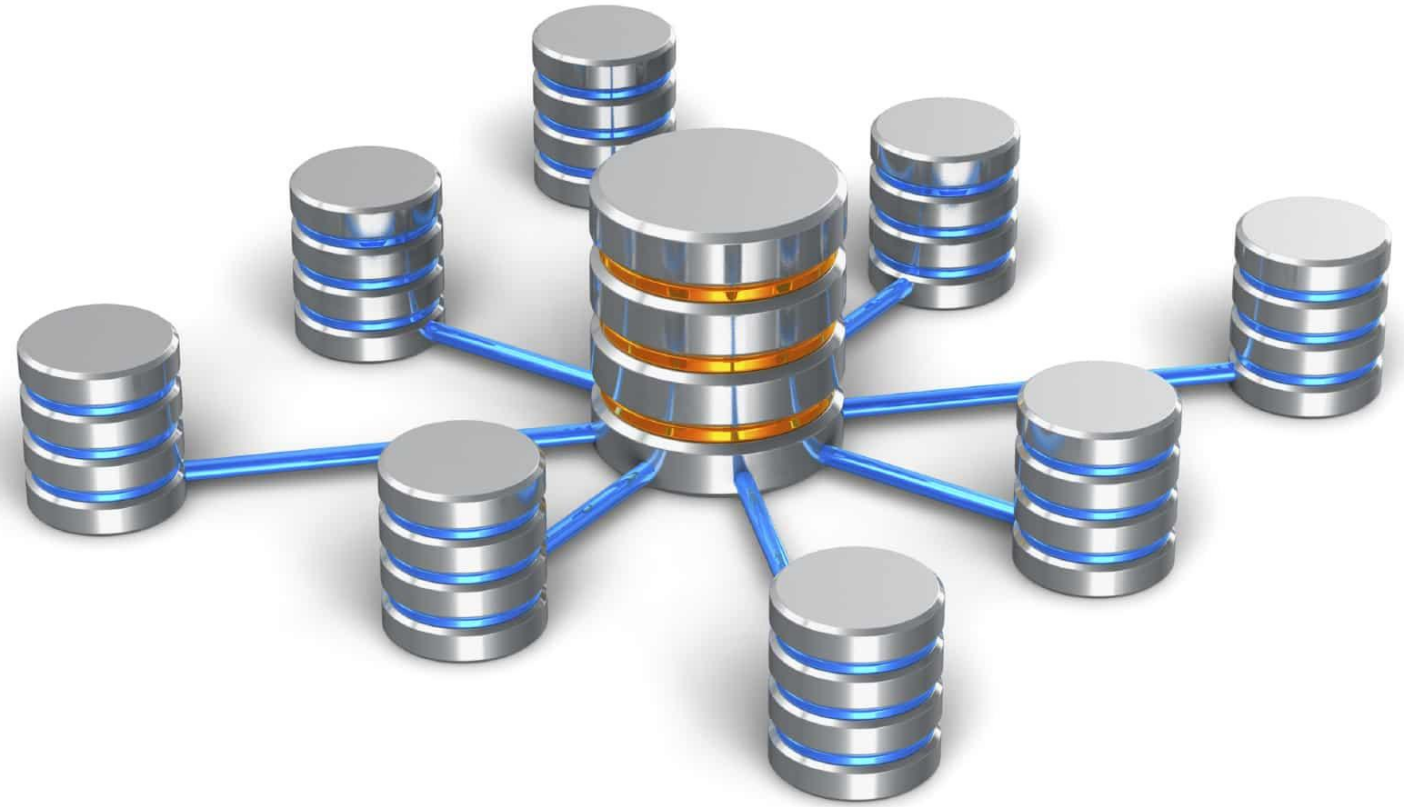
Unified Data Platform: University data management



Objective: A unified data-driven architecture

Common Data Platform

Seamless data integration
Single Source of truth
Master data management

Unified Data Store for analytics

Data analytics governance
Analytical data models
Predictive analysis using AI

Data sources

Batch processing:

