Dimensional Modeling
Sample Case Studies (Solution)

You are required to design a Dimensional Model in the way that it fulfills the requirement for the following systems.

Grocery System (POS)
The following queries shall be generated through your design:

1. Total sales of a particular product from all stores in the last quarter
2. Total sales by product by store by month
3. Yearly profit generated by stores in the north region
4. How customer deviates from store to store with particular products
5. When I promote one thing how does it affect the other
6. Check if more Products are sold on 1st 10 days and 20th to 25th date of the month than the whole month.
7. Average daily sales (in dollars) of product categories.
8. The total number of customers purchasing a particular product.
9. The total number of customers visiting a particular store in a month.
10. Count how many people buy with coupon.

For detail please refer to [1].

Bank (Account System)
The following queries shall be generated through your design:

1. We want to see five years of historical data on every account. For all prior months it will be sufficient to see the end of month snapshot.
2. For the current month, we want a valid snapshot as of yesterday. We don't need the other prior days in the current month.
3. Every type of account has a primary balance. There is a significant need to group different kinds of accounts in the same analyses and compare primary balances.
4. Every type of account has a list of custom dimension attributes and custom numeric facts that tend to be quite different from account type to account type.
5. Every account is deemed to belong to a household. Upon studying the historical production data, we conclude that accounts and the individuals who own the accounts come and go from households as much as several times per year for each household.
6. Since some of the accounts were created many years ago, and by different production systems, we find that our records of the individual account holder' names and addresses differ from account to account in many case.
7. In addition to the household identification, we are very interested in demographic information as it pertains to both the individual account holders and the households. We also capture and store behavior scores relating to the activity in each of the accounts.

For detail please refer to [2].

Design Requirements

Here is the eight points of the complete dimensional modeling design:

1. The processes, and hence the identity of the fact tables
2. The grain of each fact table.
3. The dimensions of each fact table
4. The facts, including pre-calculated facts
5. The dimension attributes with complete descriptions and proper terminology
6. How to track slowly changing dimensions
7. The historical duration of the database
8. The urgency with which the data is extracted and loaded into the data warehouse.

For details related to design requirements points please refer to [3].

Reference

[1] Chapter 2, The Data Warehouse Toolkit, Ralph Kimpball

[2] Chapter 7, The Data Warehouse Toolkit, Ralph Kimpball

[3] Chapter 12, The Data Warehouse Toolkit, Ralph Kimpball

Grocery System (POS):

1. The processes, and hence the identity of the fact tables

Following are the major processes in the Grocery System Data ware:
Sales, inventories, Cost, revenue, Buyer (Customer) etc

Identities of Fact Tables:

Base Fact Table:

i) time_key
ii) Product_key
iii) Store_key

Quarterly Agg Fact Table:

i) Product_key
ii) Quarter_key
iii) Store_key

Monthly Aggregate Fact Table:

i) Month_key
ii) Product_key
iii) Store_key

Yearly Agg Fact Table:
i) Product Key
ii) Store Key
iii) Yearly_key
iv) Region_key

Customer Fact Table:
i) Customer key
ii) Product key
iii) Store key
iv) Promo key
v) Time key

2. The grain of each fact table:

The grain of a fact table is the least level of each dimension.

Base Fact Table:
The grain level is the fact measurements by Day by Store and By Product Wise

Quarterly Agg Fact Table:

The grain level is the sales by product by quarter by All store wise.
Monthly Aggregate Fact Table:

The grain of this table is Total sales by product by store by month
Yearly Agg Fact Table:
The gain level of this fact table is sales by year by product by Region wise.
Customer Fact Table:

The grain level of this fact table is by day ,by store by customer by promotion wise.

3. The dimensions of each fact table:

Base Fact Table:
Time, Product, Store

Quarterly Agg Fact Table:

Quarter, Product, Store

Yearly Agg Fact Table:
Year, Product, Region

Customer Fact Table:
Time, Product, Store, Promotion, Customer

4. The facts, including pre-calculated facts:

The major facts and pre calculated facts are

i) Quantity_sold
ii) Dollar profit
iii) Cost
iv) Average_sales
v) Quartely_qauntiy_sold
vi) Quartely_cost
vii) Yearly_profit
viii) Customer Count
ix) Product Count
x) Product Quantity_per Customer

5. The dimension attributes with complete descriptions and proper terminology:

i) Time Dimension

Time_key It is the primary key of the time Dimension
Day Is keeps the day info
Week_key It stores the primary key of the week dimension
Day_one_to_ten It keeps dates from 1st to 10th
Day_twenty_to_twentyfive It keeps dates from 20 to 25

ii) Product Dimension

Product_key it is the primary key of the product dimension
SKU_number It is the SKU number of the production
SKU_Desc It is the Description of the SKU

Package_size It keeps the size of package of the product.
Brand It stores the Brand of the product
Subcategory IT stores the subcategory of the product.
Category
Department
Diet It keeps the diet info of the product.
Weight_type It is the weight of the product
Cases_per_pallet It keeps cases per pallet.

