MIS 6309 Business Data Warehousing

Instructor: Kevin R. Crook

Data Warehouse Design Project

Student Name: Hemant D Patil

Spring 2017

JPMorgan Chase

JPMorgan Chase & Co. is one of the world's oldest, largest and best-known financial institutions. It was founded in New York in 1799, they have succeeded and grown by listening to customers and meeting their needs.

As a global financial services firm with operations in more than 50 countries, JPMorgan Chase & Co. combines two of the world's premier financial brands: J.P.Morgan and Chase. The firm is a leader in investment banking; financial services for consumers, small business and commercial banking; financial transaction processing; asset management; and private equity. A component of the Dow Jones Industrial Average, JPMorgan Chase & Co. serves millions of consumers in the United States and many of the world's most prominent corporate, institutional and government clients.

JPMorgan Chase & Co. is built on the foundation of more than 1,000 predecessors institutions that have come together over the years to form today's company. Their many well-known heritage banks include J.P.Morgan & Co., The Chase Manhattan Bank, Bank One, Manufacturers Hanover Trust Co., Chemical Bank, The First National Bank of Chicago and National Bank of Detroit, each closely tied in its time to innovations in finance and the growth of the United States and global economies.

As America expanded and diversified in the 1800s, new banks were formed across the nation. JPMorgan Chase has historic links to many of these early institutions, including The Western Reserve Bank, one of the first banks in Ohio when it was organized in 1812; Second State Bank of Indiana, formed in 1834 when Indianapolis still was a frontier town with a population of about 1,500; and Springfield Marine and Fire Insurance Co., which began operation in Illinois in 1851.

Abraham Lincoln was one of its first customers, depositing $310. All three banks are predecessors of Bank One, which merged with JPMorgan Chase in 2004.

JPMorgan Chase provides a variety of features as a bank. They encourage customers to open an account with them by providing attractive benefits. Also, special loan deals help customers purchase or invest money for personal or commercial gain. Company's website as well as a mobile application running on Android and IOS, to provide customers with easy accessibility for managing their accounts and transactions with the bank, from anywhere they wish to.

JPMorgan Chase competes with other banks by providing seamless customer service also they believe that if the customer is treated well then, we will have less chance of abandoning them in future.

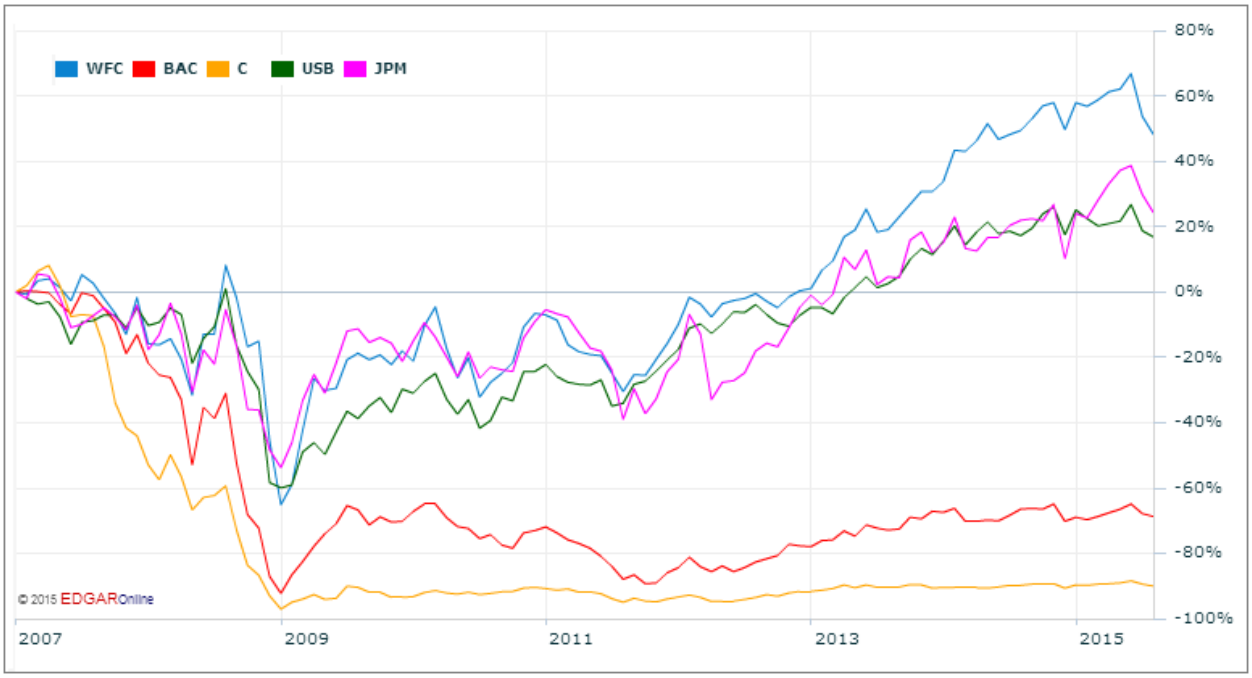
However, the competition is undeniable with all business and with different sectors! We can say that JPMorgan handled its competition squarely. But as there are new upcoming banks as well as relatively old banks, the bank faces new and sterner challenges than ever before.

