Business Context

Home Financial Services is a leading lender offering various loan products to clients, including:

- Cash loans
- Consumer Durable (CD) loans
- Two-Wheeler (TW) loans

Each loan product has a unique product code and specific terms, such as interest rates, fees, and repayment period (number of installments).

When a client applies for a loan, a contract is created with a unique contract number. The contract outlines:

- Principal Amount: The loan amount borrowed.
- Interest Amount: The interest charged.
- Fee Amount: Additional charges, like administrative or processing fees.
- Tenor: The number of installments for repayment.

Loan contracts progress through several stages:

- New: Contract created but not activated.
- Active: Loan in repayment with ongoing installments.
- Paid-off: Loan fully repaid.
- Written Off: Loan defaulted due to non-payment.
- Sold: Loan transferred to a third party for recovery.
- Cancel: Contract terminated before activation.
- Close: Contract finalized with no further action needed.

Clients repay loans in monthly installments, each with a due date. If an installment is paid after the due date, a penalty amount is applied to the loan.

The management of Home Financial Services is focused on analyzing and optimizing its loan portfolio. They are particularly interested in the following areas:

- Loan Contract Analysis by Region: Understanding how loans are distributed and performing across different geographic regions of client.
- Client Default Risk Analysis: Using historical repayment data to assess whether a client is likely to default on their loan. This involves analyzing payment history to identify patterns such as late payments, skipped payments, or early settlements, which help in predicting future defaults.

This item's classification is Internal. It was created by and is in property of the Home Credit. Do not distribute outside of the organization.

Data Engineering Technical Assessment

1. Data Modeling

Using the provided business context and data in the data.zip file, design a data warehouse model to support:

- Loan Distribution Analysis by Region
- Client Default Risk Analysis

2. Data Pipelines

Design and develop an end-to-end data pipeline to implement your data model using your chosen technologies. The pipeline should:

- Ingest raw data from source files with minimal transformations.
- Process and transform data into curated datasets aligned with business needs, ensuring quality through:
    - Deduplication
    - Handling missing values
    - Validating data types
- Bonus:
    - Add error handling for data quality issues.
    - Implement schema evolution to manage changing source structures, documenting versioning and compatibility.
    - Apply the Write-Audit-Publish (WAP) pattern.

Requirements

- Use open-source technologies.
- Ensure idempotency.
- Support backfilling historical data.
- Follow software engineering best practices (design pattern, testing, documentation).

3. Data Operations

Design a platform to support daily pipeline operations:

- Provide a basic UI for scheduling, monitoring, and restarting jobs.
- Enable data lineage tracking to visualize data flow.
- Implement data quality checks with alerting.
- Bonus:
    - Containerize the solution for deployment across environments (local, dev, prod) with CI/CD.

4. MLOps Solution (Bonus)

Briefly describe how your data processing solution could support ML workflows:

- Feature Engineering Pipeline and Feature Store
- Model Training Infrastructure
- Model Serving & Deployment

This item's classification is Internal. It was created by and is in property of the Home Credit. Do not distribute outside of the organization.