iii) Store Dimension

Store_number it stores the store number of the store
Store_key It is the primary key of the store dimension
Store_address it keeps the address of the Store
City
Country
State
Zip
Manager
Phone
Fax
Floor_plan_type It stores the current floor plan type of the store
Region_key It stores the Region key of the Region level dimension

iv) Customer Dimension:

Customer_key It is the primary key of the customer dimension
Customer_SSN It keeps the SSN no of the customer for tracking
Customer_name It keeps the name of the customer
Customer_phone_no
Customer_region

v) Promotions Dimension

Promotion_key It is the primary key of the promotion Dimension
Promotion_name It is the name of the promotion
Price_reduction_type
Ad_type It stores the advertisement type of the promotion
Display_type It stores the display type of the promotion
Cupon_type It stores the coupon types offered in the promotion
Ad_media_name It stores the media name of the advertisement
Display_provider It stores the provider of the displayer.
Prmo_cost It keeps the cost of the promotion
Promo_begin_date It keeps the start date of promotion
Promo_end_date It keeps the end date of promotion

vi) Quarter Dimension:
Quarter_key it is the key of the Quarter Dimension
Quarter
Year_key

vii) Year Dimension

Year_key
Year
viii) Region Dimension

Region_key
Region

ix) Week Dimension

Week_key
Week
Month_key

x) Month Dimension

Month_key
Month
Quarter_key

6. How to track slowly changing dimensions:

We will track slowly changing Dimensions by Type TWO approach where we generate a new account record every time a meaniningful account attribute changes.

7. The historical duration of the database:

The historical duration of database is 7 to 8 years approximately however it varies according to the type of Database and under certain requirements and constraints.

8. The urgency with which the data is extracted and loaded into the data warehouse:

The urgency with which the data is extracted and loaded into the data ware is daily that for 24/7 data ware houses mostly.

Bank (Account System):

1. The processes, and hence the identity of the fact tables

Following are the major processes in the Bank Account System:
Accounts, household, transactions, interests, debits, credits, balances, profit, etc

Identities of Fact Tables:

Base Fact Table:

i) Account_key
ii) Branch_key
iii) Household_key
iv) Time_key
v) Status_key

Checking Fact Table:

i) checkingacckey
ii) account_key
iii) status_key
iv) household_key
v) branch_key

Savings Fact Table:

i) saving_acc_key
ii) account_key
iii) status_key
iv) household_key
v) branch_key

Credit Card Facttable:

i) credit_card_key
ii) account_key
iii) status_key
iv) household_key
v) branch_key

Safe Deposit Fact Table:

i) safe_dep_key
ii) account_key
iii) Status_key
iv) Household_key
v) Branch_key
vi)

Time Deposit Fact Table:

i) Time_dep_key
ii) account_key
iii) Status_key
iv) Household_key

v) Branch_key

2. The grain of each fact table:

The grain of a fact table is the least level of each dimension.

Base Fact Table:
The grain of a fact table is by account by branch by status by product and by household. By month

Checking Fact Table:

The grain of a fact table is by account by branch by status by Checking Account product and by household by month

Savings Fact Table:
The grain of a fact table is By account by branch by status by saving account product and by household by month
Credit Card Fact Table:

The grain of a fact table is By account by branch by status by product Credit cards and by household by month.

Safe Deposit Fact Table:
The grain of a fact table is By account by branch by status by product Safe deposits and by household by month.
Time Deposit Fact Table:

The grain of a fact table is By account by branch by status by product time deposts and by household by month.

3. The dimensions of each fact table:

Base Fact Table:
Time, Product, Accounts, HouseHold, Status, Branch

Checking Fact Table:

Time, Product, Accounts, HouseHold, Status, Branch

Savings Fact Table:

Time, Product, Accounts, HouseHold, Status, Branch

Credit Card Fact Table:

Time, Product, Accounts, HouseHold, Status, Branch
Safe Deposit Fact Table:

Time, Product, Accounts, HouseHold, Status, Branch

4. The facts, including pre-calculated facts:

The major facts and pre calculated facts are

i) Quantity_sold
ii) Dollar profit
iii) Cost
iv) Average_sales
v) Quartely_qauntiy_sold
vi) Quartely_cost
vii) Yearly_profit
viii) Customer Count
ix) Product Count
x) Product Quantity_per Customer

5. The dimension attributes with complete descriptions and proper terminology:

i) Time Dimension

Time_key
Day
Week_key
Day_one_to_ten
Day_twenty_to_twentyfive

ii) Product Dimension

Product_key
SKU_number
SKU_Desc
Package_size
Brand
Subcategory
Category
Department
Diet
Weight_type
Cases_per_pallet

iii) Store Dimension

Store_number
Store_key
Store_address
City
Country
State
Zip
Manager
Phone
Fax
Floor_plan_type
Region_key

iv) Customer Dimension:

Customer_key
Customer_SSN
Customer_name
Customer_phone_no
Customer_region

v) Promotions Dimension

Promotion_key
Promotion_name
Price_reduction_type
Ad_type
Display_type
Cupon_type
Ad_media_name
Display_provider
Prmo_cost
Promo_begin_date
Promo_end_date

vi) Quarter Dimension:
Quarter_key
Quarter
Year_key

vii) Year Dimension

Year_key
Year

viii) Region Dimension

Region_key
Region

ix) Week Dimension
Week_key
Week
Month_key

x) Month Dimension

Month_key
Moth
Quarter_key

6. How to track slowly changing dimensions:

We will track slowly changing Dimensions by Type TWO approach where we generate a new account record every time a meaningful account attribute changes.

