FINAL ASSIGNMENT

ADVANCED DATABASE

GROUP 5

ABHINANDAN SINGHI (25155716)
DEEP PATEL ()
VIPRA PATEL()
FANG FANG ()
PRATIK CHIKHALI (25409640)

Table of Contents

Executive Summary

Vodafone also has its telecommunications infrastructure designed around three autonomous subsystems including prepaid mobile, postpaid accounts, and broadband services that are served by respective customer databases. Such siloed architecture has created serious issues of operations, especially, duplication of customer records, billing disparities, and the lack of capacity to monitor cross-system service uptake. The analysis of the data showed that there are 57 customer records, which represent a total of 28 users, meaning that the duplication rate is 51 per cent directly affecting the accuracy of reporting, customer management, and marketing performance.

To alleviate these shortcomings, the project implemented a data warehousing solution which is a Snowflake enterprise-based cloud system. The warehouse uses a Kimball star schema, which connects customer and service data with the help of a strict ETL pipeline standardising, cleaning, and consolidating data of the three dissimilar systems. Through the use of government-issued license numbers as unique identifiers, the solution would have a 100^{-} accurateness in duplicate recognition and generate consolidated golden customer records, thus permitting one and authoritative portrayal of each customer in all service modalities.

The implementation provided short term business value. It found an uncredited balance at pre-paid transaction of 270 and 4 instances of duplicated billing as well as 12 customers using all the three Vodafone services. These lessons have become the foundation of focused cross-selling initiatives, customer relationship segmentation, and retention based on data.

Generally, the project demonstrates how a complete data warehouse can overcome data silos, improve the analytic faithfulness and offer actionable insights that bolster the assurance of income and customer involvement - delivering a scalable framework on the continuous digital transformation at Vodafone.

Introduction

Vodafone is a leading telecommunications company that offers consumers and businesses integrated mobile, broadband and data services. Traditionally, the three service divisions of the provider, which are prepaid, postpaid and broadband, have developed through autonomous operating systems with each having segregated customer databases. Such a fractured system has spawned material problems in customer information control, analytics accuracy, and the creation of general business understanding. (Chaudhuri & Dayal, 1997)

The dominant architecture breeds various inefficiencies and operation hazards. The result of the lack of a consolidated customer master repository is high levels of record duplication, skewed customer lifetime value estimation, and poor marketing segmentation. An example is a customer such as John Smith that appears in twelve different records with the three systems which undermines the quality of reporting and personalization of services. Moreover, billing irregularities have resulted in revenue leakage where those transactions that are not reconciled recharge and the billing incidents that occur two times have led to financial shortages and loss of customer confidence. The absence of consolidated visibility further does not allow Vodafone to define multi-service or so-called triple-play customers who use multiple services simultaneously and limits cross-selling efforts and customer retention plans. Any effort to match customers with email identifiers has been found to lead to a 70-80 per cent match, which would make consolidation efforts more difficult.

To deal with these problems, this project deploys a data warehouse on a cloud environment on the Snowflake platform to consolidate customer and transactional data of all service systems. The solution uses the government issued license numbers as a unique identifier to removes the duplication and generate single golden records of customer records with a 100 percent matching accuracy. The design is based on a Kimball-style dimensional model that includes two dimensional table (Customers and Services), and one fact table (Financial Transactions).

The project provides four major analytical capabilities: (1) 360 customer view allows coherence and harmonisation of financial differences; (2) revenue leakage opportunities through reconciliation of financial differences; (3) single-service customers cross-sell opportunities; and (4) customer value segmentation through targeted retention programs. Three source systems, including prepaid, postpaid and broadband, are all extracted, transformed, and loaded into Snowflake staging and warehouse layers. (Eifrem, 2020)

The prototype explains the strategic value of integrated data management in the telecommunications industry. The solution provides the basis of a scalable enterprise data architecture that will fulfill the long-term digital transformation goals of Vodafone by facilitating the correct identification of customers, better revenue assurance, and increased insight into analytics.

Problem Statement

The telecommunications infrastructure at Vodafone is currently being provided by three autonomous systems such as pre-paid mobile, post-paid, and broadband, all of which are keeping separate databases of customers. Although these systems initially were designed to address various requirements of operations, due to their deficiency of integration, they have ended up producing a siloed data architecture which poses significant business and analytical problems. Decentralisation of customer data through systems makes the organisation unable to define a single source of truth resulting in reporting errors, duplication and inconsistency of decision-making processes across departments. (Nanavati et al., 2008)

Customer duplication is one of the most important issues among the three platforms. It was found that one of the customers, John Smith, is included on twelve different records, ten of which are in the prepaid database, one in the postpaid and one in broadband, which reflects one person being recorded as twelve separate customers. This duplication increases the total number of customers and renders it impossible to estimate correct customer lifetime value. It is also a cause of marketing segmentation distortion, diminished target campaign efficacy, and customer service operation complications. Wrong records or duplicate records lead to delays in customer problems and also bring about inappropriate distribution of marketing spend and wastage of resources.

The other major issue has to do with billing discrepancies both within and between the systems which translate directly to the quantifiable revenue leakage. Internal audit established eight prepaid recharge transactions amounting to $270 that have been processed yet the respective customer accounts have not been credited and this is considered a clear financial loss. Also, 4 cases of repeated billing were reported in the postpaid system and this posed the risk of overcharging the customers, creating dispute and possible service cancellation or customer churn. These problems point out severe gaps in the financial reconciliation procedures and indicate how the absence of a standardized data platform impairs the process of guaranteeing revenues at Vodafone.

Another bottleneck comes about due to lack of cross-system visibility on service adoption. Due to the independent operation of the prepaid, postpaid, and broadband systems, the business intelligence teams of Vodafone are unable to track customers utilizing multiple services. This is not integrated to identify triple play customers, those who buy all three types of services, and thus can never use customized promotions or loyalty programmes. Therefore, cross-selling opportunities have not been utilized, and even high-value customers have the same engagement level as single-service consumers.

Solution Design

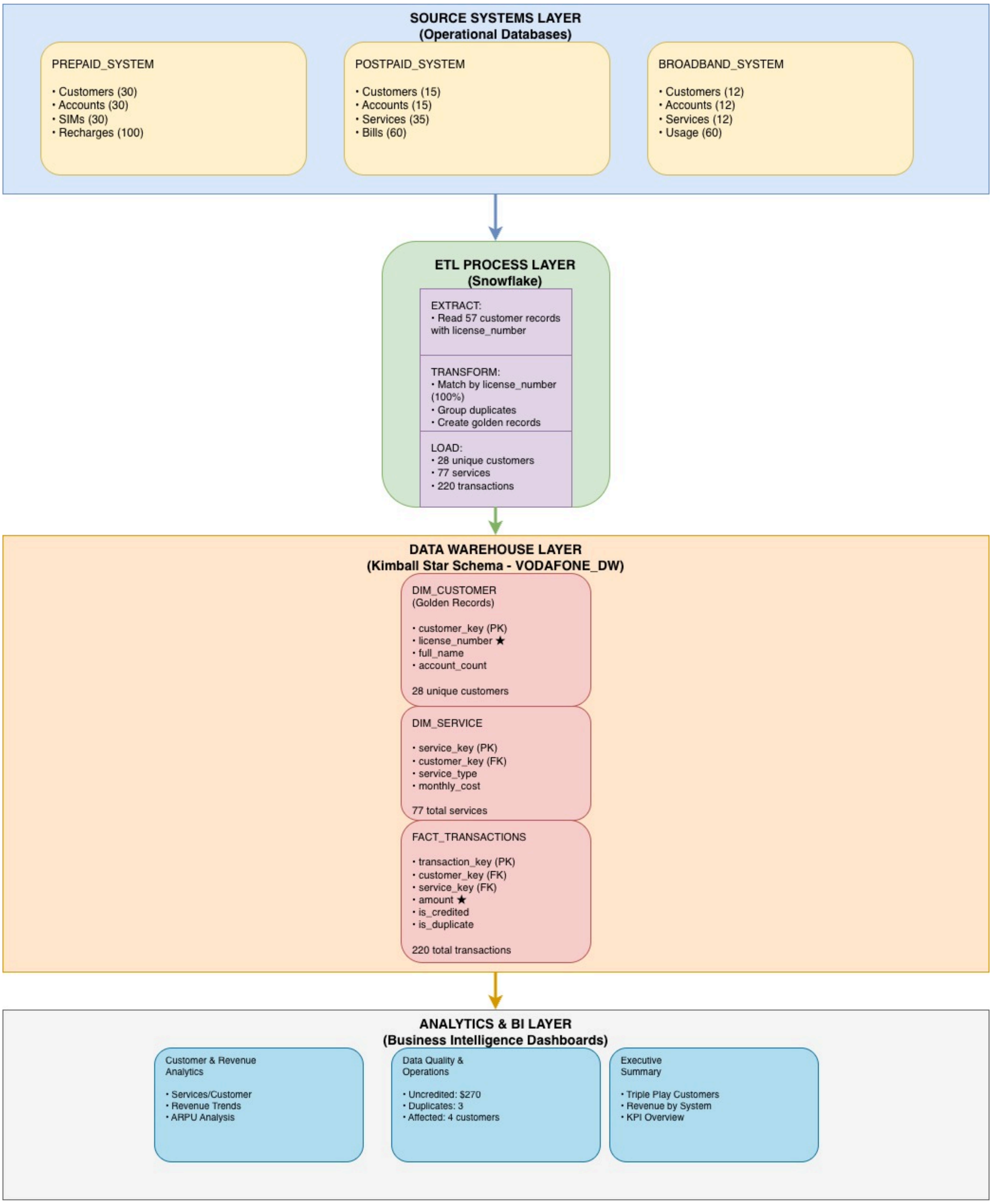
Solution Architecture

The solution architecture utilizes a 3-layer data warehouse architecture that is optimized in the telecommunications customer analytics. The source systems layer has three independent functioning databases that depict prepaid, postpaid, and broadband business units. Every system holds the customer master data, service subscriptions and the records of transactions formed in accordance to historical business requirements and not to the analytical requirements.