Student Data

Faculty and Staff Data

Administrative Data

Course and Curriculum Data

Research Data

Library Data

Event and Calendar Data

Infrastructure Data

Learning Management System (LMS) Data

Real time processing:

Sensor Data (if applicable)

External Data

User Interaction Data

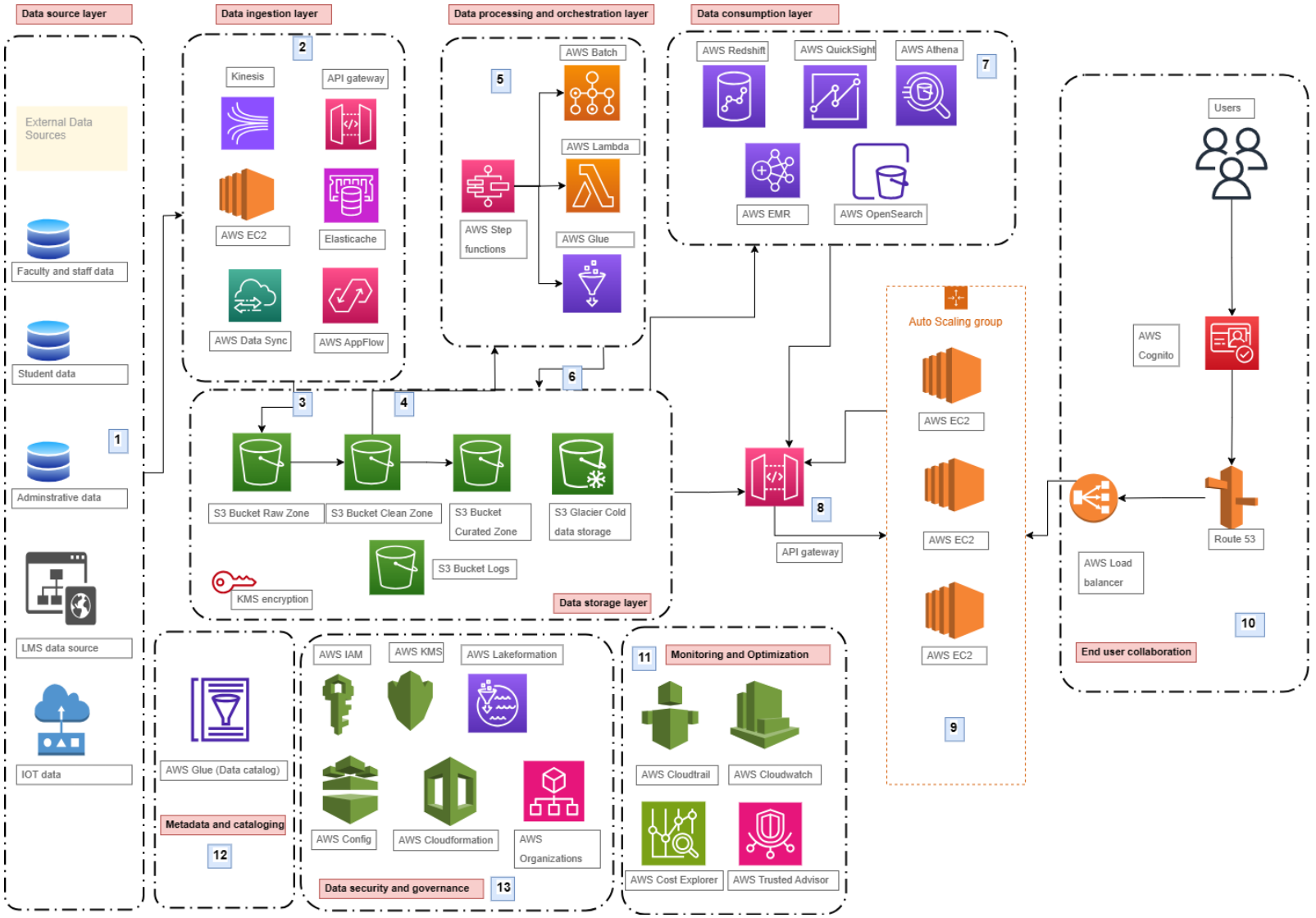
Communication Data

Biometric Data (if applicable)

