Optimising Credit Card Onboarding for Reem Finance

ISIT 312 Project Overview

Muhammad Shaheer Kashif – 7877146
Mikaeel Faraz Safdar – 8074689
Muhammad Bisham Adil Paracha - 793540
Rabail Keshav Lal - 7778144
Varun Tulsiyani - 8044661

Overview

1. Introduction
2. Key Findings
3. Solution
4. Conclusion

INTRODUCTION

Executive Summary

Reem Finance, a leading financial services provider, manages extensive data on credit card customers, transactions, and rewards. However, reliance on static data sources limits scalability, operational efficiency, and the ability to extract real-time insights.

This project redesigns Reem Finance's data infrastructure by introducing a scalable data lake architecture combined with automated ETL pipelines. It integrates structured and semi-structured data from diverse sources, including customer details, transaction logs, and behavioral insights, ensuring consistency and accuracy.

The goal is to deliver actionable insights through interactive Power BI dashboards, enhancing decision-making, customer engagement, and risk management while supporting future scalability and regulatory compliance.

Project Objectives

	Objectives
1	Build a scalable and fault-tolerant data storage architecture.
2	Automate data cleaning, processing, and integration pipelines.
3	Generate actionable insights into volunteer engagement and event performance.
4	Simulate and test the system's ability to handle larger datasets.

Leading Questions

	Questions
1	Data Volume?
2	Collection Methods?
3	Storage Systems?
4	Processing Pipelines?
5	Integration Challenges?
6	Potential Issues?

Existing architecture

Current State	
1	Data collected manually and through CRM forms
2	Government API integrations provide structured data.
3	Semi-structured and unstructured data stored in separate systems (e.g., Document Management System for PDFs).

Limitations	
1	Data silos prevent real-time insights.
2	Lack of scalability for growing data volumes.
3	Manual processes increase operational inefficiencies.

KEY FINDINGS

Data Volume?

#	Insights
1	Significant data generated during onboarding, including **structured**, **semi-structured**, and **unstructured** datasets.
2	Example: Government API integration (e.g., **EFR**) provides **73 fields** per customer, covering name, date of birth, nationality, employment details, and residence status.
3	High volume during digital campaigns, especially targeting Gen Z and Millennials via platforms like **Instagram**, **LinkedIn**, and **X**.

Collection Methods?

#	Insights
1	**Digital Campaigns**
	Target-specific audiences on social media platforms based on demographics and psychographics (e.g., gamers, travellers). First data points captured: Name, mobile number, email, and product interest via CRM forms.
2	**Government APIs**
	Data from entities like EFR, Al Etihad Credit Bureau, and MOHRE for biometric verification, credit history, and employment records.
3	Customer inputs include biometric selfies, OTP verifications, and consent for data retrieval.

STORAGE SYSTEMS

#	Insights
1	**Digital Warehouse**
	Structured, field-level data (e.g., customer profiles, transaction summaries).
2	**Government APIs**
	Stores unstructured data such as scanned PDFs of bank statements.
3	All data linked via a **unique customer identifier (CIF)** to ensure seamless integration and accessibility.

PROCESSING PIPELINES

#	Insights
1	**Automated workflows validate, transform, and integrate data using APIs:**
	E.g., salary details cross-verified across four sources (EFR, bank statements, credit bureau, and MOHRE).
2	**OTP Validation**
	Ensures accuracy of contact details.
3	**Credit Decision Engine**
	Automatically calculates credit limits based on salary, credit score, and existing liabilities.

Integration Challenges?

#	Insights
1	**Schema Inconsistencies**
	Combining structured and semi-structured data from APIs and manual inputs.
2	**Privacy Compliance**
	Strict UAE regulations require anonymization of Emirates ID and other sensitive fields.
3	**Real-Time Processing**
	Challenges in synchronizing multiple data sources, especially during high-traffic campaigns.
4	**User Experience**
	Ensuring minimal input while maximizing automated data retrieval for a seamless onboarding journey.

