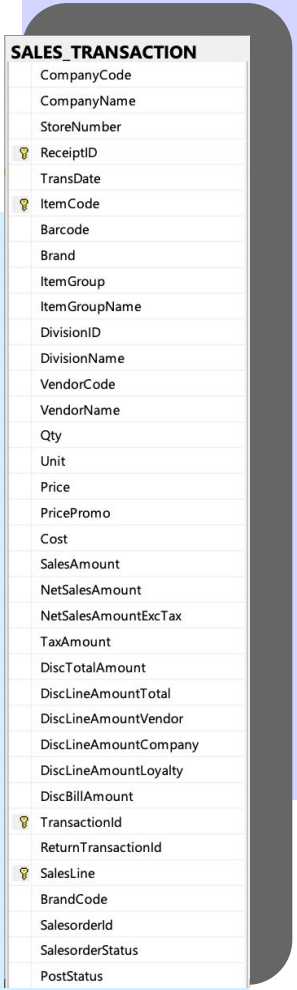
Data Warehouse Design and Implementation

For Retail Company

Nattharika Sae Tang 6031748621
Natakorn Ammy Kam 6031772621
Phongsun Worrawattanapreecha 6031795021
Pachara Pattarabodee 6031796721

Data Inspect

Clean, Mock, Analyse



Sales_Transaction

INSPECT

Primary Key

TransDate

ReceiptID

SalesLine

CLEAN

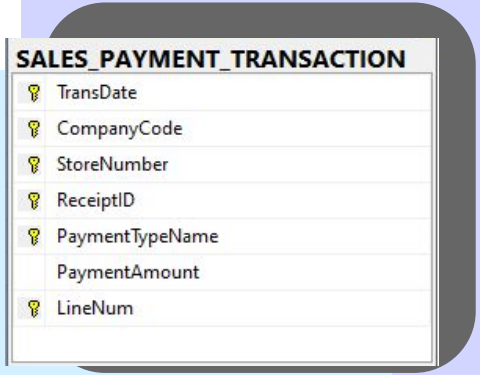
No data clean

MOCK

Add column (**TransactionType**)

Type	Condition
Buy	(**Qty**) > 0, (**NetSalesAmount**) >0 , (**Price**) > 0
Full Discount	(**SalesAmount**) = (**DiscTotalAmount**), (**Price**) > 0
Giveaway	(**NetSalesAmount**) = 0, (**Price**) = 0, **Cost** = 0
Defected	**NetSalesAmount**<0, **Cost**>0
Return	Have (**ReturnTransactionId**)
Error	Other

Sales_Payment Transaction



INSPECT

- This table tells the payment type (**PaymentTypeName**) and its number of payment in a receipt (**LineNum**)

TransDate	CompanyCode	StoreNumber	ReceiptID	PaymentTypeName	PaymentAmount	LineNum
2016-04-13 13:16:35.000	200	2022	2022102-1002845	Credit/Debit Card	27000.0000000000000000	1.0000000000000000
2016-04-13 13:16:35.000	200	2022	2022102-1002845	Credit/Debit Card	106650.0000000000000000	2.0000000000000000

- At the same receipt, it can have multiple number of payment with the same payment type (ex. First credit card met its limit)

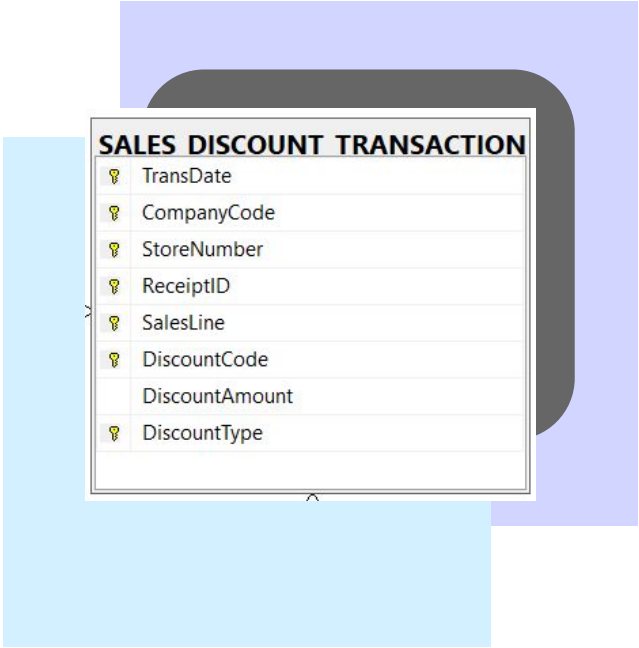
What did we clean?

No data clean

What did we mock?

No data mock

Sales_Discount Transaction



INSPECT

- Data contains only the receipt where discount is involved
- Connected to Discount Master table

What did we clean?

No data clean

What did we mock?

No data mock

Sales_Transaction Summary

INSPECT

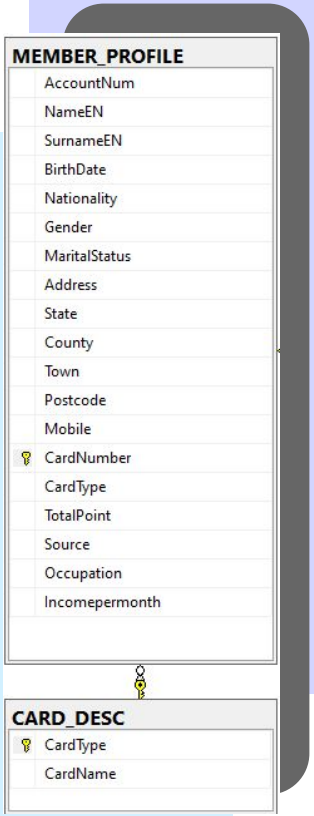
- Data contains only summary of those with membership.
- Redundant (**StoreNumber**) and (**TenantID**) column as the type of shop is stated in (**RetailType**) column.
- (**CompanyCode**) column exists only when the transaction summary comes from Retail

What did we clean?

No data clean

What did we mock?