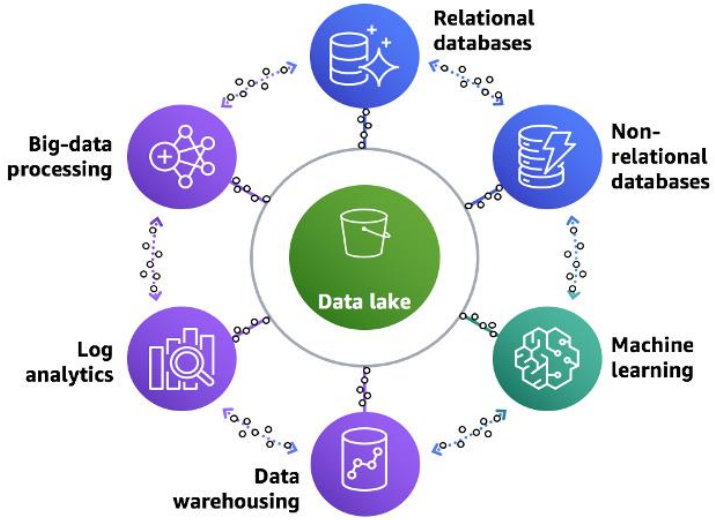
Solution roadmap

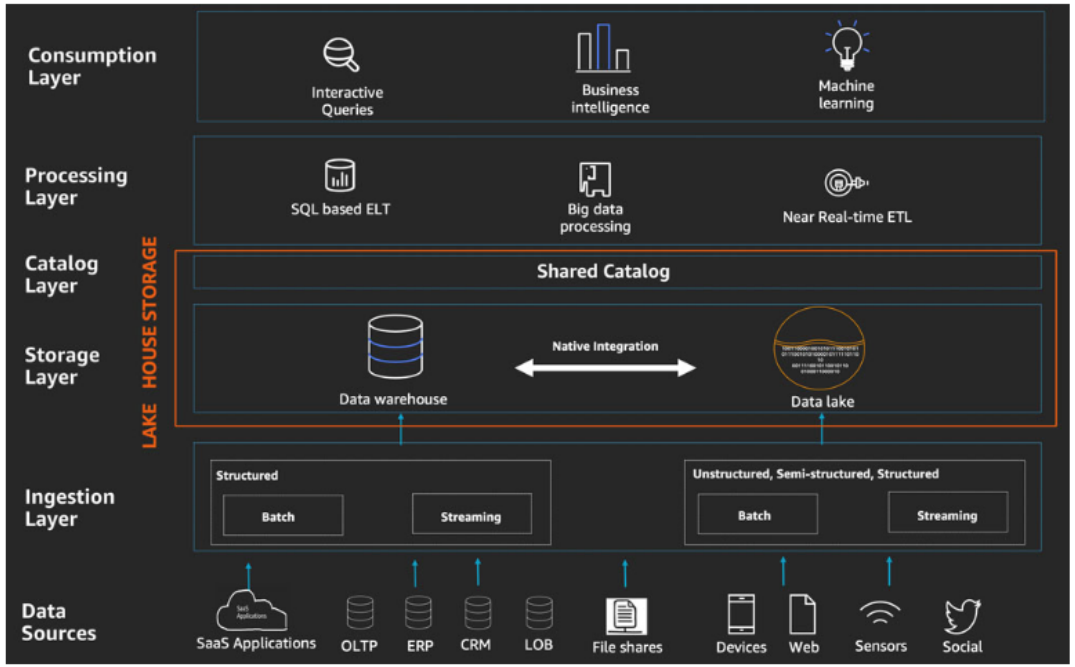
- A data lakehouse architecture
- Hybrid approach (Data sources partly on-prem)
- Cloud native solution with a data migration strategy
- Addressing bottlenecks
- Optimization solutions
- Implementation plan
- AI/ML use cases