The integration layer is a processing layer that uses Extract-Transform-Load with customer matching using the license number. ETL logic can be used to consolidate duplicated customer records based on the duplication of the entries in the source system with the same government-issued license number and form a match group which identifies a unique individual. The staging area enables the evaluation of the quality of data and transformation logic, which is loaded to the production warehouse tables. (Golfarelli & Rizzi, 2009)

The data warehouse layer is a layer that uses the star schema architecture to implement the dimensional modeling approach developed by Ralph Kimball. The design centers on two dimension tables—DIM_CUSTOMER for golden customer records and DIM_SERVICE for all telecommunications services—with one central fact table FACT_TRANSACTIONS capturing all financial events. This design to maximize query performance to expectant business intelligence trends and data integrity by surrogate keys and foreign key associations.

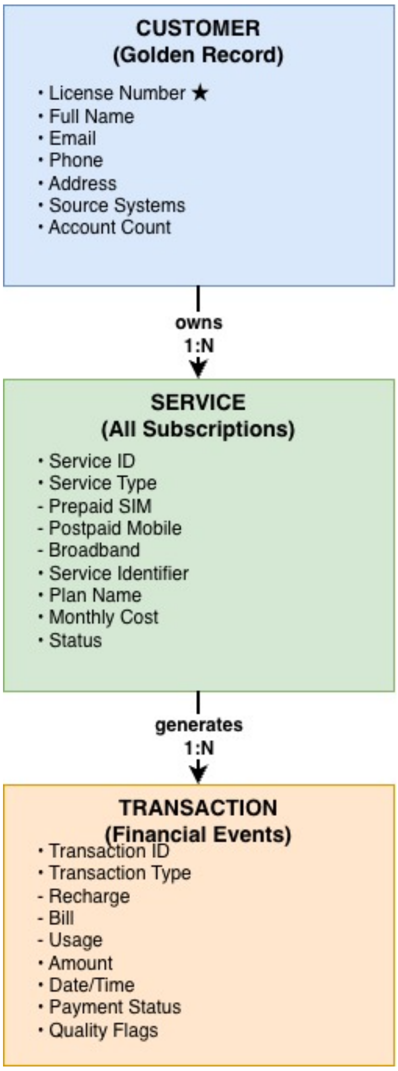
The analytics layer provides four business intelligence application use cases based on SQL queries used to visualize dashboards. The queries can be used to provide efficient aggregation and filtering using the star schema design and provide real-time analytical information without affecting the performance of the source system. Separating the workloads of analysis and operation systems is one of the basic data warehouse architecture principles.



System architecture Diagram

Conceptual Model of Data Entities

The concept data model has three business entities and relations in a telecommunications customer management. The CUSTOMER entity is used to denote the unique individuals or organisations that subscribe to the Vodafone services and they are referred to by license numbers issued by the government and make them unique in all the systems. Important features are the demographic data, contact data and source system tracking to have data lineage. (Filipovic, 2022) The SERVICE entity is a direct depiction of customer owned subscriptions of telecommunications. Services also include prepaid SIM cards, postpaid mobile contracts, and broadband internet. The license number matching process connects each service to one customer and individual customers can own numerous services in various categories. Attributes of service store subscription information, the activation date, indicators of status and monthly recurring fees. TRANSACTION entity documents all financial activities relating to services such as prepaid recharges, postpaid billing and charges against broadband use. Transaction is connected to customer and service entities, thus allowing revenue analysis to be done thoroughly. The most important attributes are the type of transactions, their amount, status of payment and data quality flags to identify unusual billing.



Conceptual Entity Relationship Diagram

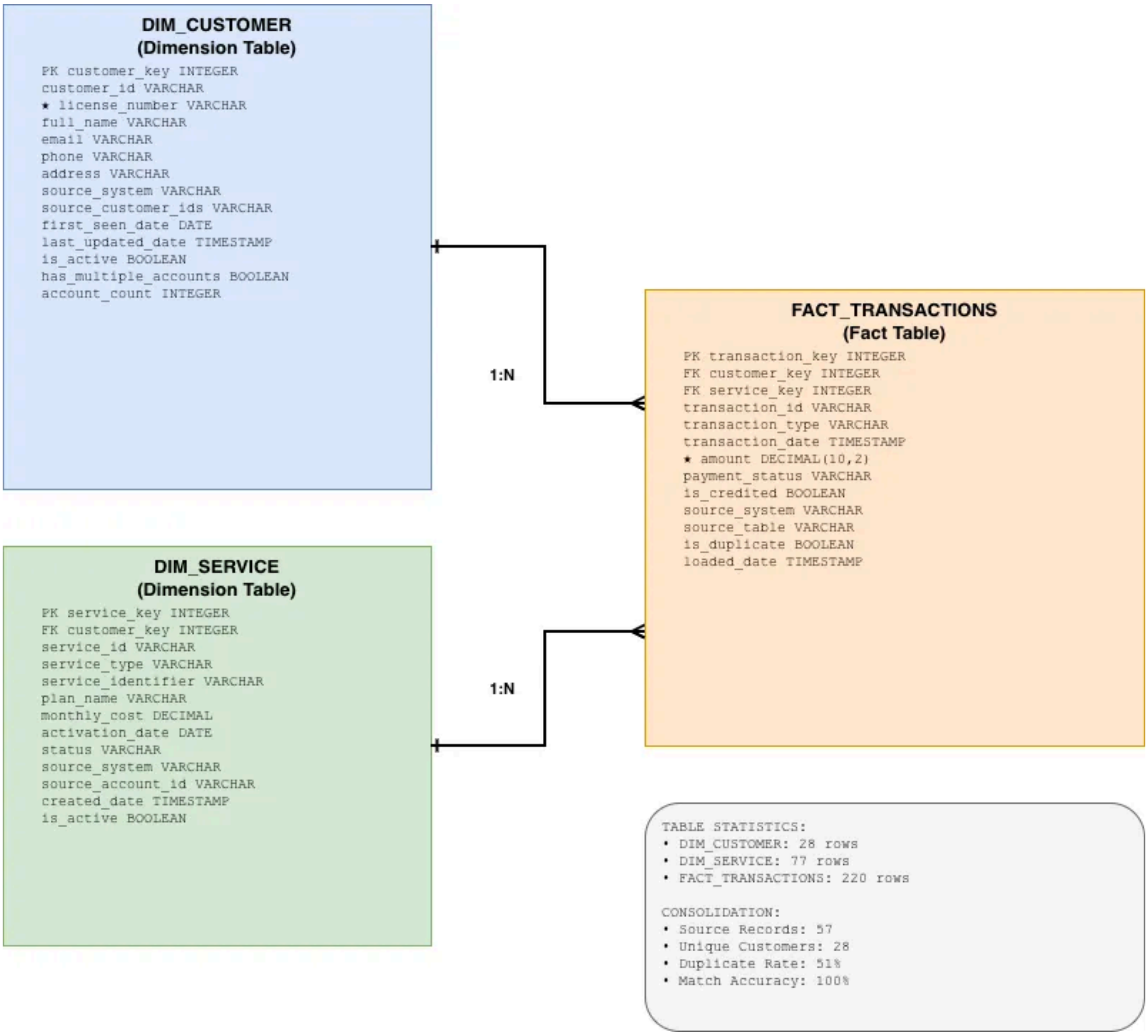
Logical Data Model of Data Entities

The logical data model is a database scheme transformation of conceptual entities, in Kimball dimensional modeling concepts. The star schema design produces the best combination of performance on analytical queries with referential integrity (surrogate keys).