No data mock



Member_Profile

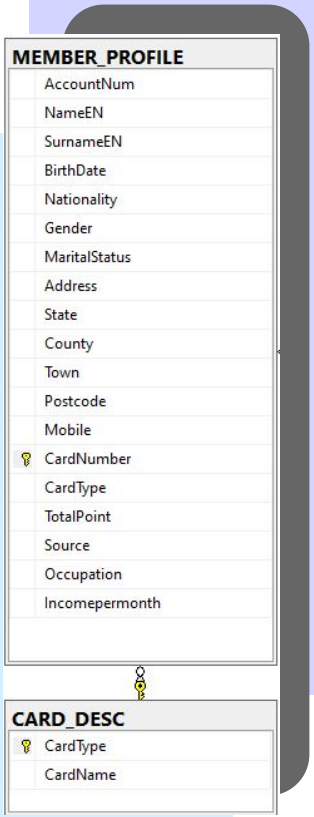
INSPECT

AccountNum	NameEN	SurnameEN	BirthDate
-	Seve	Madura	1954-03-17
-	Seve	Madura	1954-03-17
BN	Julia	Butel	1944-02-18
999999	Mahmud	Phi	NULL
92040000253	Scevola	Vanvuren	NULL

- AccountNum has missing number and contain **invalid** value
- BirthDate has multiple **NULL**

Nationality	Gender	MaritalStatus	Address	State	Postcode	CardNumber	CardType	TotalPoint	Source	Occupation
TH	Female	Married	2 Soccer Embankm...	CO	80539	15028163IM	15	17217.0000000000000...	Application	NULL
THA	NonSpecific	Single	32 Shulman Rd.	RI	02816	15000198IM	15	0.0000000000000000	Website	NULL
NULL	Male	None	NULL	NULL	NULL	0000000241	01	0.0000000000000000	Event	NULL

- Nationality has redundant value → **TH, THA**
- Nationality, Address, State, Postcode, Occupation has **NULL**



Member_Profile

CLEAN

- Replace invalid AccountNum
- Replace nationality from having both TH and THA to only **TH**
- Replace **NULL** value with 'NonSpecific''

MOCK

- Mock NULL BirthDate with **random date** from year 1961 to 2006 which is range of member at age 15 to 60

Card_Desc

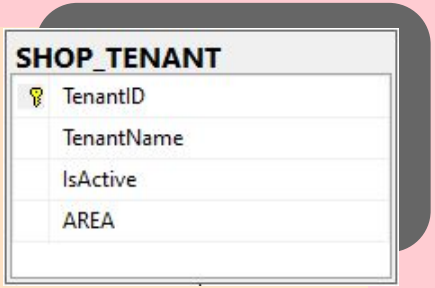
CLEAN

No data clean

MOCK

No data mock

Shop_Tenant



INSPECT

- (**TenantName**) column contains redundant information as (**AREA**) column.

2	1	ก่วยเตี๋ยวเรือยกพลๆ - Oasis*	0	Oasis
3	864	World Study Center-Oasis Haha*	0	Oasis

- (**TenantName**) column contains area name (**4AM**) in (**TenantName**) column.

175	749	Love Mo4AMhino	0	4AM

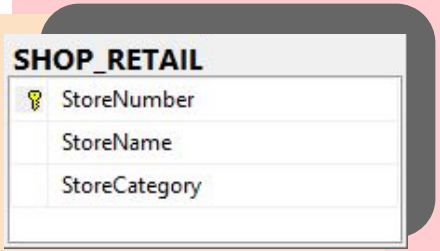
What did we clean?

- Remove area information from (**TenantName**) column
- From inspection, Replace '4AM' with 'sch' in (**TenanName**) column

What did we mock?

No data mock

Shop_Retail



INSPECT

- (**StoreName**) column contains description for the store state.

11	209920	(NOT USE)PM 4PM Outside	PM
12	209921	(NOT USE)PM The Square Outside	PM
13	209922	(NOT USE)PM The Park Outside	PM
14	209924	(NOT USE)PM ZPELL Outside	PM

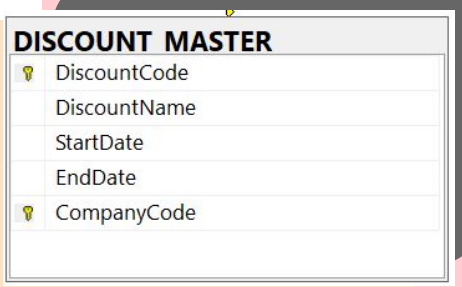
What did we clean?

- Remove description for store state in (**StoreName**) column.

What did we mock?

- Add (**isActive**) column to describe the store state.

Discount_Master



INSPECT

- Discounts that have no (**StartDate**) and (**EndDate**) are considered as discounts that last forever.

DiscountCode	DiscountName	StartDate	EndDate	CompanyCode
BKS_disc5p	NULL	NULL	NULL	200
CITI_10p	NULL	NULL	NULL	203
CITI_disc10p	NULL	NULL	NULL	203
CITI_disc5p	NULL	NULL	NULL	204

- Discounts that that have an empty (**DiscountCode**) are considered as discounts from Pink Pomelo.

DiscountCode	DiscountName	StartDate	EndDate	CompanyCode
	NULL	NULL	NULL	200
	NULL	NULL	NULL	203
	NULL	NULL	NULL	204

What did we clean?

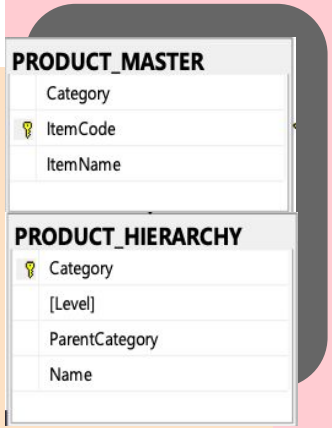
No data clean

What did we mock?

No data mock

Product_Master

Product_Hierarchy



INSPECT

- (**Category**) in Product_Master is not always Level 6

Category	ItemCode	ItemName
5637242849	1081814	NBC008 ของเล่น Alpaca 495

Category	Level	ParentCategory	Name
5637242849	2	5637145423	90000 - ECO

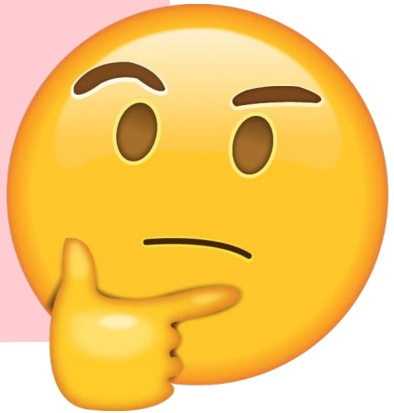
What did we Clean?

No data clean

What did we mock?

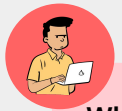
No data mock

Curiosity!?

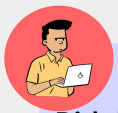


Did the fraud happen in the retail store?