JPMorgan Chase and its subsidiaries and affiliates operate in a highly competitive environment. Competitors include other banks, brokerage firms, investment banking companies, merchant banks, hedge funds, commodity trading companies, private equity firms, insurance companies, mutual fund companies, investment managers, credit card companies, mortgage banking companies, trust companies, securities processing companies, automobile financing companies, leasing companies, e-commerce and other Internet-based companies, and a variety of other financial services and advisory companies. JPMorgan Chase's businesses generally compete based on the quality and range of their products and services, transaction execution, innovation and price. Competition also varies based on the types of clients, customers, industries and

geographies served. With respect to some of its geographies and products, JPMorgan Chase competes globally; with respect to others, the Firm competes on a regional basis. The Firm's ability to compete also depends on its ability to attract and retain professional and other personnel, and on its reputation.

The financial services industry has experienced consolidation and convergence in recent years, as financial institutions involved in a broad range of financial products and services have merged and, in some cases, failed. This is expected to continue. Consolidation could result in competitors of JPMorgan Chase gaining greater capital and other resources, such as a broader range of products and services and geographic diversity. It is likely that competition will become even more intense as the Firm's businesses continue to compete with other financial institutions that may have a stronger local presence in certain geographies or that operate under different rules and regulatory regimes than the Firm.

Although JP Morgan Chase faces stiff competition from the wide of array banking institutions in the U.S. and globally, it continues to command market share and remains the biggest bank in terms of consolidated assets, according to Federal Reserve statistics. As per statistics released in August 2015, JPMorgan Chase had assets worth $2.10 trillion, compared to those of Bank of America Corporation (BAC) at $1.6 trillion, Wells Fargo and Co. with $1.57 trillion (WFC) and Citigroup Inc. (C), with $1.34 trillion. Other banks on the top ten list were US Bancorp (USB), PNC Financial Services Group Inc. (PNC), Bank of New York Mellon Corp. (BK), State Street Corp. (STT), Capital One Financial Corp. (COF) and the Toronto-Dominion Bank (TD).



The financial crisis of 2008-09 was a test for U.S. financial institutions, which JPMorgan Chase could endure better than peers like Citigroup, Wells Fargo, Bank of America and US Bancorp. The graph above shows the movement of these stocks together for the period starting January 2007 through August 31, 2015. The maximum dip was seen in the stocks of Citigroup and Bank of America while Wells Fargo, US Bancorp and JPMorgan Chase were able to limit the fall to some extent. The latter three have also had a stronger and faster recovery period compared to Bank of America and Citigroup.

