement Amazon Kinesis for immediate data streaming and analysis.
- Create custom machine learning models for specific university needs like personalized learning.
- Integrate AI services with AWS data storage and processing for efficient workflows.
- Leverage AWS's scalable infrastructure for evolving AI and analytics needs.
- Use advanced AI capabilities to empower research and improve teaching methodologies.

# HARNESSING GENAI FOR UNIVERSITY DATA MANAGEMENT

**Predictive Analytics Enhancement**

GenAI enhances predictive analytics for personalized learning and resource optimization.

**Advanced Data Processing**:

Integrates with AWS Glue for efficient data cleansing using GenAI pattern recognition.

**Content Personalization**:

Utilizes GenAI for generating customized online course materials.

**Security Enhancement with GenAI**

Employs GenAI in AWS IAM for dynamic threat detection and enhanced security measures.

**Resource Optimization**

Applies GenAI for efficient faculty scheduling, budgeting, and classroom allocation.

**Improved Engagement with AI Assistants**

Implements GenAI-driven chatbots and virtual assistants for real-time student and staff support.

# REFERENCES


- https://odsc.medium.com/choosing-a-data-lake-format-what-to-actually-look-for-edc6b87b01aa
- https://aws.amazon.com/aws-cost-management/cost-optimization/
- https://aws.amazon.com/big-data/real-time-analytics-featured-partners/
- https://docs.aws.amazon.com/wellarchitected/latest/framework/welcome.html
- https://docs.aws.amazon.com/glue/latest/dg/catalog-and-crawler.html
- https://docs.aws.amazon.com/whitepapers/latest/building-data-lakes/building-data-lake-aws.html
- https://docs.aws.amazon.com/sagemaker/