Did the retail store make profit?



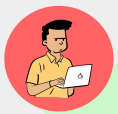
What type of customers spend more money?



Did the product sales increase during weekend?

Did the tenant store sell better than retail?

Which promotion is the most popular?



Which season (month) has the most sales?

Which vendors are heavily depend on?

What type of product sell the best?

4-Step Design

Business Process, Grain, Dim, Measure,
Design Techniques

BUS MATRIX

Business Process \ Common Dimension	Account Dim	Account to Member Bridge	Company Dim	Date Dim	Discount Master Dim	Member Dim	Member Demographic Dim	Month Dim	Payment Type Dim	Product Dim	Salesline Dim	Store Dim	Store Retail Dim	Store Tenant Dim	Store Tenant Area Dim	Store Tenant Status Dim	Time Dim	Transaction Type Dim	Vendor Dim
Promotion Fact				✔	✔	✔				✔	✔		✔				✔	✔	✔
Product Fact				✔						✔	✔		✔				✔	✔	✔
Aggregate Product Fact						✔		✔		✔			✔						✔
Member Fact	✔	✔	✔	✔		✔	✔		✔			✔					✔		
Store Fact				✔										✔	✔	✔			

Store Fact

1 Business Process

Tenant Store Analysis

2 Granularity

1 row
per 1 sales summary in each store
per day

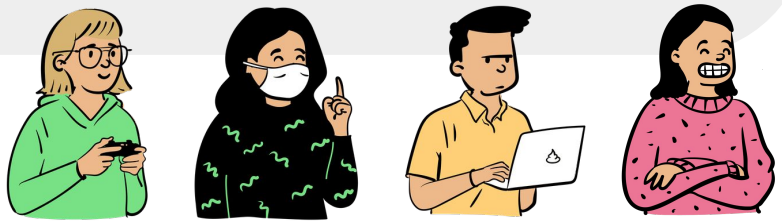
3 Dimensions

- StoreTenant_Dim
- StoreTenantArea_Dim
- StoreTenantStatus_Dim
- Date_Dim

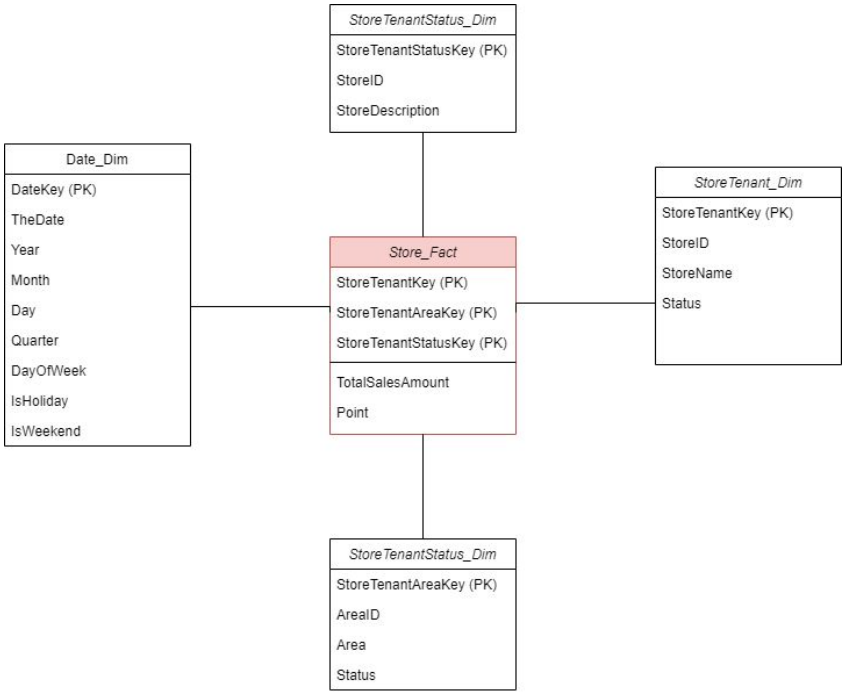
4 Measures

- TotalSalesAmount
- Point

1 Business Process
2 Granularity
3 Dimensions
4 Measures

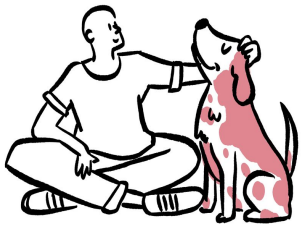


Store Fact

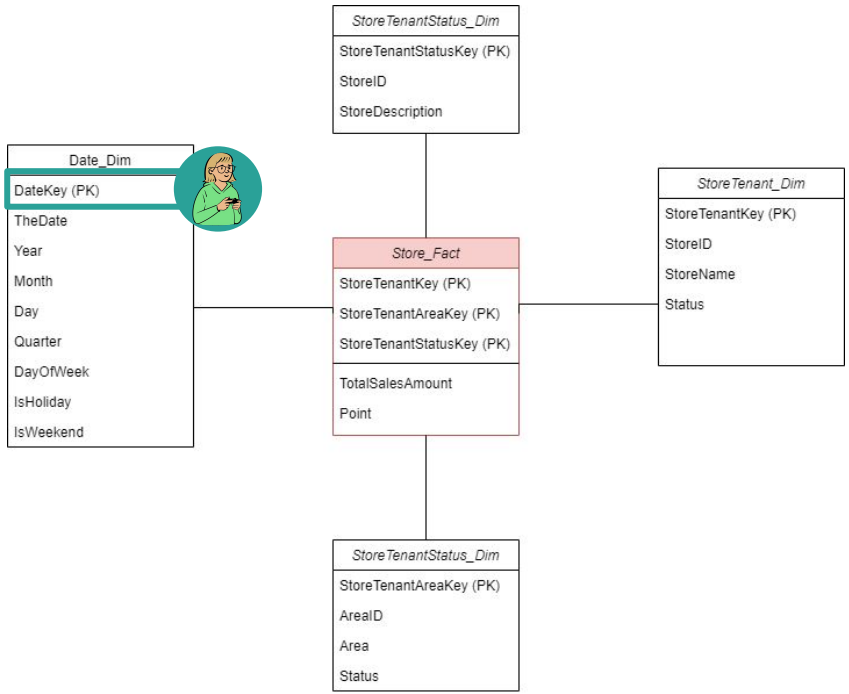


1 What can "Store Fact" tell us?

- Which tenant stores sell the highest?
- What is the highest tenant store sell?
- How many tenant store are active or inactive?
- Which mall contains the most tenant store?
- How much point does the member get from tenant store?
- Does the tenant sell better in weekend?
- Does the tenant sell better in holiday?
- Does the area affects the sale for tenant store?