The stock of JPMorgan Chase was up by 21.69% during 2006, while it fell by 9.63% and 27.77% in 2007 and 2008 respectively. The year 2009 witnessed the stock rebound with 32.16% returns, while returns were flattened at 1.8% in 2010. The year 2011 wasn't a happy one for JP Morgan Chase, as its stock fell by 21.62%. However, it sprang up again with 32.24% and 33% returns in

2012 and 2013, respectively, and it closed 2014 at $62.58, posting a 7.01% return for the year. In 2015, the stock is down 1.77% YTD (as of September 27, 2015).

Keeping a strong hold on your present customers becomes an increasingly difficult task now days and also coming up with new ideas while maintaining uniformity across the bank is again something which is tedious. Here, having a second opinion on challenges is also worth noting. Having to confront challenges of different varieties actually improves your organization as a whole. It uncovers previously undiscovered loopholes and makes you self-aware of improvement areas and you end up at a better position than where you started.

JPMorgan Chase needs to incorporate strategies to halt systems from getting haywire while responding to suddenly increased workloads. They want to have different systems regulating load and increasing throughput rates.

One of the solution for this can be Data mining. But if we want a long-term solution for this then implementation of data warehouse make sense. This will result in creating important reports, dissolve complexity, mine data and display comparisons. Another advantage of business data warehouse is that it is cheap and affordable option of JPMorgan Chase.

Data Warehouse Architecture:

Looking at the problems we can say Kimball method will be the best architecture for the design of data warehouse. Since the data is available at enterprise level which is dimensional and it will be easy to implement using Kimball architecture. We want to access the data directly which can be done easily by Kimball than Inmon's methodology.

We categorize data into either facts or dimensions, where facts contain all numeric data and dimension refers to information that gives context to the facts. Here data marts need not to be logical at subject area level and hence this method is being implemented for this data warehouse. Kimball's approach is a bottom-up approach which will be added advantage while being implemented for this warehouse. Here all the data will be organized dimensionally. This makes it easy to understand. Due to this retrieval of data is quick.

As it is a bottom-up approach data marts are created first which provide reporting and analytical capabilities for specific business areas such as accounts or transactions. These data marts together form the data warehouse using a bus architecture, which consist of conformed dimensions between all the data marts. Data warehouse in this case is nothing but a collection of logically self-contained and consistent data marts.

Advantage: Dimensional approach makes data warehouse easy to understand.

Disadvantage: Integrity of facts and dimensions, loading data warehouse with the data from different operational system is complicated.

We have chosen Kimball's out of several reasons mentioned below:

1) First, the data marts are created for different business process like customer, account, household etc.

This provides a thin view and understanding of the organizational data.

Then we combine all these data marts into one single atomic data warehouse.

2) In Inmon's method, users have access only to data marts and the data is brought into data marts from the enterprise wide data warehouse.

   In Kimball's method a single atomic repository of all the logical data marts is generated which the user can directly access to run reports and analytics.

3) Due to less time and less cost of implementation we prefer Kimball's method.

Business Problems:

Structure of banks:

It is more interesting, therefore, to establish why banking systems in the several countries do differ from one another, sometimes in quite material respects. We can consider the large bank as the data warehouse and the branches as the data mart to work on. Because at the end after combining large data marts we get data warehouse. The problems that banks faces are much the same the world over but there is considerable variety in the solutions that are put forward to resolve them. Hence, it is in the details of organization and technique that one tends to find the differences. Yet there is a tendency for the differences to become less pronounced because of growing efficiency in international communication and the disposition to emulate practices that have proved successful elsewhere. These similarities and differences can be discussed in terms of: (a) the structure of commercial banking systems; and (b) the varying emphases in the types of business that are done by banks in different countries.

Although one must be careful not to oversimplify, it is possible to classify banking structures as falling within one or another of certain broad categories. For example, data mart. In a number of other countries, it is more usual to find a small number of commercial banks, each of which operates a highly-developed network of bank branches. As well as one should understand why we need the data warehouse?