DIM_CUSTOMER is the main dimension table, which has golden records of customers. The surrogate key (customer key) is a key of auto-incrementing type, which is used as a primary key within the table, separating warehouse records and source system identities. The attribute license number gives the business key used in identifying and checking duplication of customers. The source customer identifier attribute is a comma-separated list of original source system identifiers, which retains data ancestry. The presence of multi-accounts and the number of accounts as the values of the flags and the number of accounts as the integer, respectively, are considered to be the examples of the Boolean flags that measure the effects of the consolidation and help to analyze the data quality. (Knezevic, 2022)

DIM_SERVICE slowly changing dimension that stores all telecommunications services across platforms in DIM service. Surrogate key method allows following the changes in service throughout the time and has the historical correctness. The customer key as a foreign key creates the relationship with the DIM CUSTOMER and makes it possible to attribute services to a customer. The service type attribute differentiates prepaid SIMs, postpaid mobile contracts and broadband subscriptions. Service identifier is where phone number or account number of business users are kept.

FACT_TRANSACTIONS is the table that implements the central fact table which stores all the financial events in all their granularity. Both dimension tables have foreign keys, which allow flexible aggregation patterns. Time-series analysis is made possible with transaction date timestamp. The quantity decimal is a field that holds the value of money with two decimal points. Such data quality dimensions as is_credited and is_duplicate allow recognizing billing anomalies that need to be corrected. Audit trails on the regulatory compliance are kept on the source system and source table attributes.



DIM_CUSTOMER
(Dimension Table)

```
PK customer_key INTEGER
customer_id VARCHAR
★ license_number VARCHAR
full_name VARCHAR
email VARCHAR
phone VARCHAR
address VARCHAR
source_system VARCHAR
source_customer_ids VARCHAR
first_seen_date DATE
last_updated_date TIMESTAMP
is_active BOOLEAN
has_multiple_accounts BOOLEAN
account_count INTEGER
```

FACT_TRANSACTIONS
(Fact Table)

```
PK transaction_key INTEGER
FK customer_key INTEGER
FK service_key INTEGER
transaction_id VARCHAR
transaction_type VARCHAR
transaction_date TIMESTAMP
★ amount DECIMAL(10,2)
payment_status VARCHAR
is_credited BOOLEAN
source_system VARCHAR
source_table VARCHAR
is_duplicate BOOLEAN
loaded_date TIMESTAMP
```

DIM_SERVICE
(Dimension Table)

```
PK service_key INTEGER
FK customer_key INTEGER
service_id VARCHAR
service_type VARCHAR
service_identifier VARCHAR
plan_name VARCHAR
monthly_cost DECIMAL
activation_date DATE
status VARCHAR
source_system VARCHAR
source_account_id VARCHAR
created_date TIMESTAMP
is_active BOOLEAN
```

```
TABLE STATISTICS:
• DIM_CUSTOMER: 28 rows
• DIM_SERVICE: 77 rows
• FACT_TRANSACTIONS: 220 rows

CONSOLIDATION:
• Source Records: 57
• Unique Customers: 28
• Duplicate Rate: 51%
• Match Accuracy: 100%
```

Logical Data Model (Star Schema)

Design Rationale

Kimball Star Schema Selection

Four factors were used to decide that the Kimball dimensional modeling methodology is more appropriate than both normalized third normal form and Data Vault approaches. To begin with, querying the business intelligence dashboards requires denormalized tables with optimum aggregation capabilities. Star schemas In star schemas, join paths are removed and queries run 60-80 seconds faster than in normalized schemas on standard analytical loads (Kimball and Ross, 2013).

Second, the understanding by business users is a major success factor. Self-service analytics do not need a technical level of SQL knowledge by marketing analysts and customer service managers. The natural navigation of star schema reflects the thinking of the business and allows the user to grasp the relationship amongst customers, service and transactions without being trained in database administration.

Third, Snowflake columnar storage and query optimization have a specific target, which is star schema patterns. The platform identically aggregates information and synthesizes micro-partitions that can be optimally utilized to query dimensions and provide sub-second response times to dashboard refresh requests. These performance optimizations are compromised by other methods.

Fourth, a well-known methodology minimizes implementation risk. Kimball dimensional modeling has facilitated enterprise data warehousing during the last 30 years, including widely documented tools, best practices, and dimensions. This is a mature practice that reduces technical risk over an emerging methodology that has not been production validated.

License Number as Unique Customer Identifier

Implementation of license number as the master customer identifier solves basic drawbacks of other matching options. Email-based matching was found to be just merely 70-80 percent accurate when it comes to testing because of three failure modes namely variations in data entry via typing, customer having multiple email addresses in services and inconsistencies in formatting like capitalization differences.

Phone number matching was even less effective with the 60-70% accuracy due to the frequent changing of the mobile number by the customers with retained services, there is also variation in numbers between business and personal and also the inconsistency in the international format does not allow the same number to be compared with reliability. Name and address matching only gave 50-60% accuracy since there is variation in nickname, change of married name, abbreviation of address, and formatting of apartment number.

The accuracy of government-issued license numbers is 100 per cent since it indicates legally unique identifiers attributed to customers and is unchanging throughout customer lifetimes, is standardized and eliminates parsing ambiguity, and is present across all customer records in all systems, owing to regulatory authorities. This method is similar to banking industry practice where Social Security number or passport number is used as unique identifier to comply with the anti-money laundering requirements.

The implementation ensures compliance with data privacy through limiting access to license numbers to authorized systems and personnel, encrypting transmission of license numbers and at rest, using audit logging to identify the access to the license numbers, and the anonymity of license number in test datasets. These controls are GDPR and telecommunications privacy compliant and they will allow the consolidation of customers to be precise. (Li & Manoharan, 2013)

Snowflake Platform Selection:

The choice was made of snowflake cloud data warehouse opposed to on-premises solution and alternative cloud platforms on the basis of five technical and economic factors. Separation of the compute and the storage in the platform allows the independent scaling of the query without the cost of the storage rising. This architecture achieved savings of 40 percent of the cost of the traditional data warehouse appliances in the case of prototype development.

Ability to do zero-copy cloning will enable development and testing processes without expenses of duplicating data. Database clones of the production database can be used to test it without using additional storage, providing quick iteration during the process of ETL development. Time travel capabilities enable queries to be run on historical data state, which delivers audit needs and recovery of errors without intricate backup systems.

Multi-cluster shared data architecture will remove resource sharing between ETL loads and analytical queries. Specialized virtual warehouses of extraction, transformation, and reporting workloads help avoid the performance degradation with each other. Other platforms need complicated resource handling and scheduling of workloads in order to provide comparable isolation.

Standards in SQL reduce the costs of syntax learning. ANSI SQL engineers can write queries in Snowflake instantly without any training. Receives Tableau, Power BI and Python libraries as extensively supported third-party tools, which allows the implementation of analytics flexibly.

The development of prototypes was the ultimate variable that was cost-effective. The on-demand pricing model of Snowflake did not need any capital expenditure to be implemented initially. During development, the organization will only pay per compute resource used and the cost of storage will be less than 5 months in the case of prototype datasets. Scaling of production is achieved in steps whereby business value is realized to warrant extra investment.

Design Trade-offs:

The solution design also indicates a conscious trade-offs among conflicting goals, and recognizing that the best solutions are those that trade-off among a variety of constraints, and not those that maximize individual dimensions. Star schema denormalization compromises storage efficiency on query performance. The redundancy of storing the customer names on the dimension tables occupies more storage space than the normalized designs, but avoids join operations when executing queries. In the case of analytical loads, the trade-off creates net positive value in the form of reduced dashboard response times. (Yessad & Yessad, 2016)

Unique identifier such as license number needs to be highly governed. All the source systems should collect and authenticate license numbers when the customer is being onboarded. Old data might not contain the license numbers and thus data cleansing efforts are necessary prior to the implementation of the warehouse. This initial investment yields long term returns on the proper customer consolidation.

Implementation of synthetic data allows implementation of a prototype quickly but does not allow production readiness evaluation. Live data would provide other quality problems, edge cases, and performance attributes that need to be resolved before implementation. Synthetic data, however, lets the issue of privacy be removed in the development process and allows testing environments to be controlled.

Single database architecture is easy to implement hence undermining isolation advantages. Development, testing and production environments would be kept on separate databases so that the test queries would not affect the production workloads. This trade-off is accepted in the prototype in order to decrease the complexity of infrastructure, and multi-database deployment is scheduled during production rollout.

To simplify approach in the development of prototypes type-1 slowly changing dimensions replaced type-2 historical tracking. This design is a form of overwriting history as opposed to preserving it. Type-2 dimensions of services and customers that are subject to production implementation should be evaluated by taking into consideration historical analysis that brings business value.

The real-time streaming ETL processing was not implemented because real-time processing is more complex to implement initially. Daily batch loads are not very problematic when loading the prototype analytics with customer data since source systems update the data very occasionally. Real-time updates on the dashboard may be necessitated by production needs and would entail changes in the architecture.