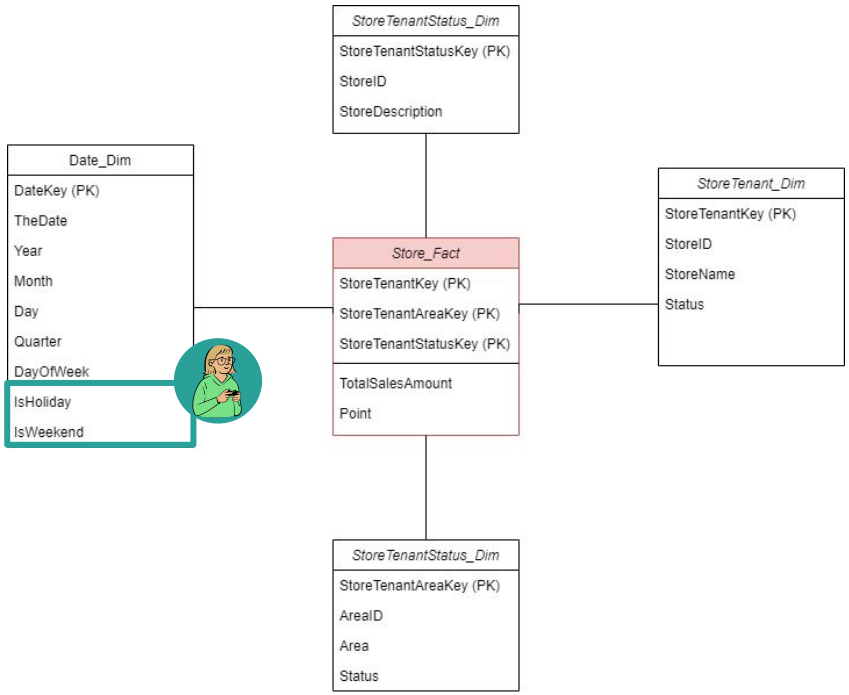
Store Fact



2 Techniques

- **Smart Key** (Date_Dim: DateKey)

  Instead of joining back to Date_Dim to get the date, we could use the datekey as the date to avoid slow query from joining.



Store Fact

2 Techniques

- **Smart Key** (Date_Dim: DateKey)

  Instead of joining back to Date_Dim to get the date, we could use the datekey as the date to avoid slow query from joining.

- **Textual Attributes** (Date_Dim: isHoliday, isWeekend)

  Both column uses the descriptive text instead of flags or indicators.

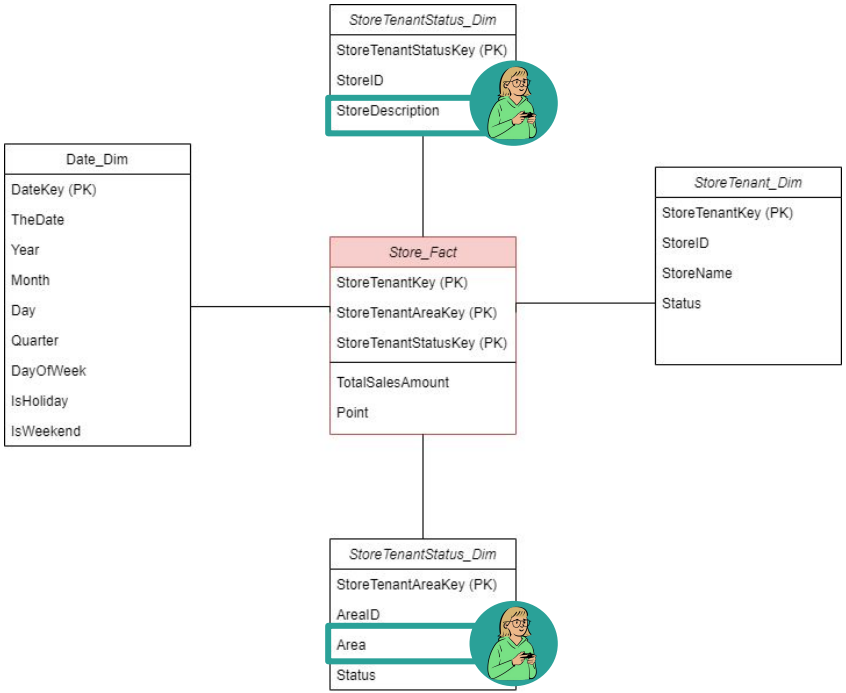
	DateKey	TheDate	Year	Month	Day	Quarter	DayOfWeek	IsHoliday	IsWeekend
1	20160127	2016-01-27	2016	January	27	Q1	Wednesday	Non-Holiday	Weekday
2	20160129	2016-01-29	2016	January	29	Q1	Friday	Non-Holiday	Weekday
3	20160201	2016-02-01	2016	February	1	Q1	Monday	Non-Holiday	Weekday

Smart Key

Textual Attribute

103	20150413	2015-04-13	2015	4	13	Q2	Monday	Holiday	Weekday
104	20150414	2015-04-14	2015	4	14	Q2	Tuesday	Holiday	Weekday
105	20150415	2015-04-15	2015	4	15	Q2	Wednesday	Holiday	Weekday

Store Fact



2 Techniques

- **Avoid Too Few Dimension** (StoreTenant_Dim)

  At first, table StoreTenant_Dim contains area and status information, but in order to avoid too few dimension those information are separated from the table.