HYBRID SOLUTION: ARCHITECTURE



The Data Lakehouse approach



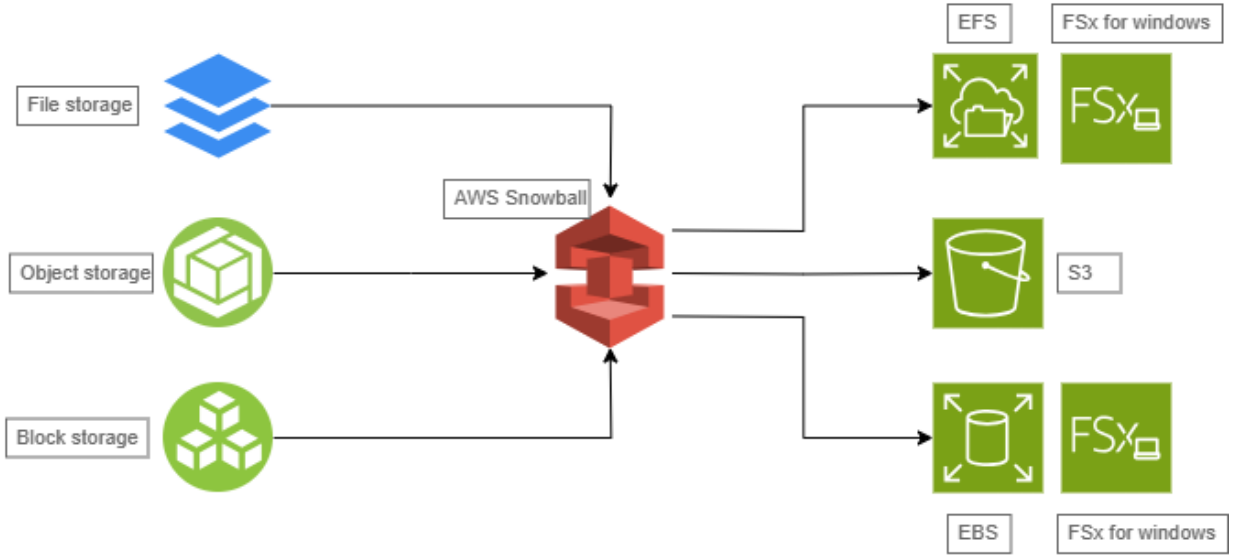


CLOUD SOLUTION: MIGRATION PLAN

Migration Strategy

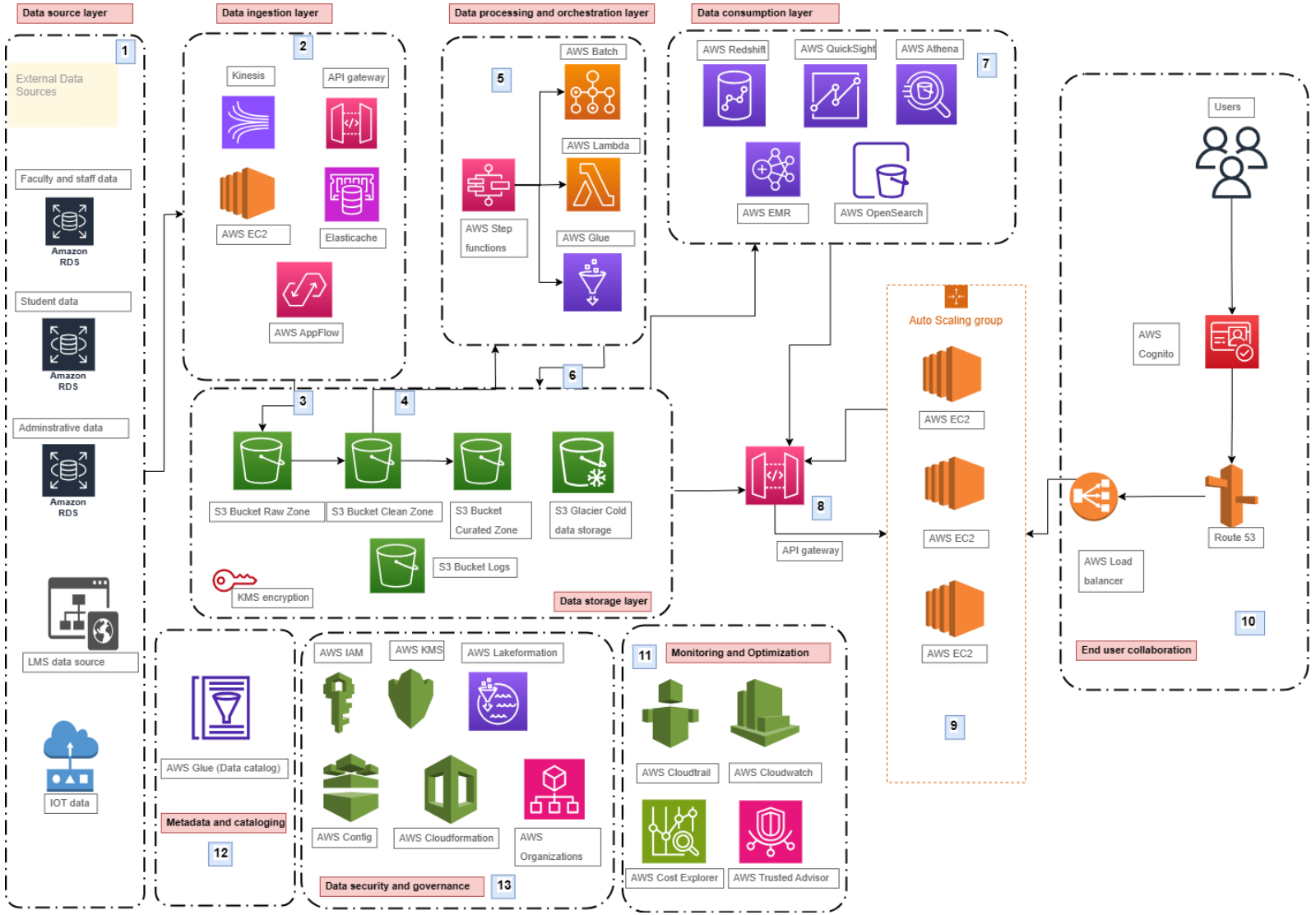
Step 1: Database Migration

Step 2: Extraneous Data Migration