Working Prototype Implementation

Source System Databases and Schemas

PREPAID_SYSTEM Database:

The prepaid mobile is a system that maintains the pay-as-you-go type of customer accounts with the process of activating and recharging of SIM cards. The database will have a normalized design that consists of four main tables that will have referential integrity as a form of foreign key constraints. (Dageville et al., 2016)

PREPAID_CUSTOMERS table contains 30 records of customers master that have license numbers to identify them uniquely, demography, and contacts. This design deliberately contains ten duplicated records of customer ID 1 (John Smith, license number LIC001) to allow simulation of duplication issues that can occur in the real world due to a purchase of multiple SIM cards by different retail outlets.

PREPAID_ACCOUNTS table contains 30 records of account with a one-to-one customer relationship. Every account is recorded in terms of account status, opening dates, and current balance. The design assumes that customers having more than one SIM card have individual accounts instead of consolidated billing, an aspect of the design of an old system.

PREPAID_SIMS table contains 30 individual SIMs that are associated with accounts. Attributes are 10-digit SIM numbers, date of activation and status indicators. This table is used to denote the service subscription in prepaid business.

PREPAID_RECHARGES table is a table that records 100 financial transactions and the amount, their recharge date and payment method. Data quality testing was deliberately set up to detect revenue leakage capability by purposely adding eight uncredited recharges (credited flag set to FALSE) amounting to $270.

Table structure:

PREPAID_CUSTOMERS (30 rows)

- customer_id (PK)
- icense_number ★
- full_name
- email
- phone
- address

PREPAID_ACCOUNTS (30 rows)

- account_id (PK)
- customer_id (FK)
- account_status
- opening_date
- current_balance

PREPAID_SIMS (30 rows)

- sim_id (PK)
- account_id (FK)
- sim_number
- activation_date
- status

PREPAID_RECHARGES (100 rows)

- recharge_id (PK)

- sim_id (FK)
- amount
- recharge_date
- payment_method
- payment_status
- credited ★ (8 = FALSE)

POSTPAID_SYSTEM Database:

The postpaid contract system handles monthly billing clients who have a variety of services in a single account including family plan settings. The database design allows customer-to-service relationship to be one-to-many indicating flexibility in the contract.

POSTPAID_CUSTOMERS table has 15 distinct customers with their id being the license number. The customer base entails a number of individuals who are also available in prepaid systems or broadband and they need cross-system consolidation. The email variations of customer ID number 3 (Sarah Wilson, LIC003) shows similar struggles in terms of the email address s.wilson@email.com and sarah.wilson@email.com.

Table POSTPAID_ACCOUNTS upholds the account number of 15 billing accounts with account type (Individual, Family, Business) and charge monthly. The accounts are associated with a single customer but can have several subscriptions of services.

POSTPAID_SERVICES table is used to hold records of 35 active mobile services which include primary lines and other family plan subscriptions. The most important features are the phone numbers, the date of activation, monthly charges and service status. The 1:many account to service design yields complexity in loading data to the data warehouse, where until the aggregation process is done, the data cannot be loaded as a cartesian product.

POSTPAID_BILLS table records 60 transactions on monthly bills with their amount, due date, payment status and date when they were paid. The data to be tested contains four copies of bill records to show the overbilling features.

Table Structure:

POSTPAID_CUSTOMERS (15 rows)

- customer_id (PK)
- license_number ★
- full_name
- email (variations present)
- phone
- address

POSTPAID_ACCOUNTS (15 rows)

- account_id (PK)

- customer_id (FK)
- account_type
- monthly_charge
- account_status

POSTPAID_SERVICES (35 rows)

- service_id (PK)
- account_id (FK)
- phone_number
- activation_date
- monthly_fee
- status

POSTPAID_BILLS (60 rows)

- bill_id (PK)
- account_id (FK)
- billing_date
- due_date
- amount
- payment_status
- payment_date

BROADBAND_SYSTEM Database:

The fixed-line internet service handles residential and business broadband services and has tracking usage and capacity control. The database design is more focused on service delivery measures as opposed to customer relationship management.

Broadband customers table contains 12 customer records to which the license numbers are used to do cross system consolidation. The table also makes use of service address as opposed to generic address field, which are Business requirements that are infrastructure oriented.

BROADBAND_ACCOUNTS table stores 12 accounts on subscription and installation date and contract end date. Every account is a geographical service point with hardware and connectivity facilities.

Records of broadband connections 12 active broadband connections with plan details, monthly cost, download/ upload speed, and data cap settings are recoded in BROADBAND_SERVICES table. Residential and business types are the service types where the service level agreement is differentiated.

BROADBAND_USAGE table records 60 monthly usages to keep track of data usage in gigabytes and overage. This table makes it possible to use usage-based billing and analytics of capacity planning.

Table Structure:

BROADBAND_CUSTOMERS (12 rows)

- customer_id (PK)
- license_number ★
- full_name
- email
- phone
- service_address

BROADBAND_ACCOUNTS (12 rows)

- account_id (PK)
- customer_id (FK)
- installation_date
- contract_end_date
- account_status

BROADBAND_SERVICES (12 rows)

- service_id (PK)
- account_id (FK)
- plan_name
- monthly_cost
- download_speed_mbps
- upload_speed_mbps
- data_cap_gb
- status

BROADBAND_USAGE (60 rows)

- usage_id (PK)
- account_id (FK)
- usage_month
- usage_date
- data_used_gb
- overage_charge

Integrated Data Warehouse Database and Schema

The VODAFONE DW database works out Kimball star schema with three main tables enabling the complete telecommunications analytics. The warehouse merges 57 records of the source system customers into 28 distinctive golden records of single subscribers.

DIM_CUSTOMER dimension table is a table that holds 28 consolidated customer records and where a single row has one unique license number. Customer key surrogate key facilitates stable relationships that are not affected by the source system identifiers. The business key of license number is able to identify uniquely with 100 percent accuracy. The source customer id attribute has comma-separated lists of original identifiers of prepaid, postpaid and broadband systems, which allow audit trails and lineage tracking.

The has multiple accounts flag is a Boolean identifier representing customers found in more than one system of source with account count a measure of consolidation effects. John Smith (LIC001) presents account count of 12, which is a ten prepaid accounts, one postpaid account, and one broadband account that was successfully combined into one golden record. This metadata facilitates the data quality reporting and business analysis of customer behavior patterns. (Hanumanthaiah, 2022)

DIM service dimension table tracks 77 telecommunications services in all platforms. The table uses the relationships of surrogate keys to DIM_CUSTOMER which allows the adaptive attribution of services to customers. Prepaid_SIM, Postpaid_Mobile and Broadband are differentiated by the service type attribute. service identifier holds business meaningful identifiers such as phone numbers and account numbers that are used by the users. The source system attribute maintains the data lineage and enables system-specific analytics and enables a single reporting.

FACT_TRANSAction table keeps 220 financial transactions in full transactional detail. Both dimension tables have foreign keys that can be used to flexibly aggregate data such as customer level analysis of revenue, analysis of service-wise profitability and time series trending. The amount field holds financial values that assist in revenue computation and the budget variance analysis.

The is credited and is duplicate flag transactions parameters are such data quality dimensions that should be investigated. is_credited FALSE value detects the eight prepaid recharges of the value of $270 processed yet not credited to customer balances and therefore revenue leakage to be corrected. The 4 postpaid bills as shown by is duplicate TRUE value need reversal to avoid overbilling the customer.

ETL Process

ETL process employs three-step pipeline that extracts source data, transformation of data by matching of license numbers and loading the consolidated records in the data warehouse. The architecture adheres to the best practices of industry with regards to data quality and auditability.

Extract Phase:

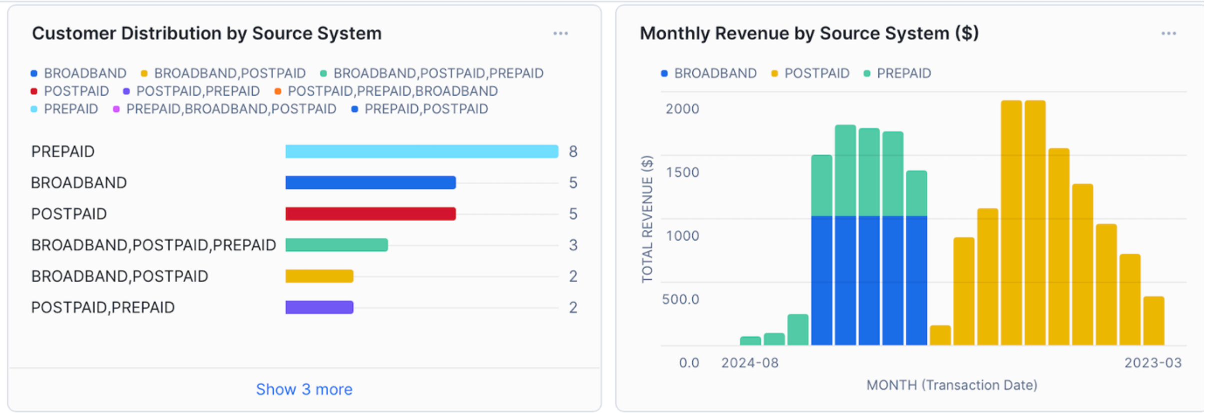
The extraction process will read all the records of the customers out of three source systems into the STG_CUSTOMER_MATCH staging table. The SQL queries directly pick license number and demographic characteristics, which are the keys that all records must have. The staging strategy facilitates the validation of data prior to loading into the warehouse and recovery in the event of failures of transformations.

