Case 1 – Data Model Design

1. Introduction:

In this project, the bank is implementing a digital process for customers using eKYC (electronic Know Your Customer) technologyThis process involves customers going through the onboarding process, creating a digital account, and making transactions. To effectively track the progress of the eKYC process, as well as monitor the customer's account creation and transaction behaviour, we require a robust data warehouse solution.

2. Entities:

2.1. Customers:

Represents digital individuals who are engaging with our banking services. With basic customer information such as name, contact details, identification documents, and eKYC status.

2.2. Accounts:

Represents the accounts created by customers during the digital onboarding process → Category 1001, and if the customer has done KYC at branch → upgraded to 1002. With account details such as account number, category, creation date, and associated customer information

2.3. Transactions:

Represents the digital financial transactions conducted by customers. With Transaction details such as transaction ID, timestamp, amount, type, and associated account and customer information.
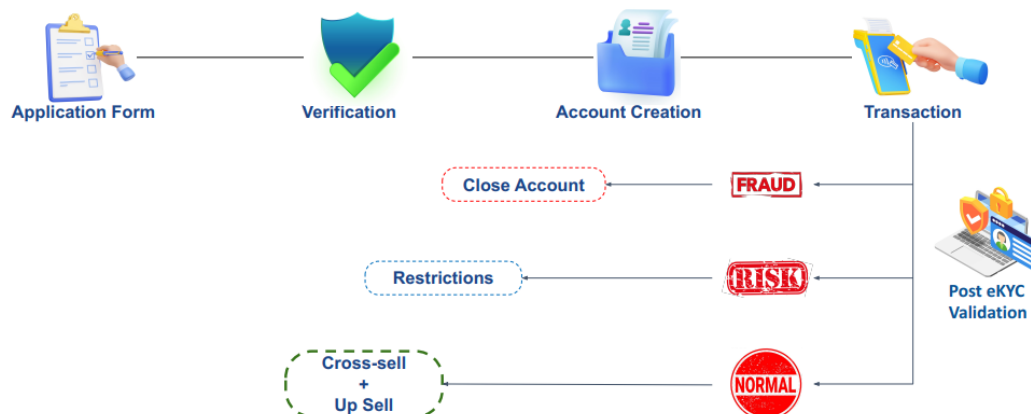
3. Customer journey analysis:

Figure 3-1: Customer journey.

3.1. Application Form:

First, customers download the bank's application and fill in all the initial required information fields. Customers can choose the document they want to verify, such as Citizen Identification Card/ Passport. Then, they take clear photos of both sides of the document.

3.2. Verification:

At this step, customers will perform eKYC according to the following process:
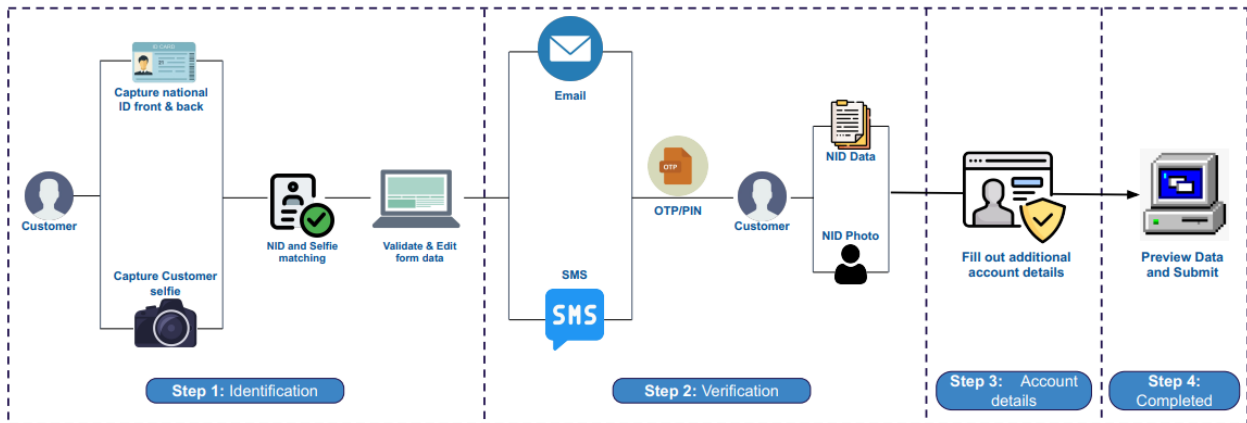
The algorithms will compare the Citizen Identification Card/ Passport photos taken by the customer with the information they have edited to check for consistency.

In this process, the data warehouse will track the status of the eKYC process for each customer, including collecting information about the eKYC verification progress, such as pending, verified, or rejected. Capturing the eKYC status must be timestamped to maintain a historical record.

3.3. Account Creation:

After the successful verification process, the customer will have their account created. During this process, the data warehouse will track whether the customer has successfully created an account. This information must be updated in real-time and include details such as the account creation time, linked account information, and customer details.

3.4. Transaction:

Currently, at most banks, once customers successfully create an account, they are usually allowed to perform basic transactions immediately. However, these transactions will be limited to a certain threshold.

The data warehouse will track customer transaction behavior. This involves capturing the transaction time, transaction details, and linking it to the relevant customer and account

3.5. Post eKYC Validation:

In this process, the bank will review customer information. After the review, the outcome will determine whether the customer is classified as Fraud, Risk, or Normal.