Step 3: Complete Integration with Data Lakehouse

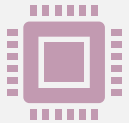
CLOUD SOLUTION : ARCHITECTURE



Addressing bottlenecks and optimizations

Bottlenecks:

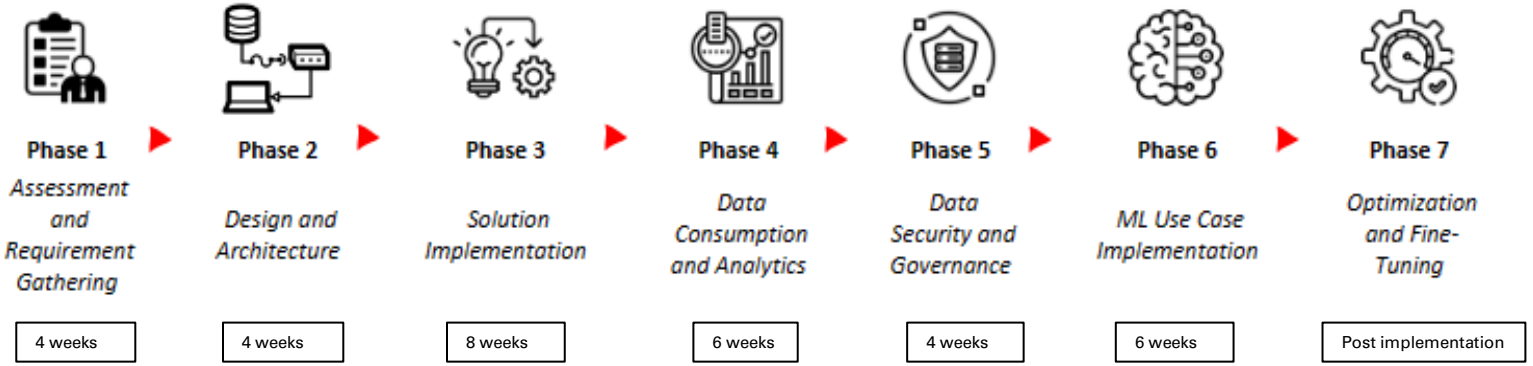
Horizontal Vs. Vertical scaling
Database partitioning and sharding
Caching strategies