Table StoreTenant_Dim

- contains basic information for tenant store

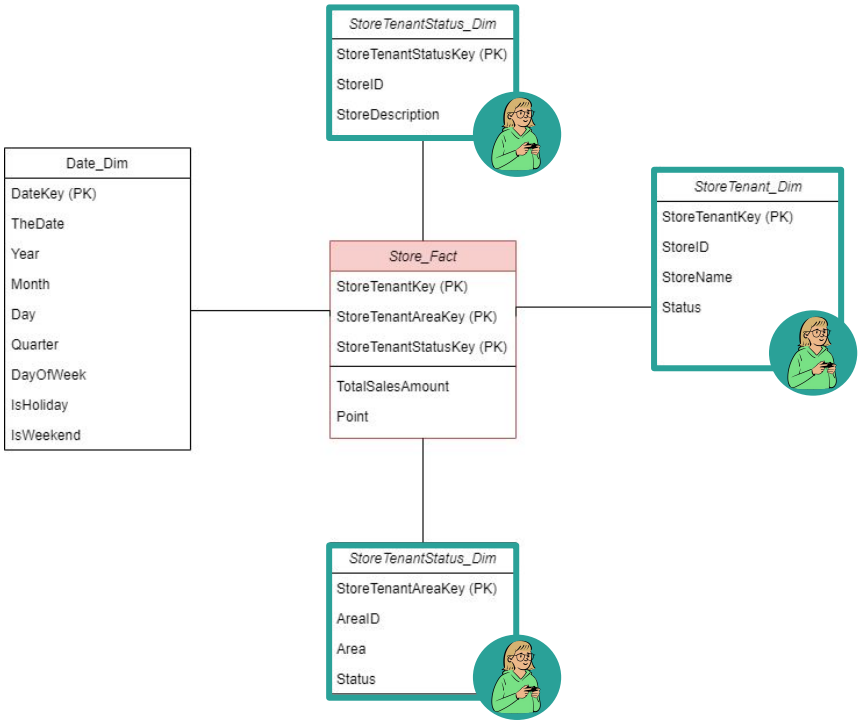
Table StoreTenantStatus_Dim

- contains status information for tenant store

Table StoreTenantArea_Dim

- contains area information for tenant store

Store Fact



3 Slowly Changing Dimension

- **SCD2** (*StoreTenant_Dim*)
  - *Track of the Store Name*
- **SCD2** (*StoreTenantArea_Dim*)
  - *Track of the Area*
- **SCD1** (*StoreTenantStatus_Dim*)
  - *Track of the Status Description*

4 Remark

- **(StoreID) column in StoreTenantStatus_Dim:**
  - *1 → Active*
  - *0 → InActive*

Product Fact

1 Business Process

Product Transaction Analysis

2 Granularity

1 row per 1 item in a receipt

3 Dimensions

- Product_Dim
- Date_Dim
- Time_Dim
- SalesLine_Dim
- Vendor_Dim
- StoreRetail_Dim
- TransactionType_Dim

4 Measures

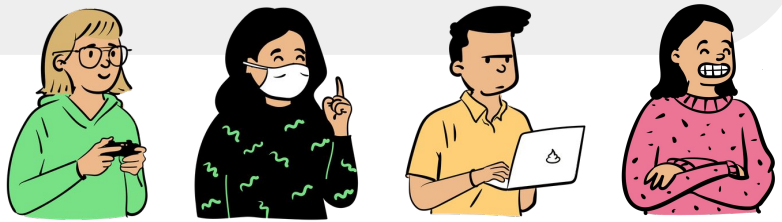
- QTY
- NetSalesAmount
- NetSalesAmountExcTax
- TotalCost = ABS(QTY) * Cost
- GrossProfit = NetSalesAmount - TotalCost
- NetProfit= NetSalesAmountExcTax - TotalCost