To answer the question, we should understand the structure of the bank which is in 3NF. The main problem while doing the analysis with respect to current as well as historical data was 1)

Normalized databases needs clear and broad understanding of business. 2) It takes more time to analyze and understand the business.

Similarly, if bank has large number of tables in their database, it consumes a large amount of space. Given the large amount of space it also takes a lot of time to develop and implement those tables. Querying on those tables is not easy because not a single table gives us the whole information. Banks must use complex queries such as joins to get what they want in such scenarios.

A customer of a bank always wants to view the current transaction as well as the previous transactions done. Hence, bank must maintain current as well as historic data. Operational systems don't store all the historic data which has a good analytical value.

For large data size the performance of operational systems is going to grow linearly or worse as it requires full table scan. For this reason also data ware house should be used by banks.

Customer rate of a bank:

JPMorgan Chase have grown to more than 325,000 CPC clients, up 51% from 2013. Customers have a lot of options while selecting a bank and hence it is necessary for Chase to be at the top to attract more and more customers.

Customer satisfaction is the main thing which attracts more customers. So, what should a bank do to keep their customers satisfied.

Data warehouse plays an important role in this manner. As the bank generates lots of data from its operations every day, it must be stored along with historical data. From the historical data, we can find out what our customers are looking for or what services they are using most of the time. Bank can improve those services so that it will be easy for customers to do the transaction more efficiently and softly. From the historical data, understanding customers' requirement is most important which bank should take care of. Bank can also upgrade the services which the customers are using the most for efficient use.

Services:

In today's world, nobody choose to walk in to the bank. This is because the services that bank is providing to their customers so that customer can do the banking by sitting at home. Customer really gets upset if some service didn't work due to some technical issue at bank side. In this case bank, may lose the potential customer. That's why services play an important role in keeping customers happy.

The more the number the services the more the space allocated for that customer because we want to keep current as well as the historical information when a customer uses a service. Services include mobile banking, net banking, sms/email alerts etc.

Bank uses marketing strategies to introduce the newly launched products among its customers or new potential customers. Bank uses their demographic data to find out the target customers. Census data or purchase data also plays an important role in such scenarios.

Some banks use one to one interactions with their customers or new potential customers to know their prospects or meet their demands. It is observed that interactive marketing has higher

Demography of branches:

It is very important for a bank to open branch at all possible locations. Usually bank opens their branches where a lot of residential houses and sometimes in downtown of cities where most of the offices are located for the convenience of their customers.

Bank also consider to open their branches in a neighborhood whose income is medium because people with medium income uses banks more that people with higher income or low income. Hence the demographic data plays very important role.

Credit:

Credit plays an important role in banking industry. It is observed that 90% customers use credit card for purchase or banking purpose. Usage of credit card creates credit history which brings lot of historical data (right from beginning of opening of an account of customer). For this the database size must be large enough for all customers of bank.