Optimizations:

Multi-AZ deployment
S3 data life cycle policies
Intelligent pricing : S3 and EC2
CAP theorem

Implementation plan and cost estimation



Group	Monthly	First 12 months Total
Data Ingestion	$112,193.45	$1,346,321.39
Data Processing and Orchestration	$2,560.08	$30,720.96
Data Consumption Layer	$112,610.57	$1,351,326.84
Auto Scaling Group	$6.57	$78.84
End User Collaboration	$27,049.10	$324,589.20
Monitoring and Optimization	$1,692.08	$20,304.98
Data Security and Governance	$217.17	$2,606.04
Data Storage	$3,319.05	$39,828.60
Metadata and Cataloging	$2.96	$35.52
Total	**$259,651.04**	**$3,115,812.48**

AI/ML use cases

- Predictive Analytics for Student Success
- Alumni Donation Prediction
- Identifying students at risk (dropouts, failures, financial or similar hardship)
- Gauging effectiveness of staff and curriculum
- Segmentation for student services/counselor organizations
- Sales/marketing insights and alerts

Conclusion

- We started with a comprehensive evaluation of the university's existing data landscape by closely looking at how the university deals with data today, understanding data types, and the data sources.
- 2 kinds of solutions providing the university more flexibility in choosing the architecture
- The modern data strategy (Data lakehouse) forms the core of efficient data management, incorporating scalable data lakes, performance optimization, serverless architecture, unified data access, and machine learning integration
- Optimization Techniques and Enhancements to address bottlenecks
- Implementation plans and costs involved (first 12 months and monthly costs)
- The proposed solution provides transformative vision for the university's data infrastructure—a harmonious blend of existing reliability and the dynamism of cloud-based solutions. This sets the stage for a modern, scalable, and secure foundation, primed for future advancements in data management and analytics.

References

- https://aws.amazon.com/blogs/big-data/build-a-lake-house-architecture-on-aws/
- https://docs.aws.amazon.com/wellarchitected/latest/analytics-lens/batch-data-processing.html
- https://aws.amazon.com/caching/database-caching/
- https://aws.amazon.com/getting-started/decision-guides/analytics-on-aws-how-to-choose/
- Transitioning objects using Amazon S3 Lifecycle - Amazon Simple Storage Service
- Sharding with Amazon Relational Database Service | AWS Database Blog
- Horizontal scaling - AWS Well-Architected Framework (amazon.com)
- CAP theorem - Availability and Beyond: Understanding and Improving the Resilience of Distributed Systems on AWS (amazon.com)
- Data encryption - Amazon Redshift
- Encrypt Your Amazon Redshift Loads with Amazon S3 and AWS KMS | AWS Big Data Blog
- AWS Glue DataBrew now provides detection and data masking transformations for Personally Identifiable Information (PII) (amazon.com)
- Efficient Infrastructure Management with AWS CloudFormation | by Brandon Damue | AWS in Plain English
- https://aws.amazon.com/blogs/architecture/overview-and-architecture-building-customer-data-platform-on-aws/
- https://aws.amazon.com/solutions/case-studies/Universidad-Francisco-de-Vitoria-UFV-case-study/?did=cr_card&trk=cr_card
- https://aws.amazon.com/blogs/big-data/category/case-study/