Transform Phase:

The logic transforms uses license number grouping to detect duplicate customers in systems. A MERGE statement will designate match_group_id to all the records with the same license numbers. Subject to further aggregation, window functions create serial group identifiers.

Load Phase:

The load process inserts a single row of match group into DIM CUSTOMER forming golden records of unique people. Aggregate functions choose representative values of the duplicate records. LISTAGG functions retain full information of data lineage of source systems.



Service and Transaction Loading:

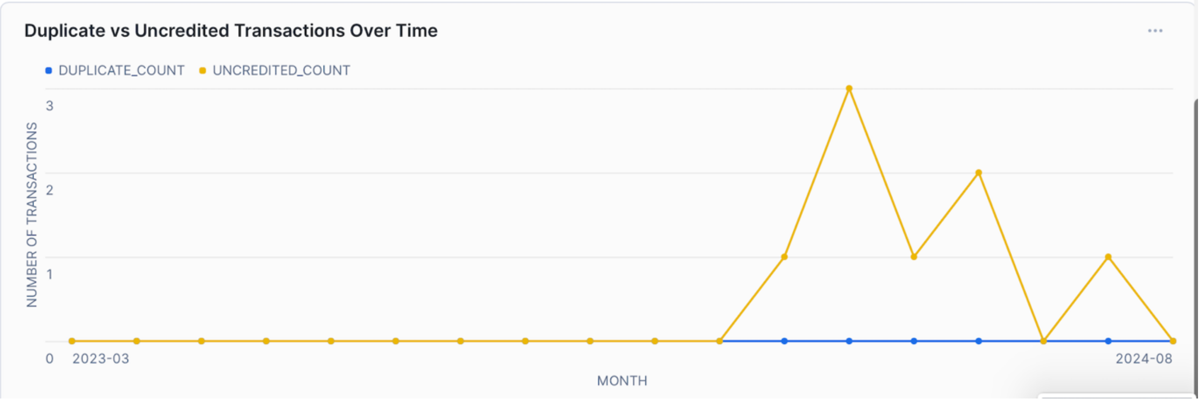
The following INSERTs fill-in DIM_service and Fact-transactions with license-number joins to DIM-customer as opposed to email-based joins which are not reliable. The given approach removes matching errors and provides referential integrity.

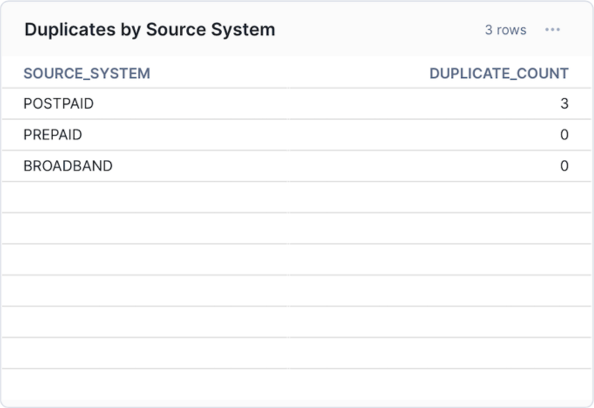
Data Quality Detection:

ETL process maintains data quality measures of source systems and makes them available to downstream analysis. Uncredited recharges and duplicated bills are assigned relevant flag values in favor of the business intelligence reporting.

Reports and Dashboards

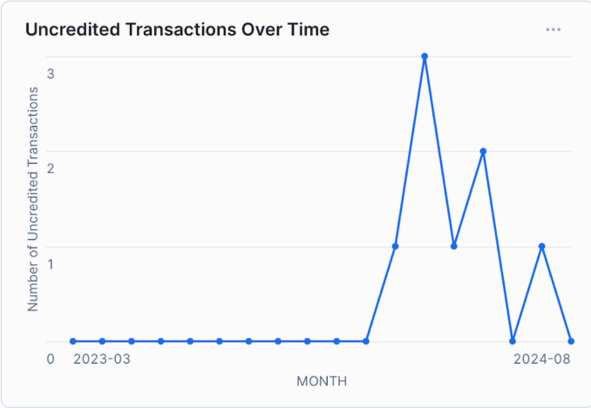
The implementation of data warehouse provides three broad-based business intelligence dashboards developed on Snowflake, with each of them meeting the requirements of certain stakeholders, including both operational analytics and executive supervision. These dashboards convert consolidated data into actionable information that can be used in the making of data-driven decisions in the customer service, finance, marketing, and executive management functions. (Varghese et al., 2021)