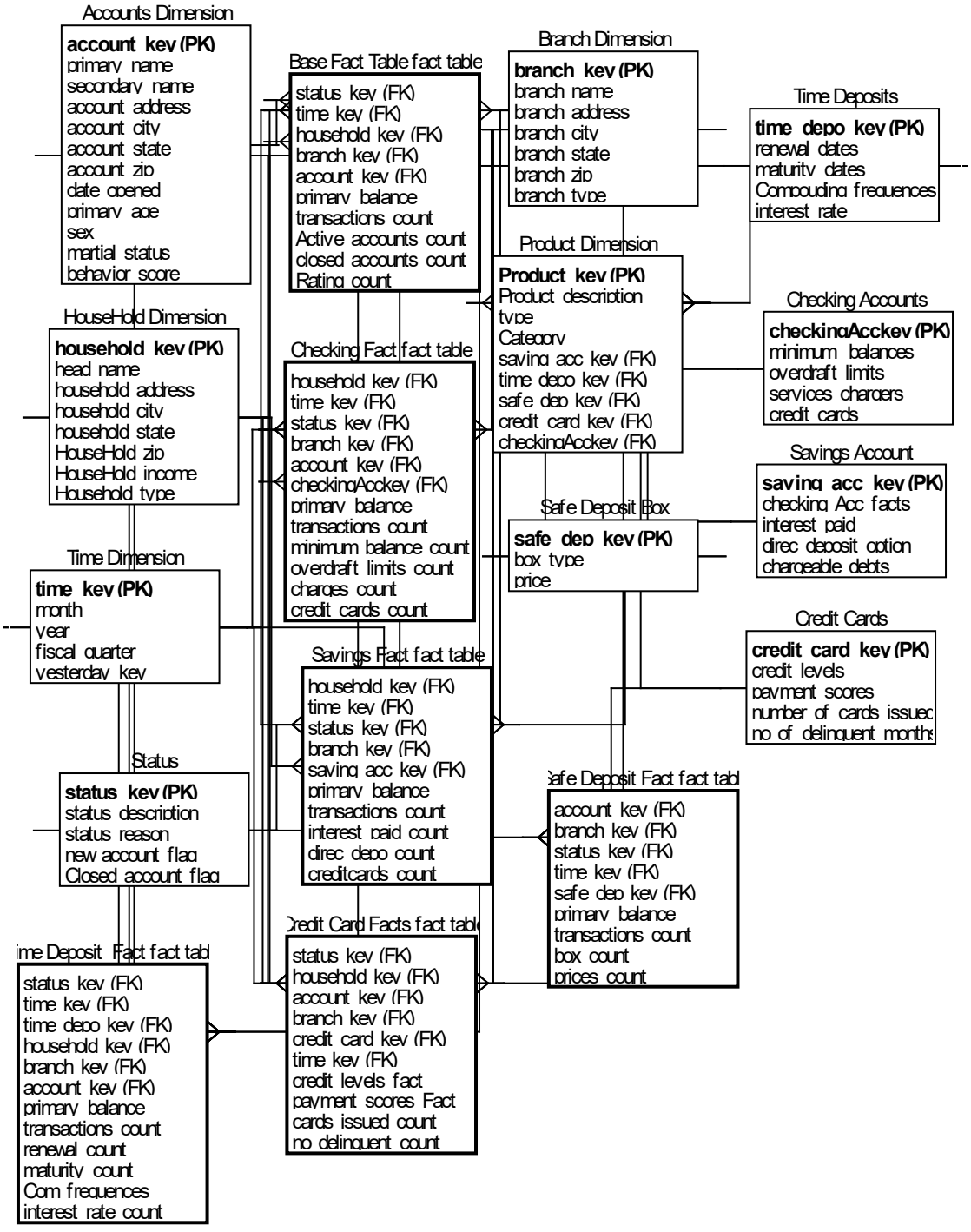
7. The historical duration of the database:

The historical duration of database is 7 to 8 years approximately however it varies according to the type of Database and under certain requirements and constraints.

8. The urgency with which the data is extracted and loaded into the data warehouse:

The urgency with which the data is extracted and loaded into the data ware is daily that for 24/7 data ware houses mostly.

Dimensional data Model - Grocery System (POS)



Dimensional data Model - Bank (Account System)