4. Data sources:

Source	Description
Onboarding eKYC verification system	Collect information about the customer's eKYC process.
Core T24 account system	Collect information about the customer's accounts and transactions.
Operational transaction system	Collect information about the customer's transactions.
CRM system	Collect and tracking the customer's interactions with the application.
Post-eKYC tracking	These are manual Excel files provided by the operations team.

Table 4-1: Data source description.

5. Data warehouse design:

For easier management, I use a 3-tier architecture for the Data Warehouse to separate and organize the main components of the system. This architecture includes three layers: Source layer, Reconciled layer, and Data Warehouse layer.
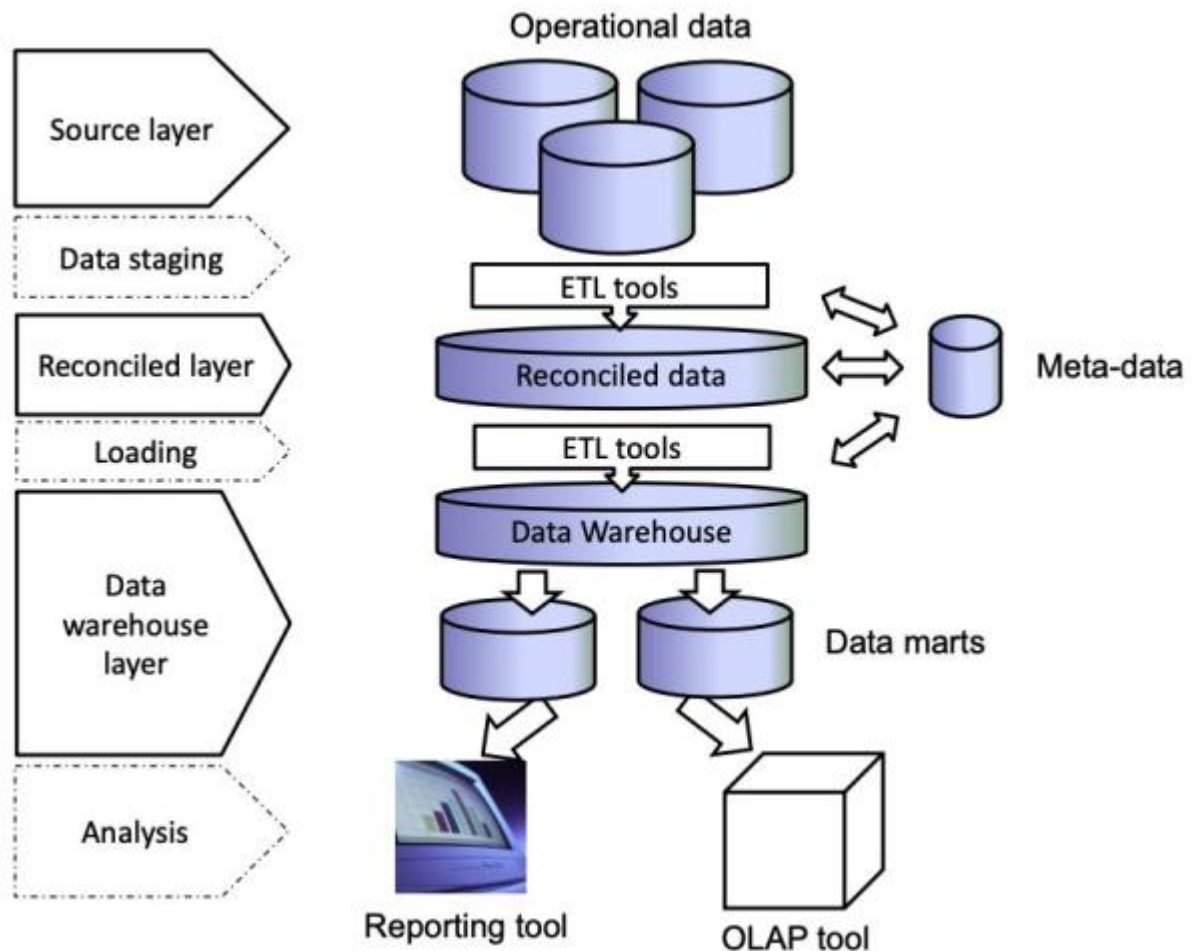
Figure 5-1: Data warehouse structure.

5.1. Sourse layer:

You can refer to the data source description in Table 4.1. This layer includes a data warehouse server. The task of this layer is to collect, clean, and transform data from various sources.

You can refer to the data source description in Table 4.1. This layer includes a data warehouse server. The task of this layer is to collect, clean, and transform data from various sources.

5.2. Reconciled layer:

The data from the Source layer, after undergoing the Data Staging process where it is filtered according to the Business Analyst's (BA) requirements, will be moved to the Reconciled layer.

5.3. Data warehouse layer:

Including querying tools, reporting, analysis, and data mining.

6. Build Data warehouse Model:

Table	Description
DIM_ACCOUNTS	Customer account information
DIM_CRM	Customer information on CRM system
DIM_EKYC	Customer eKYC process infomation.
DIM_POST_EKYC	Customer posrt eKYC process ìnomation
DIM_TRANSACTION_TYPE	Transsaction type infomation
DIM_TRANSACTIONS	Transaction information
FACT_DIGITAL_PROFILES	Customer profile information

Hình 6-1: Data warehouse table information.