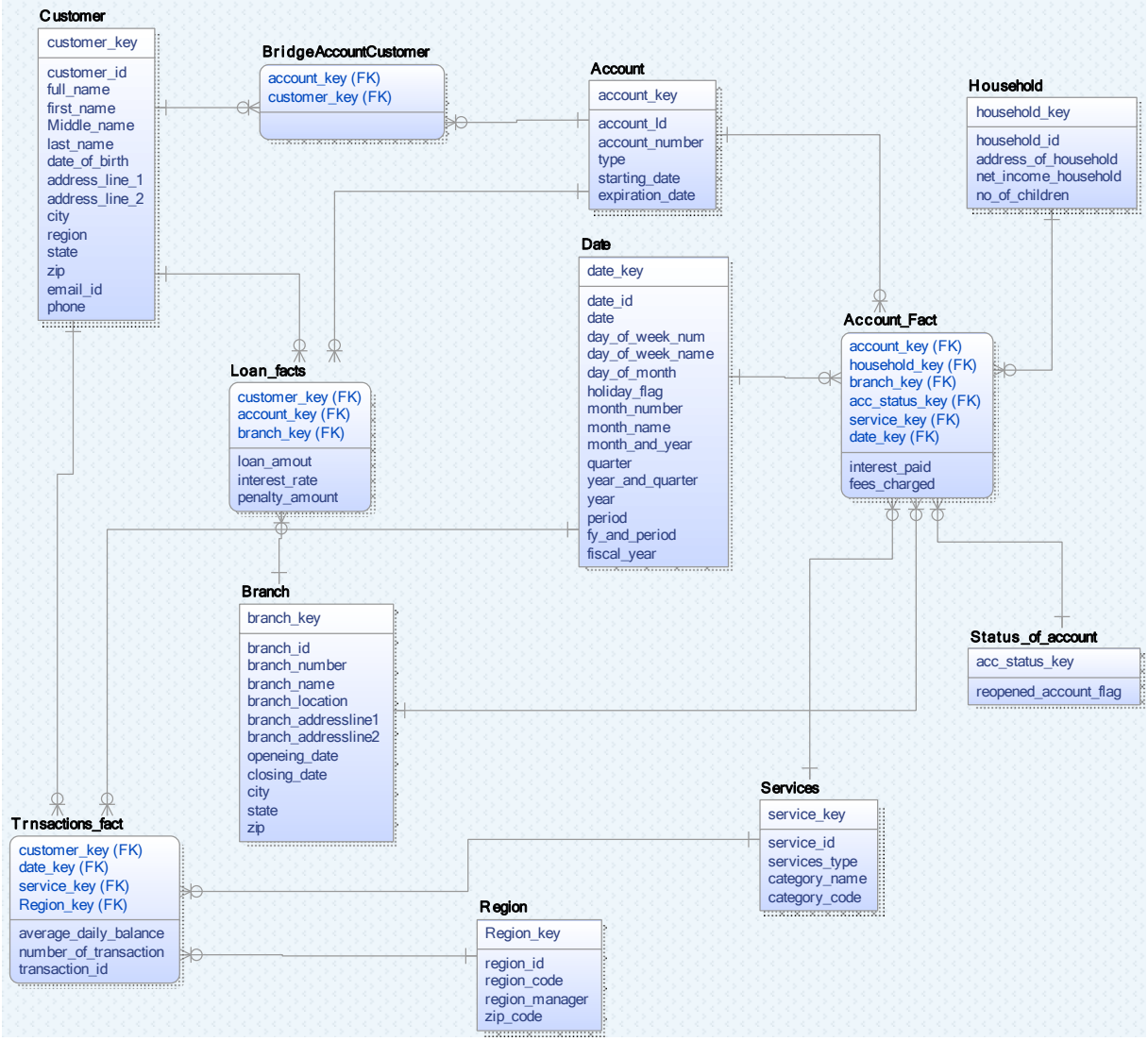
A customer can have multiple credit cards of different banks. Every customer has given a grace period for payment of credit card. If a customer fails to pay that then penalty is applied on that customers account. Bank should have an architecture which will effectively work on this but in the same case should not lose its customer by considering the past payment history of that customer. Loyalty of the customer plays important role in this case. If the customer is loyal then he will make

the payment on time. But sometimes due to emergencies if the customer fails to make payment on time and the bank charged that customer for that then there is a slight possibility of losing that customer. So in this case we need historical as well as current data. Timestamped dimension will do the best job in this scenario.

Narrow specific business problems:

a. Which services are most popular among the customers?
b. How much interest should be charged for loan amount for a customer?
c. Which branch of the bank charges the highest rate of interest for given account type?
d. What should be the interest rate applied to customer for account type?
e. What date had the maximum number of transactions under which region?

Star Schema:



Fact tables:

1. **Account_fact**:

    - Account_facts stores information of various accounts of customer.