Potential issues

#	Challenges	Insights
1	**Data Quality**	Missing or inconsistent data can lead to inaccurate credit decisions.
2	**API Dependency**	Reliance on third-party APIs (e.g., EFR, credit bureau) can cause delays.
3	**Real-Time Processing Limitations**	Interruptions in data streaming may impact dashboard updates.
4	**Security and Compliance**	Ensuring UAE regulatory compliance for sensitive data (e.g., Emirates ID).
5	**Scalability Challenges**	Growing transaction volumes may strain infrastructure over time.

SOLUTION

Main construct

#	Feature
1	Data-Driven Demand Forecasting
2	Interative Dashboard for Insights and Alerts
3	System Integration
4	Scalability and Adaptability
5	Competitor Analysis-Driven Enhancements

Overcoming Problems

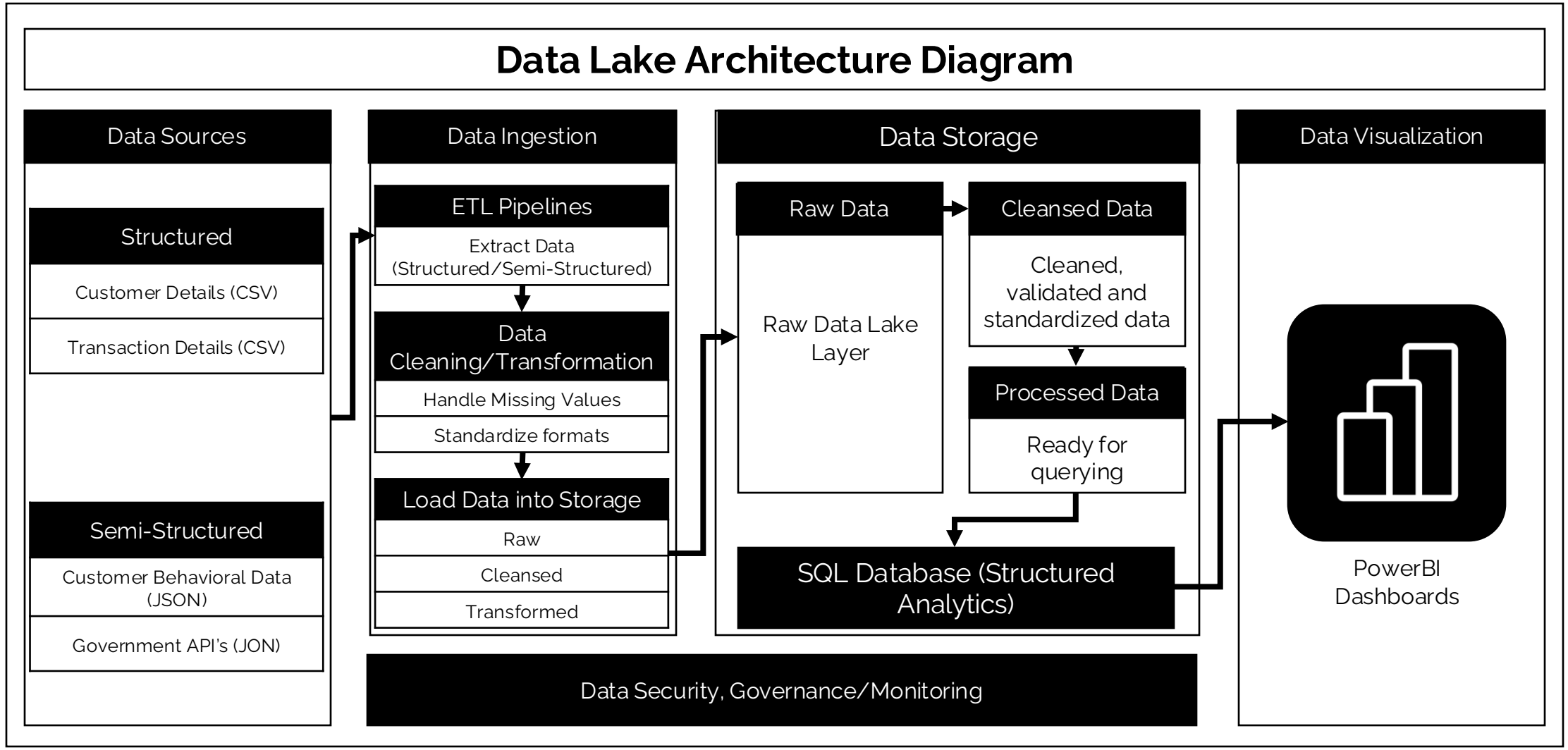
#	Challenges Addressed
1	Data silos across systems.
2	Manual workflows causing inefficiencies.
3	Real-time data processing limitations.