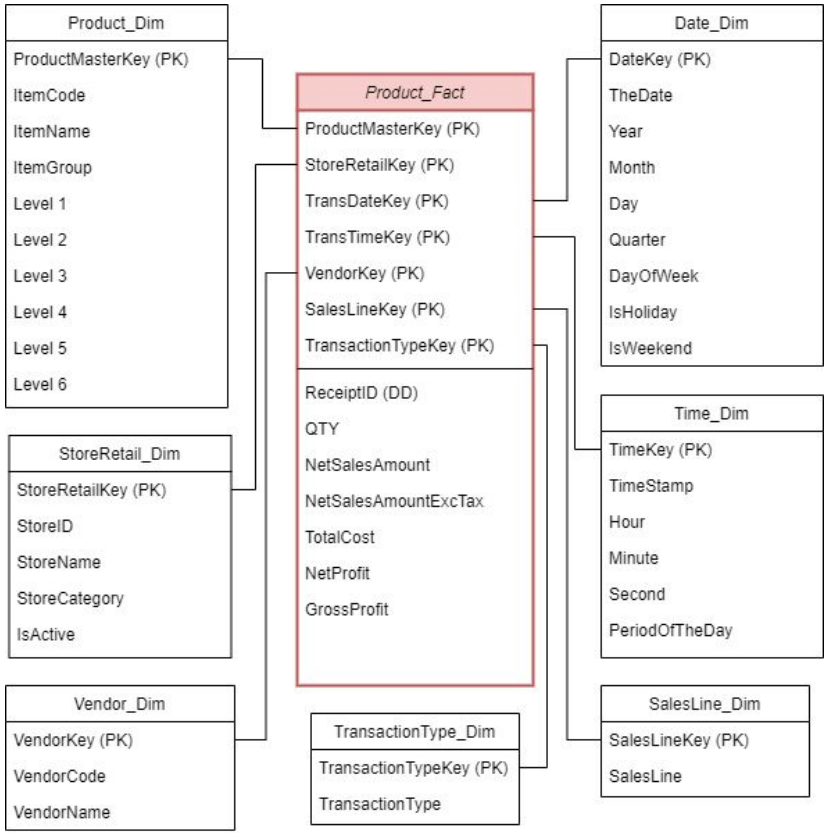
1 Business Process

2 Granularity

3 Dimensions

4 Measures

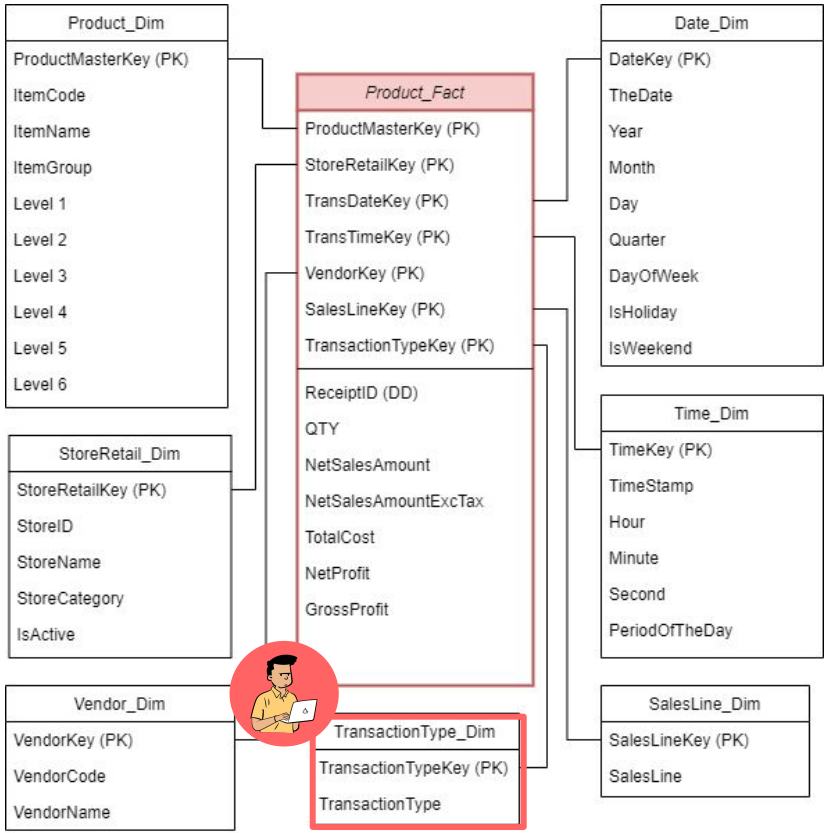




Product Fact

1 What can "Product Fact" tell us?

- Which products are frequently sold in a particular store?
- Are profits from selling products being generated more on weekends or holidays?
- Which product generated the most revenue in a particular store?
- Which product generated the most profit in a particular store?
- What is the best selling product?
- Which products are frequently bought from a particular vendor?

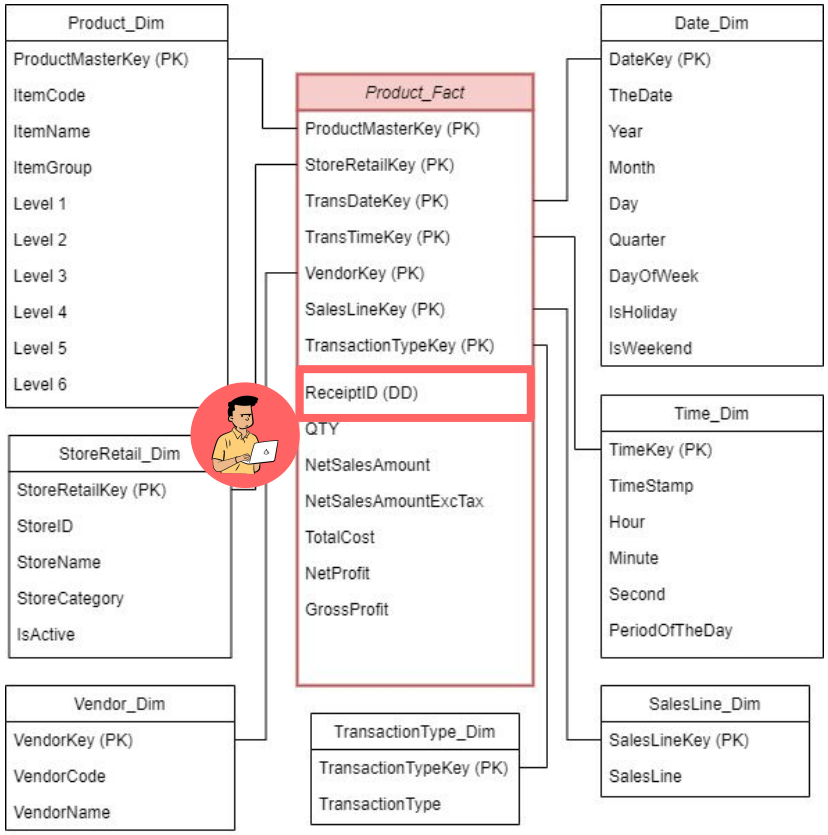


Product Fact

2 Techniques

- **Transaction Type Fact** (Product_Fact, TransactionType_Dim)

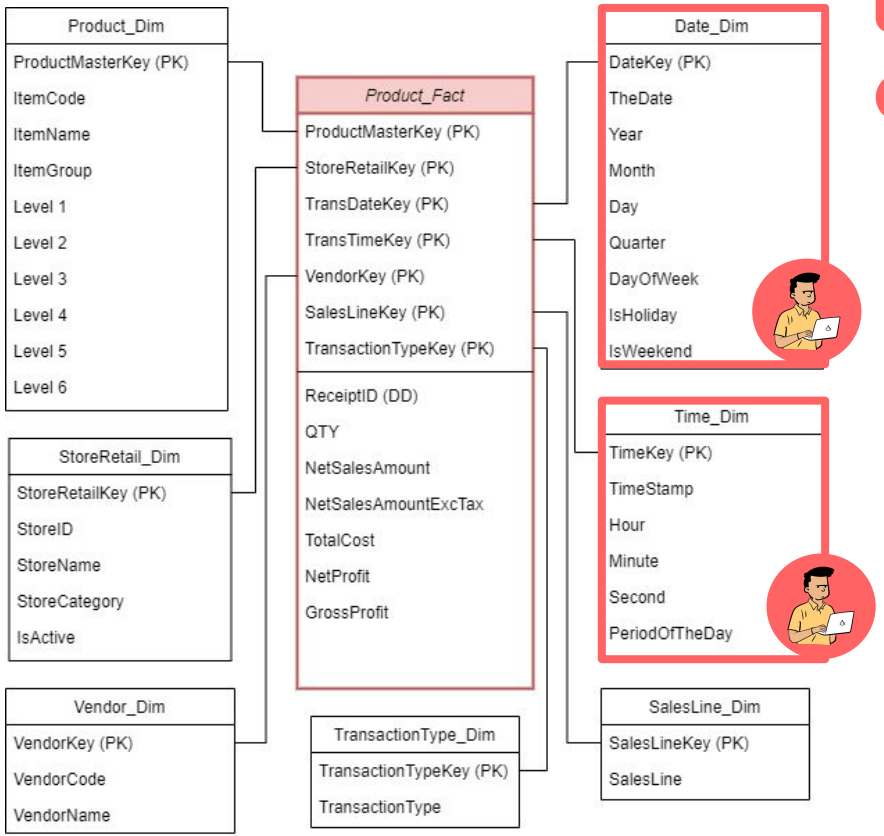
Type	Condition
Buy	(**Qty**) > 0, (**NetSalesAmount**) >0 , (**Price**) > 0
Full Discount	(**SalesAmount**) = (**DiscTotalAmount**), (**Price**) > 0
Giveaway	(**NetSalesAmount**) = 0, (**Price**) = 0, **Cost** = 0
Defected	**NetSalesAmount**<0, **Cost**>0
Return	Have (**ReturnTransactionId**)
Error	Other