- Grain here tells us about account details for customer as well as services activated for that customer. We can also find the interest rate applied for any customer and fees applied.
- Here the interest_paid and fees_charged are the two facts, which both contains the numeric values. We can say that fees_charged are additive facts, intrest_paid is non-additive fact.
- Business Problem:
  - Problem: what should be the interest rate applied to customer for account type?
  - Solution: Grain here will tell us about the household information related to the account of customer. From here bank can find out what interest rate should be applied for given account of the customer. Through various queries we can find the interest rate applied to different types of account as well as fees charged for those accounts.

2. **Transaction_facts**:
   - It stores information about the transactions done by customer.
   - The grain here is the transactions done by customers of different regions on different dates.
   - Here average_daily_balance and number_of_transaction are two facts. As both can be aggregated we can say that they are additive facts.
   - Transaction Id used in this table acts as degenerate dimension.
   - Business Problem:

- Problem: Which services are most popular among the customers?
- Solution:
- Problem: What date had the maximum number of transactions under which region?
- Solution: We can find the maximum number of transactions happened from the fact number_of_transaction, then the bank can find the date associated with that as well as the region related.

3. **Loan_facts:**

- It stores the information related to loan taken by customer.
- The grain here is interest rate applied to customers as per their account type as well as bank branch location.
- Loan_amount, Intrest_rate and penalty_amount are three facts for this table.
- Business Problem:
  - Problem: Which branch of the bank charges the highest rate of interest for given account type?
  - Solution: Bank can find the highest interest rate from the fact and the branch associated with it. Then, bank can find the account type from the account dimension.
  - Problem: How much interest should be charged for loan amount for a customer?

- ♦ Solution: Bank can find the account type of a customer as well as when he opened the account. If the customer is old bank may charge him less interest rate. It also depends on the branch location.

Dimensions Used:

1. **Customer:**
    - This dimension holds information about all customers of bank.
    - customer_key is a surrogate key for this dimension which as a primary key for this dimension. It has attributes such as cutomer_id, full_name, first_name, middle_name, date_of_birth, address_line_1, address_line_2, city, region, state, zip, email_id, phone.
    - Here phone is type 1 change since it is not necessary to maintain history for analytical purpose. State can be considered as type 2 change because it is necessary to preserve history that's why new rows are inserted into the dimensions.
    - As we can get lot of analytical value from customer dimension we can say It is highly browsable dimension. Using customer dimension, we can get how many transactions a customer performs on a daily basis or weakly basis, we can also find the account number associated with maximum number of transactions and highest average daily balance.
    - As this dimension contains all similar logical attributes, this is an affinity dimension. As it is a subset of account dimension, it acts as a conformed rollup dimension.

- It has a rich attributes address_line_1, address_line_2. These attributes give detailed location of the customer rather than overview of where the customer lives.
- A customer can have multiple accounts so connected vie a bridge(BridgeAccountCustomer).

2. **Account**:
    - It stores the account related information of a customer.
    - Account_key is a surrogate key here which acts as primary key for account dimension. It has attributes as account_id, account_number, type, starting_date and expiration_date.
    - It has no rich attribues.
    - It is conformed shared dimension between account_fact and loan_facts.

3. **Household**:
    - It stores information of household of a customer. Household information can give many insights about the customer which a bank can use to improve their knowledge about the customer.
    - Household_key is a surrogate key here which acts as a primary key for household dimension. It has attributes as household_id, address_of_household, net_income_household, no_of_children.
    - Here we capture the demographics of the household. This is highly browsable dimension.

- This is also affinity dimension as it contains all similar logical attributes.

4. **Date**:

    - It stores information related to date factor of transactions as well as account of customer.
    - Here date_key is a surrogate key which acts as a primary key for date dimension. It has attributes as date_id, date, day_of_week_num, day_of_week_name, day_of_month, holiday_flag, month_number, month_name, month_and_year, quarter, year_and_quarter, year, period, fy_and_period, fiscal_year.
    - It is a rich dimension as it contains combinations of attributes(year_and_quarter).
    - This dimension is a shared conformed dimension between account_fact and transaction_fact.