#	Solutions
1	Automated ETL pipelines for data validation and transformation.
2	Real-time data streaming implemented using **Azure Event Hub** for seamless ingestion.
3	Centralized data storage in Azure Data Lake with Raw, Cleansed, and Processed partitions to ensure scalability and structure.

Tech stack

Layer	Tech
Ingestion	Azure Data Factory (for data ingestion pipelines)
Storage	Azure Data Lake Storage (centralized storage for raw, cleansed, and processed data)
Metadata	Azure Data Catalogue (managing metadata for datasets)
Processing & Transformation	Python (data cleaning and ETL workflows), Azure Synapse Analytics (for data processing)
Data Governance & Security	Azure Purview (for data governance), Role-based access control (RBAC) in Azure
Analytics & Query	Power BI (data analysis and querying), Azure SQL Database (for structured queries)
Consumption	Power BI Dashboards (for interactive insights)
Monitoring & Optimization	Azure Monitor and Log Analytics (for monitoring pipelines and system performance)

High-Level Architecture

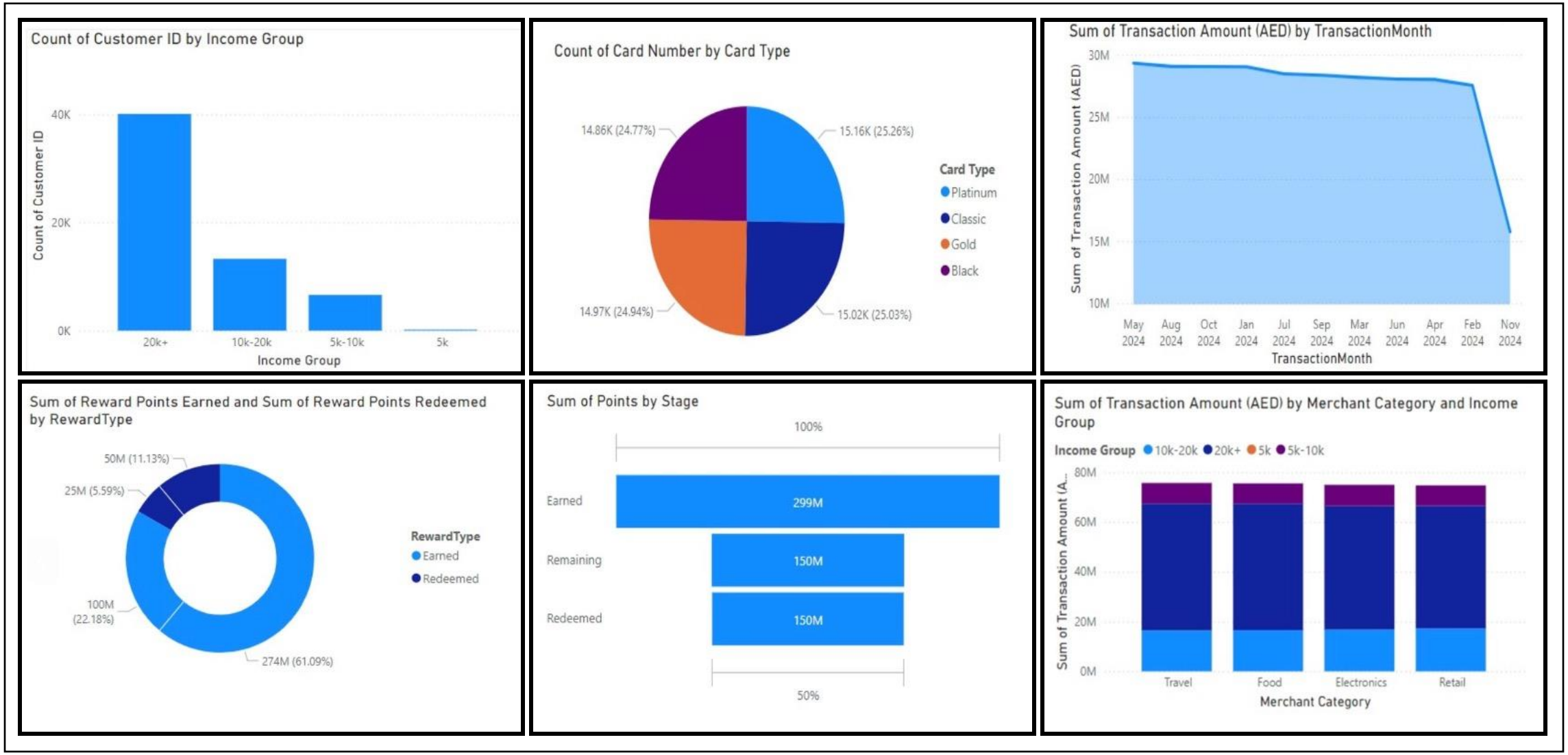


New vs Old Architecture

#	Old Architecture
1	Data stored in silos, requiring manual validation.
2	Limited scalability for growing data volumes.
3	No real-time data processing capabilities.

#	New Architecture
1	Centralized data lake for seamless data integration.
2	Automated workflows ensure faster and error-free onboarding.
3	Real-time insights enabled by streaming tools.

POWER BI ANALYTICS



PowerBI Analysis

Key Dashboards and Insights:

- **Customer Demographics:**
  - Visualization: Bar chart for income group segmentation.
  - Insight: Over 60% of customers belong to the highest income group (20K+ AED).
- **Credit Card Analysis:**
  - Visualization: Pie chart displaying card distribution by type.