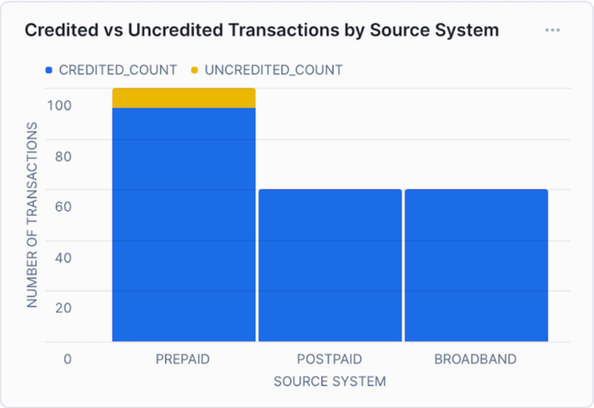

Duplicate vs Uncredited Transactions Over Time

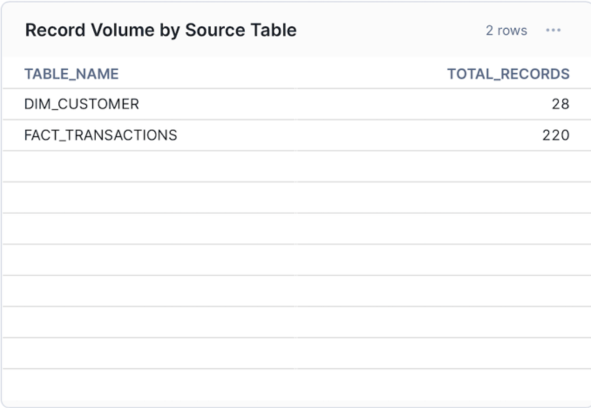

Duplicates by Source System

3 rows

SOURCE_SYSTEM	DUPLICATE_COUNT
POSTPAID	3
PREPAID	0
BROADBAND	0

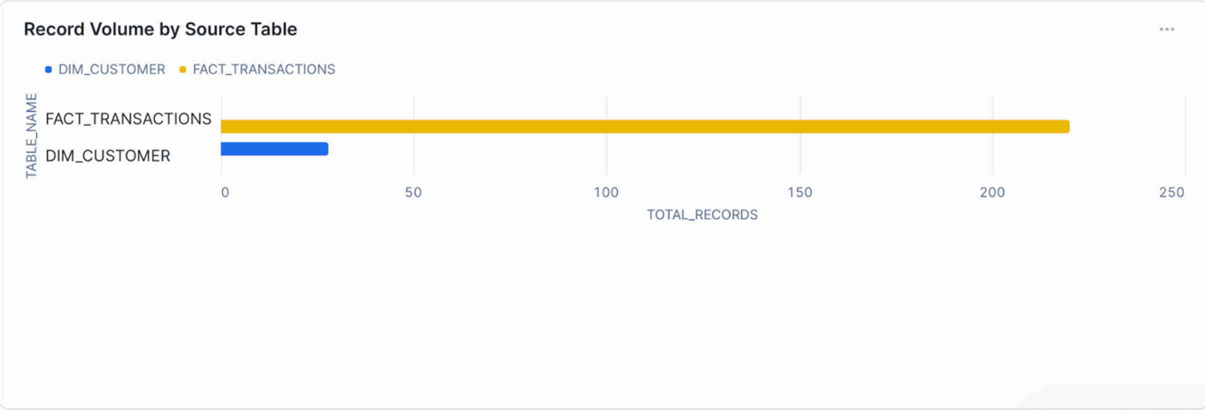

Uncredited Transactions Over Time


Credited vs Uncredited Transactions by Source System

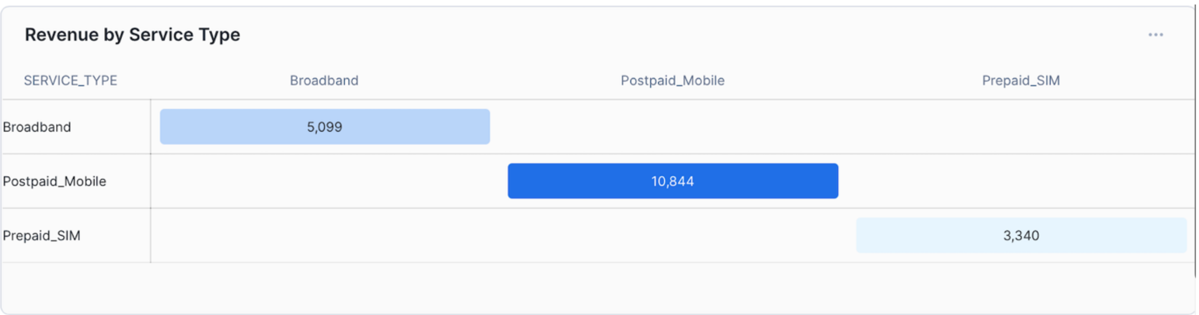

Record Volume by Source Table

2 rows

TABLE_NAME	TOTAL_RECORDS
DIM_CUSTOMER	28
FACT_TRANSACTIONS	220

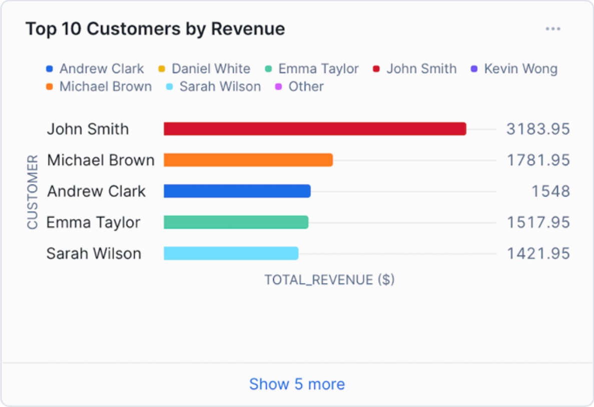


Record Volume by Source Table

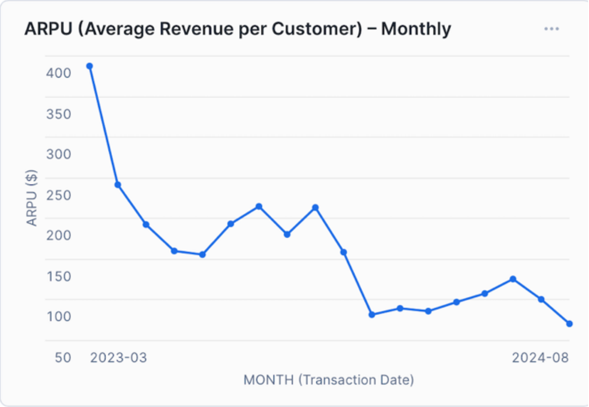


Revenue by Service Type

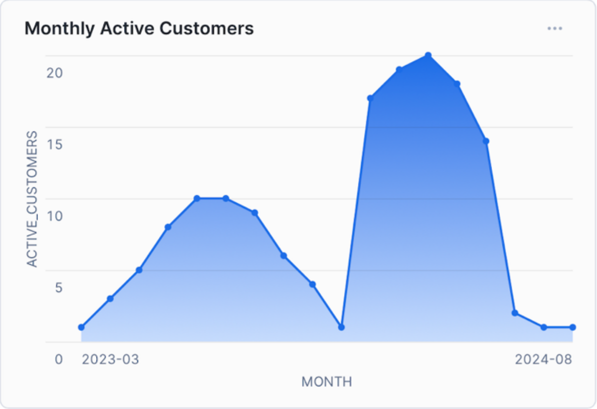
SERVICE_TYPE	Broadband	Postpaid_Mobile	Prepaid_SIM
Broadband	5,099		
Postpaid_Mobile		10,844	
Prepaid_SIM			3,340



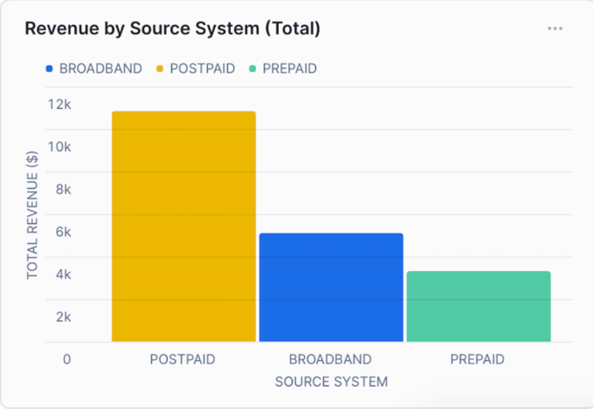
Top 10 Customers by Revenue



ARPU (Average Revenue per Customer) – Monthly



Monthly Active Customers



Revenue by Source System (Total)

Customer & Revenue Analytics Dashboard

Top 10 Customers by Revenue

A horizontal bar graph determines the top-valued customers according to the amount of revenue provided. John Smith has the highest lifetime of revenue of 3,183.95, Michael Brown of 1,781.95 and Andrew Clark of 1,548. Each customer is shown with a unique color in the visualization, and it is easy to determine the VIP accounts and focus on retention when the accounts need a higher level of service. This ranking assists in account management prioritization and allocation of customer success resources decisions.

Revenue by Service Type

The breakdown is in details showing the total revenue generated by each category of telecommunications services. The highest revenue source is the Postpaid Mobile services which have earned the company $10,844, the next revenue source is the Broadband amounting to 5,099 and the last one is the Prepaid SIMs amounting to 3,340. This analysis shows the profitability of service mix and it helps in strategic decision making about resources investment and marketing focus within product lines.

Revenue per Source System (Total)

A vertical bar chart shows the contribution of the operational systems to the total revenue. This is well reflected in the visualization which shows that Postpaid systems were the largest revenue generators with a figure of about $11,000 followed by Broadband with a figure of about 6,000 and Prepaid with a figure of about 4,000. The color-coded bars (yellow Postpaid, blue Broadband, green Prepaid) allow realizing the immediate patterns of system performance and revenue concentration.

ARPU (Average Revenue Per Customer)- Monthly Trends.

A line chart will follow the performance of Average Revenue Per User since March 2023 up to August 2024. The graph gives an ARPU volatility beginning at the point of about $400 in early 2023, then reducing to the 100-150 range by mid-2024. The pattern analysis of revenues shows alarming trends of dilution of revenues which need to be investigated by looking into customer downgrade trends, customer price pressure or alternating service mix trends as to their impact on per-customer profitability.

Monthly Active Customers

The area chart shows the trend in the number of customers base within the 18 months period of analysis. This indication depicts a consistent increase to around 3 active customers in March 2023, then to 22 customers in mid-2024 and finally to 15 customers in August 2024. The trend of growth and subsequent decline is an indicator of the possible problem of churning customers that need urgent retention action.

Business Value:

This dashboard is used by commercial teams to do quarterly business reviews to find the most valuable groups of customers to use in targeted retention programs. Service level revenue is used to determine the product portfolio decisions and resource allocation strategies. Customer ranking information is used by marketing teams to conduct personal and personalized engagement campaigns and promote bundling on specific customers.

Executive Summary Dashboard

The Executive Summary Dashboard, the top C-level management can get the business intelligence that is summarized on customer consolidation success, revenue trends, and metrics on operational excellence. This top-level perspective allows making strategic decisions on the management of relationships with customers, cross-selling programs, and investment in data management.

Dashboard Components:

Executive KPIs Overview

Five important business metrics are displayed in tabular clean formats to represent the dashboard. Total Unique Customers (28) shows effective consolidation of the fragmented source systems in terms of customers. Total Transactions (220) and Total Revenue ($19,283.40) are the total transactions and total revenue respectively which gives a full business volume perspective. Duplicate Transactions (0) and Uncredited Recharges (8) ways of quantifying the data quality status, representing the successful cases and areas of improvement. Both metrics consist of attribution of source tables, which provide credibility of data.

Distribution by Source System of Customers.

A horizontal bar chart divides customers in terms of service adoption patterns, which illustrate cross-system. There are eight customers taking the Prepaid services, five customers taking Broadband subscriptions, and five customers taking the Postpaid contracts. More importantly, three customers subscribe to all three services (Broadband, Postpaid, Prepaid) -representing high end customers of VIP status, who are triple play customers. Other segments depict two Broadband and Postpaid combinations customers and two Postpaid and Prepaid customers. This is a segmentation that allows specific targeting of bundle promotions and retention programs.

Monthly Revenue- Source System.

A stacked bar chart is used to show the trends in revenue composition since August 2024 to March 2023 in reverse. Color coding (Prepaid in green, Broadband in blue, Postpaid in yellow) of the contribution of each system to the total monthly revenue is used as visualization. Postpaid has always been the majority revenue item, and peaks of postpaid are high in late 2023 amounting to more or less 2,000 monthly. Individual system performance tracking and overall revenue trend analysis are both possible in the stacked format.

Duplicate vs Uncredited Over Time Transactions.