5. **Branch**:

    - It stores the information about various branches of the bank.
    - The surrogate key used here is branch_key. It has attributes as branch_id, branch_number, branch_name, branch_location, branch_addressline1, branch_addressline2, opening_date, closing_date, city, state, zip.
    - Information we can get from this dimension is what's the branch id for this particular location?, what is the branch number associated with this customer?
    - This dimension is a shared conformed dimension between account_fact and loan_fact.

- opening_date and closing_date stores the information on which the branch of the bank was opened and closed. From this we can find out the duration for which the branch was in service.
- This type of change is known as Timestamped Dimension.

6. **Status_of_account**:
    - acc_status_key is the surrogate key used here. It has attribute as account_flag.
    - Fields from this dimension can't be grouped with any other dimension by affinity. Hence it is a junk dimension. reopened_ccount_flag is a flag which is present if the account was closed previously for critical reasons and then opened again after solving the issues.

Junk Dimension:

Attributes which doesn't belong to any dimension are grouped together logically and then placed in a dimension called junk dimension. Junk dimensions are mostly avoided.

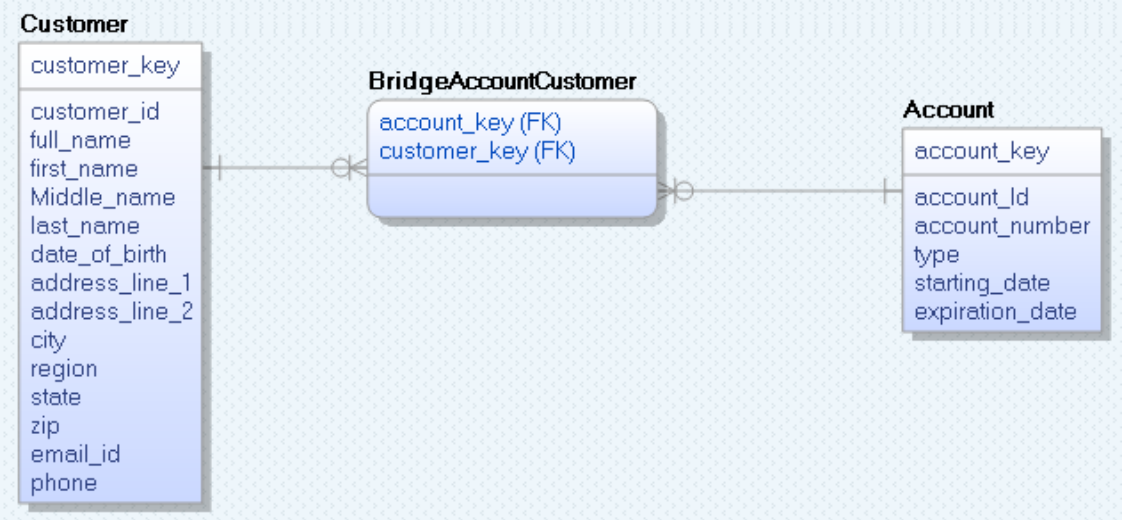
7. **Services**:
    - It stores information of services provided by bank as per account type of customer.
    - service_key is the surrogate key for this dimension. It contains attributes like service_id, services_type, category_name and category_code.
    - This dimension is a shared conformed dimension between account_fact and transaction_fact.
    - This is also an affinity dimension as it contains all similar logical attributes.

8. **Region**:

- It stores the demographic information of bank.
- Here region_key is a surrogate key which acts as a primary key for region dimension. It has attributes as region_id, region_code, region_manager and zip_code.
- Using this dimension we can get information like which region has which manager. Under which manager the maximum number of transactions occur. We can also find which region provides which services.

Bridge:

- BridgeAccountCustomer is used to establish a relationship between Customer and Account.



- A Bridge is a special table that is a Child Table of 2 Dimension Tables.
- As the customer is a child of account, it could not be directly linked to fact table. So the bridge was introduced between account and customer dimension. Here the bridge contains only foreign keys of customer and account.
- If a customer has multiple accounts, we can find them easily using bridge.