  - Insight: An even distribution of Classic, Platinum, Gold, and Black cards highlights customer diversity.
- **Transaction Trends:**
  - Visualization: Line graph for transaction amounts over time.
  - Insight: Seasonal spikes observed during holidays (November–January), indicating high-value periods for targeted marketing.
- **Rewards and Benefits:**
  - Visualization: Doughnut chart for earned vs. redeemed reward points.
  - Insight: 60% of reward points remain unredeemed, signaling opportunities to improve customer engagement.
- **Spending Patterns:**
  - Visualization: Stacked bar chart of transaction amounts by merchant category and income group.
  - Insight: High-income groups spend the most on travel and retail, identifying premium service opportunities.

Conclusion

Conclusion

Successfully implemented a scalable data lake architecture for Reem Finance's onboarding process.

- Automated ETL pipelines to unify data from CRM systems, government APIs, and unstructured sources.
- Integrated Power BI dashboards to provide real-time insights into customer demographics, onboarding metrics, and credit decisions.

Key Outcomes

- 50% reduction in onboarding time, improving customer satisfaction.
- 90% decrease in manual errors, enhancing data quality and decision accuracy.
- System validated to handle 10x data growth, ensuring future scalability.
- Fully compliant with UAE data privacy regulations through secure storage and encryption.

Business Impact

- Enhanced operational efficiency and regulatory compliance.
- Improved customer segmentation and targeting through actionable insights.
- Empowered data-driven decision-making for credit approvals and risk assessment.

"Our solution not only resolves current challenges but also positions Reem Finance as a leader in digital banking innovation. Thank you!"