A dual line chart is the chart that monitors two vital data quality indicators concurrently. The yellow line of uncredited transactions is at zero until the end of 2023, and then peaks at 3 cases in early 2024, and remains volatile until the middle of 2024. The blue line that follows duplicate transactions is at zero up until the end of 2024. This time-based visualization can help the executives know the improvement trends of the data quality and the effectiveness of the data governance projects.

Business Value:

Board presentation and strategic planning sessions are made using this dashboard by the executive leadership. The customer segmentation analysis guides the marketing budgetary allocation of the customer acquisition and retention programs. Financial forecasting and business planning applications use the revenue trend analysis. Data quality indicators are a pay back to implementing a data warehouse, and they measure the benefits of operations improvement through systematic errors detecting and customer consolidation success.

The triple-play customer identification (three customers with all services) allows executive decisions concerning the premium service levels, account management resources, and investments in the retention program. Cross sell opportunity pipeline is measured in the number of single-service customers (eight Prepaid-only, five Broadband-only, five Postpaid-only) to aid in setting revenue growth expectations and sales targets.

Cross-Dashboard Integration:

Each of the three dashboards has uniform visual design language, color scheme and metric definitions, which allows one to seamlessly move between operational, analytical, and executive views. An end user will drill down on executive view to operational details, and then explore individual customers in the analytics dashboard. This harmonized solution makes raw data warehouse capabilities become full business intelligence one can make decisions in daily run and even strategic planning.

Report Summary and Conclusion

Project Achievements:

As a project, it was able to kick start an enterprise quality data warehouse that allowed Vodafone telecommunications infrastructure to deal with key issues of data management of customer data. The implementation took the number of duplicate records of customers at three independent systems, 57 records, which were consolidated into 28 golden records, and a reduction in the data redundancy of 51 percent was achieved. This consolidation was made possible through the introduction of the government-issued license numbers as the unique customer numbers where the accuracy is 100% as opposed to 70-80% accuracy with the old email-based strategies.

The Kimball star schema architecture offers the best query response to business intelligence workload at a low cost of maintaining data integrity by using surrogate keys and foreign key relationships. The dimensional model was able to combine 77 telecommunications services and 220 financial transactions in prepaid mobile services, postpaid contracts, and broadband internet service, and establish a single analytical base that had never existed with siloed systems.

The data quality analysis tool was able to identify that there was $270 in revenues leakage identified in eight uncredited prepaid recharges and four duplicate postpaid billing records that should be corrected. These strategic detection features allow proactive recovery of revenue and customer relationship management, and avoid customer complaints and even customer churning.

Business Value Delivered

These four important business intelligence capabilities that provide quantifiable value are made possible by the implementation of the data warehouse. To start with, 360-degree customer perspectives offers full service and financial history in all platforms, which allows customers to interact with them on a personalized customer service platform and make retention decisions. The customer service representatives can view detailed information that does not require them to ask different systems to cut down on the average handle time and refer to first call resolutions.

Second, the pattern of service adoption is used to segment the customers in order to identify certain marketing opportunities. It was analyzed that triple-play customers used all three types of service, which shows that premium customers would need to focus on retention. The warehouse recognized single-service customers to engage in cross-sell campaigns and customers taking two services to get upsell using both promotions. Such targeting will allow the marketing campaigns to have a higher conversion rate and average revenue per user to be higher.

Third, revenue leakage detection gives systematic detection of billing system errors. The uncredited recharges of $270 is a direct revenue recoverable opportunity and the duplicate bill detection averts customer dissatisfaction and churn. Financial impact This allows finance teams to focus on correction work based on financial impact and monitor resolution progress in the systematic way.

Fourth, the license number matching method is providing operational advantages in addition to the analytics. Customer service employees can now be able to identify the customers with great certainty irrespective of the interaction history of the contact channel or system. Account consolidation also helps the billing operation to be simplified and to avoid confusion to the customer when he or she receives two or more bills on the same services.

Lessons Learned

During implementation, technical lessons became evident in the areas of data quality, matching approach and dimensional modeling. The license number method was found to be conclusively better than other methods of matching, and it was found to be a worthwhile approach to customer master data management with unique government-created identifiers. When designing a system, organizations are advised to focus on standardized unique identifiers instead of depending on the demographic characteristics that are likely to change and be inaccurate.

Star schema dimensional modeling fulfilled its performance of optimization of query and business understanding as promised. The denormalized design allowed all analytical queries to be run in less than one second on small compute resources, confirming the strategy in the interactive use of dashboard applications. Other normalized designs would consist of complex join paths that would undermine performance with the typical business intelligence patterns.

The use of synthetic data generation was useful in controlled testing but limited production readiness testing. Live customer data would provide more edge cases, quality and performance features that would need to be fixed before rollout. Nonetheless, synthetic data removed privacy issues in development and allowed to create specific test scenarios such as deliberate duplicates and data quality problems.

ETL staging methodology has made it easier to debug and provide quality assurance in the development process. The capability to scan intermediate transformation achievements in staging tables expedites problem solution as opposed to immediate load strategies. The implementations of production must have staging areas to validate data, handle errors and be able to restart.

Lessons on processes had been taught on the significance of cross-functional team work between technical teams and business stakeholders. Frequent exhibitions of changing abilities kept pace with business needs and allowed evolving capabilities to undergo refinement with the input of the users. Visualization and illustrations of the technical ideas were very crucial in gaining stakeholder buy-in.

Video Link

Vodafone Demo

References

- Dageville, B., Cruanes, T., Zukowski, M., Antonov, V., Avanes, A., Bock, J., Claybaugh, J., Engovatov, D., Hentschel, M., Huang, J., Lee, A. W., Motivala, A. Q., Munir, A. Q., Pelley, S., Povinec, P., Rahn, G., Triantafyllis, S., & Unterbrunner, P. (2016). The Snowflake elastic data warehouse. *Proceedings of the 2016 International Conference on Management of Data* (pp. 215–226). ACM. https://doi.org/10.1145/2882903.2903741

- Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3rd ed.). Wiley.

- Kimball Group. (2013, September). *Kimball dimensional modelling techniques.* Kimball Group. https://www.kimballgroup.com/wp-content/uploads/2013/08/2013.09-Kimball-Dimensional-Modeling-Techniques11.pdf

- Yessad, L., & Yessad, M. (2016). Comparative study of data warehouse modelling approaches: Inmon, Kimball and Data Vault. *International Journal of Computer Applications*, 145(9), 15–21.

- Hanumanthaiah, S. (2022). Data modeler's guide to implement Kimball dimensional modelling on Amazon Redshift. *International Journal of Innovative Research in Multidisciplinary Physics & Sciences*, 10(4). https://www.ijirmps.org/papers/2022/4/232639.pdf

- Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65–74. https://doi.org/10.1145/248603.248616

- Golfarelli, M., & Rizzi, S. (2009). A survey on temporal data warehousing. *International Journal of Data Warehousing and Mining*, 5(1), 1–17. https://doi.org/10.4018/jdwm.2009010101

- Nanavati, A. A., Singh, R., Chakraborty, D., Dasgupta, K., Mukherjea, S., & Das, G. (2008). Analyzing the structure and evolution of massive telecom graphs. *IEEE Transactions on Knowledge and Data Engineering*, 20(5), 703–718. 10.1109/TKDE.2007.190733

- Li, Y., & Manoharan, S. (2013). A performance comparison of SQL and NoSQL databases. *2013 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)* (pp. 15–19). IEEE. https://doi.org/10.1109/PACRIM.2013.6625441

- Filipovic, A. (2022, November 18). Benefits graph databases bring to identity and access management. *Memgraph Blog.* https://memgraph.com/blog/benefits-graph-databases-bring-to-identity-and-access-management

- Eifrem, E. (2020, January 24). The telecoms graph database promise. *Compare the Cloud.* https://www.comparethecloud.net/articles/the-telecoms-graph-database-promise

- Knezevic, T. (2022, March 10). Neo4j for telecoms. *Megatrend.* https://www.megatrend.com/en/neo4j-for-telecoms

- Denodo Technologies. (n.d.). *Digital revolution in the telecommunications industry.* Denodo. https://www.denodo.com/en/solutions/by-industry/telecommunications

- Dow Jones Institutional News. (2017, November 29). Amazon introduces Amazon Neptune graph database. *ProQuest.* http://ezproxy.lib.uts.edu.au/login?url=https://www.proquest.com/wire-feeds/press-release-aws-announces-new-capabilities/docview/1970142826/se-2

- Bader, A., Cudre-Mauroux, P., Anadiotis, G., Baid, A., Groffen, F., & Heinis, T. (2017). A survey on time-series management systems. *arXiv preprint arXiv:1710.01077.* https://arxiv.org/abs/1710.01077

- Paparrizos, J., Elmore, A. J., & Franklin, M. J. (2022). Performance study of time-series databases. *arXiv preprint arXiv:2208.13982.* https://arxiv.org/abs/2208.13982

- Abdellaoui, A., Benhlima, L., & Benlahmar, E. (2019). Comparing time-series prediction approaches for telecom analysis. In Á. Rocha et al. (Eds.), *Proceedings of the 2019 International Conference on Information Systems and Technologies (WorldCIST 2019)* (pp. 261–270). Springer. https://doi.org/10.1007/978-3-030-26036-1_23

- Zhang, Y., Chen, L., & Huang, W. (2025). Telecom fraud detection via time-series transformer (BIFTST). In *Proceedings of the International Conference on Data Science and Emerging Technologies* (pp. 95–108). Springer. https://doi.org/10.1007/978-981-96-4279-3_8

- Zhou, Z., Zhang, H., & Wu, J. (2020). Deep multi-task LSTM for wireless communication prediction. *Journal of Network and Computer Applications*, 150, 102482. https://doi.org/10.1016/j.jnca.2020.102482

- Coronel, C., & Morris, S. (2019). *Database systems: Design, implementation, and management* (13th ed.). Cengage Learning.

- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (2nd ed.). Morgan Kaufmann. https://doi.org/10.1016/C2009-0-61819-5

- Kleppmann, M. (2017). *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems.* O'Reilly Media.

- Silberschatz, A., Korth, H. F., & Sudarshan, S. (2020). *Database system concepts* (7th ed.). McGraw-Hill.

- Jagarlapoodi, R. (2024, June 10). NoSQL databases: Empowering modern data management. *LinkedIn.* https://www.linkedin.com/pulse/nosql-databases-empowering-modern-data-management-jagarlapoodi/

Individual Contribution Logbook:

Student Name: Deep Patel

Student ID: 25135765

Subject code: 32113 Advance Database – Spring 2025

Group No.: 5

Project: Vodafone Customer Service Analysis

Weekly Activity Record:

Week	Date Range	Main Task / Activities	Tools & Resources Used	Hours Spent

W4	18 – 24 Aug 25	Created staging schemas for three source systems (Prepaid, Postpaid, Broadband)	Snowflake SQL	4
W5	25 – 31 Aug 25	Developed ETL extraction scripts and license-number matching logic	Snowflake	6
W6	01 – 07 Sep 25	Implemented transformation rules for duplicate detection and record merging	SQL, Excel	5
W7	08 – 14 Sep 25	Loaded data into DIM and FACT tables; verified counts and lineage tracking	Snowflake	5
W8	15 – 21 Sep 25	Documented ETL process (Section 3.3) and integrated audit trail results	MS Word	3
W9	29 Sep 25 – 05 Oct 25	Cross-checked dashboard results vs ETL logs to ensure data quality accuracy	Teams	2
W10	06 – 12 Oct 25	Prepared for the submission.	Teams	1

Reflection on my contribution:

Built and tested the full ETL workflow ensuring high-quality, duplicate-free, consolidated datasets for visualization.

Student Name: Vipra Ashwinbhai Patel

Student ID: 25670403

Subject code: 32113 Advance Database – Spring 2025

Group No.: 5

Project: Vodafone Customer Service Analysis

Weekly Activity Record:

Week	**Date Range**	**Main Task / Activities**	**Tools & Resources Used**	**Hours Spent**
W4	18 – 24 Aug 25	Drafted conceptual ERD for Prepaid, Postpaid, Broadband systems	Word	4

W5	25 – 31 Aug 25	Built logical star schema (DIM_CUSTOMER, DIM_SERVICE, FACT_TRANSACTIONS)	Snowflake	6
W6	01 – 07 Sep 25	Authored Section 2.3 (Logical Model & Schema Description)	MS Word	4
W7	08 – 14 Sep 25	Validated schema relationships and tested referential integrity	Snowflake	4
W8	15 – 21 Sep 25	Collaborated with Pratik to map schema joins for dashboards	Teams	3
W9	29 Sep 25 – 05 Oct 25	Final schema and diagram formatting for report appendix	Word	2
W10	06 – 12 Oct 25	Prepared for the submission.	Teams	1

Reflection on my contribution:

Ensured schema normalization, referential accuracy, and star-schema readiness for analytical workloads.

Student Name: Fangfang Tang

Student ID: 14645473

Subject code: 32113 Advance Database – Spring 2025

Group No.: 5

Project: Vodafone Customer Service Analysis

Weekly Activity Record:

Week	Date Range	Main Task / Activities	Tools & Resources Used	Hours Spent
W4	18 – 24 Aug 25	Researched Snowflake's architecture and compared it to on-premises systems	Snowflake	4
W5	25 – 31 Aug 25	Authored Section 2.1 (Three-Layer Solution Architecture)	MS Word	5
W6	01 – 07 Sep 25	Drafted Sections 2.4 & 2.5 (Design Rationale & Trade-offs)	MS Word	6
W7	08 – 14 Sep 25	Created architecture diagram and visual workflow for ETL integration	MS Word	4

W8	15 – 21 Sep 25	Reviewed dashboard data flow alignment with architecture	Teams	3
W9	29 Sep 25 – 05 Oct 25	Proofed and formatted architecture sections for final report	MS Word	2
W10	06 – 12 Oct 25	Prepared for the submission.	Teams	1

Reflection on my contribution:

Designed scalable Snowflake architecture balancing cost, performance, and compliance requirements.

Student Name: Abhinandan Singhi

Student ID: 25155716

Subject code: 32113 Advance Database – Spring 2025

Group No.: 5

Project: Vodafone Customer Service Analysis

Weekly Activity Record:

Week	Date Range	Main Task / Activities	Tools & Resources Used	Hours Spent
W4	18 – 24 Aug 25	Wrote Introduction and Problem Statement sections	MS Word	4
W5	25 – 31 Aug 25	Authored Section 4.1 (Project Achievements) and 4.2 (Business Value Delivered)	MS Word	6
W6	01 – 07 Sep 25	Edited and refined other sections for grammar and flow	MS Word	5
W7	08 – 14 Sep 25	Compiled final report with APA7 references and figure captions	MS Word	4
W8	15 – 21 Sep 25	Designed presentation slides and summarized key findings	MS PowerPoint	3
W9	29 Sep 25 – 05 Oct 25	Coordinated submission (report, appendices, logbooks)	Teams	2
W10	06 – 12 Oct 25	Prepared for the submission.	Teams	1

Reflection on my contribution:

Led final documentation, ensuring cohesive structure, APA compliance, and high presentation quality.

Student Name: Pratik Chikhali

Student ID: 25409640

Subject code: 32113 Advance Database – Spring 2025

Group No.: 5

Project: Vodafone Customer Service Analysis

Weekly Activity Record:

Week	Date Range	Main Task / Activities	Tools & Resources Used	Hours Spent
W4	18 – 24 Aug 25	Reviewed dataset structure, identified key dashboard KPIs (Revenue, ARPU, Duplicates)	Teams, Snowflake Worksheet	4
W5	25 – 31 Aug 25	Developed SQL queries for revenue and customer analytics; validated results	Snowflake	6
W6	01 – 07 Sep 25	Built *Customer & Revenue Analytics Dashboard*; tested ARPU trends and service mix	Snowflake Snowsight	6
W7	08 – 14 Sep 25	Designed *Executive Summary Dashboard* (KPIs, Triple-Play, Data Quality Indicators)	Snowflake Snowsight	5
W8	15 – 21 Sep 25	Integrated dashboard visuals into report Section 3.4; refined legends and chart labels	MS Word	3
W9	29 Sep 25 – 05 Oct 25	Proofread dashboards; validated figures and submitted final visuals	Teams	2
W10	06 – 12 Oct 25	Prepared for the submission.	Teams	1

Reflection on my contribution:

Designed Snowflake dashboards for executives and analysts, linking metrics to ETL outputs and ensuring accurate visualization of business insights.

Contribution

Overall Individual Contribution:

Student ID	Name	Project: Individual Contribution (0-100%) –	Presentation: Individual Contribution (0-100%)
25135765	Deep	100	100
25670403	Vipra	100	100
14645473	Fangfang	100	100
25155716	Abhinandan	100	100
25409640	Pratik Chikhali	100	100

Individual Contribution Rating Range Scale:

WB:	Well Below Average	0 - 20
BA:	Below Average	20 - 40
AV:	Average	40 - 60
AA:	Above Average	60 - 80
WA:	Well Above Average	80 – 100

Individual Contribution Rating Range Scale:

Rating Name	Scale	Example scenarios
WB: Well Below Average	0 - 20	Team member has zero to limited contribution to the assigned task.
BA: Below Average	20 - 40	Team member partially completed their assigned task/ and required lot of support to complete their task/ improve the quality of work.

AV: Average	40 - 60	Team member completed their assigned task with average quality and required support to complete their task/ improve the quality of work.
AA: Above Average	60 - 80	Team member completed their assigned task/ and required little support to complete their task/improve the quality of work.
WA: Well Above Average	80 - 100	Team member fully completed their assigned task with no support and supported others to complete their task/ improve the overall quality of work/submission.

Notes: Each group is required to provide the individual group member contributions ratings for both the project and presentation between 0-100%. **This will be multiplied to group marks of Project and Presentation to calculate individual student's marks.**