Product Fact

2 Techniques

- **Transaction Type Fact** (Product_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID)

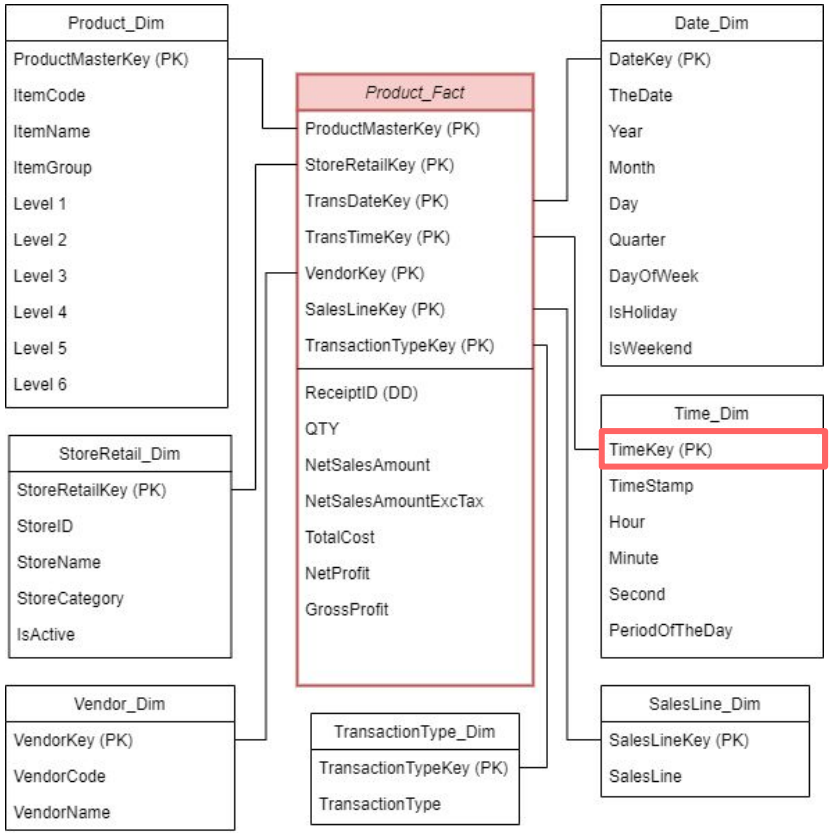


Product Fact

2 Techniques

- **Transaction Type Fact** (Product_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID)
- **Separate Date_Dim and Time_Dim**

  Since the format of the Transaction Date in the original database is datetime, we separated the datetime into Date_Dim and Time_Dim in order to reduce the number of rows in the dimension table.



Product Fact

2 Techniques

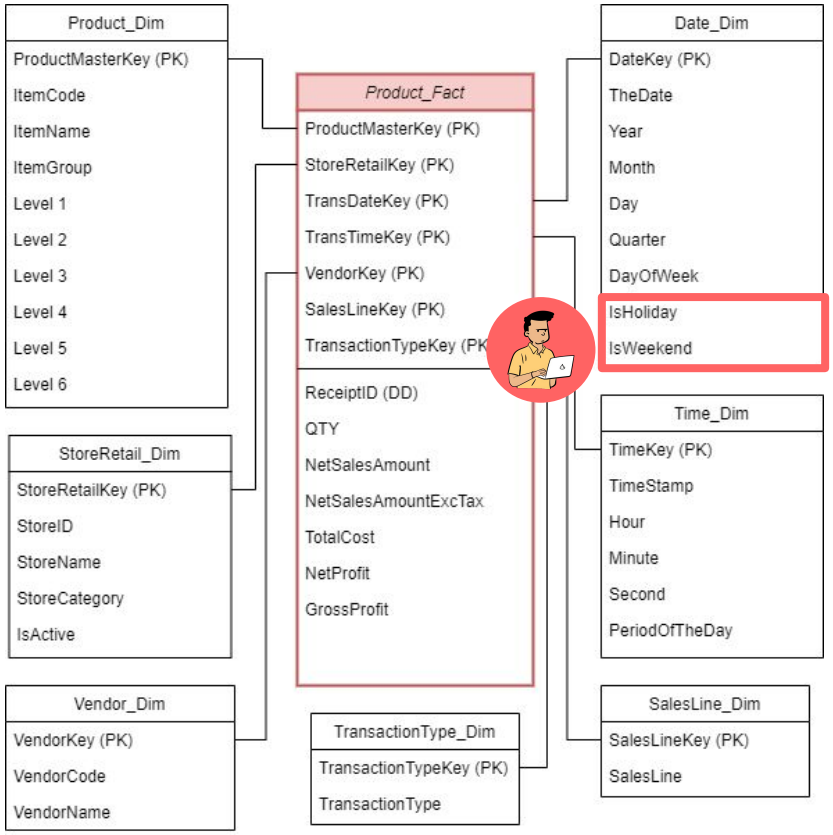
- **Transaction Type Fact** (Product_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID)
- **Separate Date_Dim and Time_Dim**

  Since the format of the Transaction Date in the original database is datetime, we separated the datetime into Date_Dim and Time_Dim in order to reduce the number of rows in the dimension table.

- **Smart Key** (Date_Dim, Time_Dim)

  Instead of joining back to Time_Dim to get the time, we also use the timekey as the time to avoid slow query from joining.

TimeKey	Timestamp	Hour	Minute	Second	PeriodOfTheDay
235944	23:59:44	23	59	44	PM
235945	23:59:45	23	59	45	PM
235946	23:59:46	23	59	46	PM



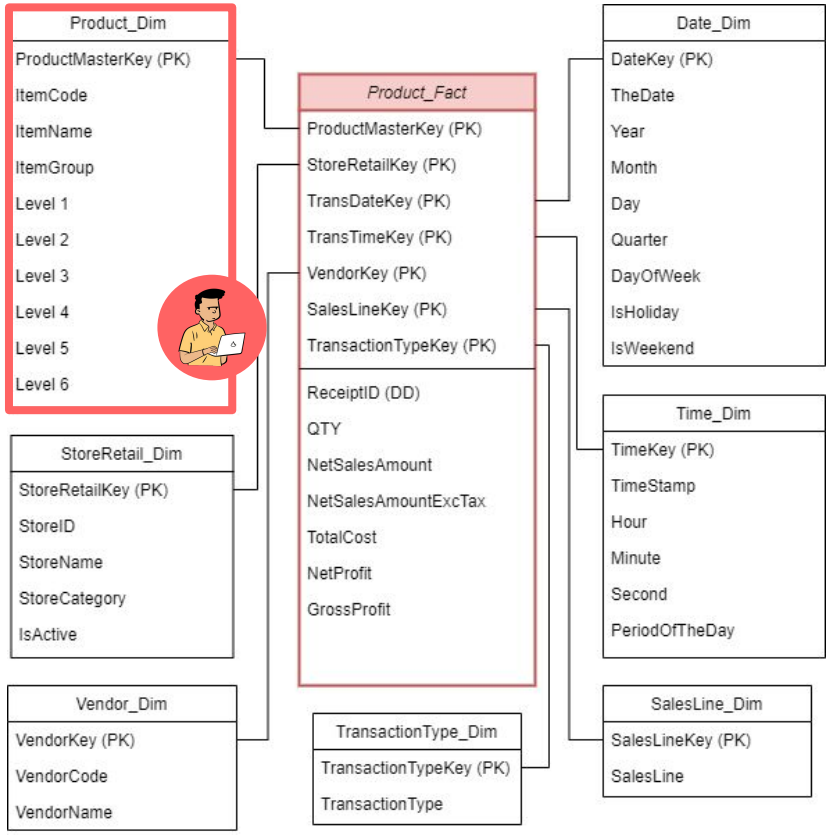
Product Fact

2 Techniques

- **Transaction Type Fact** (Product_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID)
- **Separate Date_Dim and Time_Dim**

  Since the format of the Transaction Date in the original database is datetime, we separated the datetime into Date_Dim and Time_Dim in order to reduce the number of rows in the dimension table.
- **Smart Key** (Date_Dim, Time_Dim)

  Instead of joining back to Time_Dim to get the time, we also use the timekey as the time to avoid slow query from joining.
- **Textual Attributes** (Date_Dim: isHoliday, isWeekend)



Product Fact

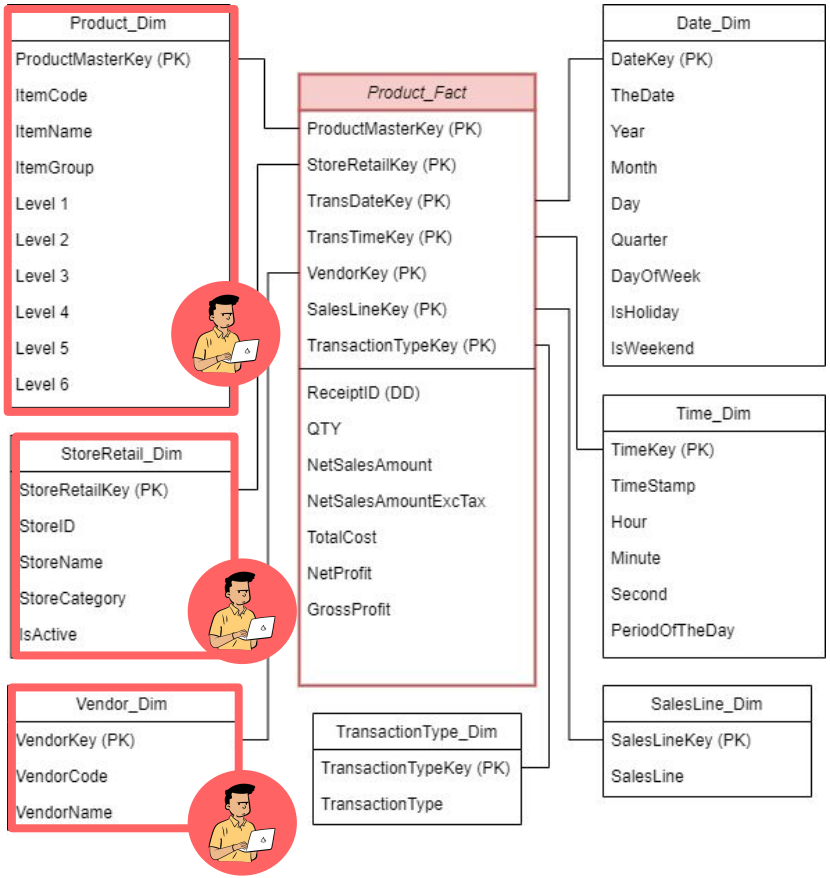
2 Techniques

- **Transaction Type Fact** (Product_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID)
- **Separate Date_Dim and Time_Dim**

  Since the format of the Transaction Date in the original database is datetime, we separated the datetime into Date_Dim and Time_Dim in order to reduce the number of rows in the dimension table.
- **Smart Key** (Date_Dim, Time_Dim)

  Instead of joining back to Time_Dim to get the time, we also use the timekey as the time to avoid slow query from joining.
- **Textual Attributes** (Date_Dim: isHoliday, isWeekend)
- **Hierarchy** (ProductMaster_Dim)

  Since there are 6 levels of product category (Fixed-Depth Hierarchies), we have added a total of 6 different attributes as a representation of each product category level.



Product Fact

3 Slowly Changing Dimension

- **Product Master Dim**
  - ***SCD1***: *Track of Item Group Name*
  - ***SCD1***: *Track of Item Name*
  - ***SCD1***: *Track of Each of Level*
- **Store Retail Dim**
  - ***SCD1***: *IsActive*
  - ***SCD2***: *Track of Store Category*
  - ***SCD2***: *Track of Store Name*
- **Vendor Dim**
  - ***SCD1***: Track of Vendor Name

4 Remark

- **TotalCost = ABS(QTY) * Cost** → total cost of the item in the receipt
- **GrossProfit = NetSalesAmount - TotalCost** → profit including tax
- **NetProfit = NetSalesAmountExcTax - TotalCost** → profit excluding tax

Aggregate Product Fact

1 Business Process

Monthly Product Sale Analysis

2 Granularity

1 row per 1 product sale in each month

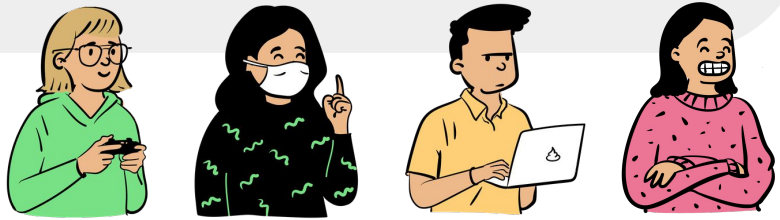
3 Dimensions

- Product_Dim
- Vendor_Dim
- StoreRetail_Dim
- Month_Dim

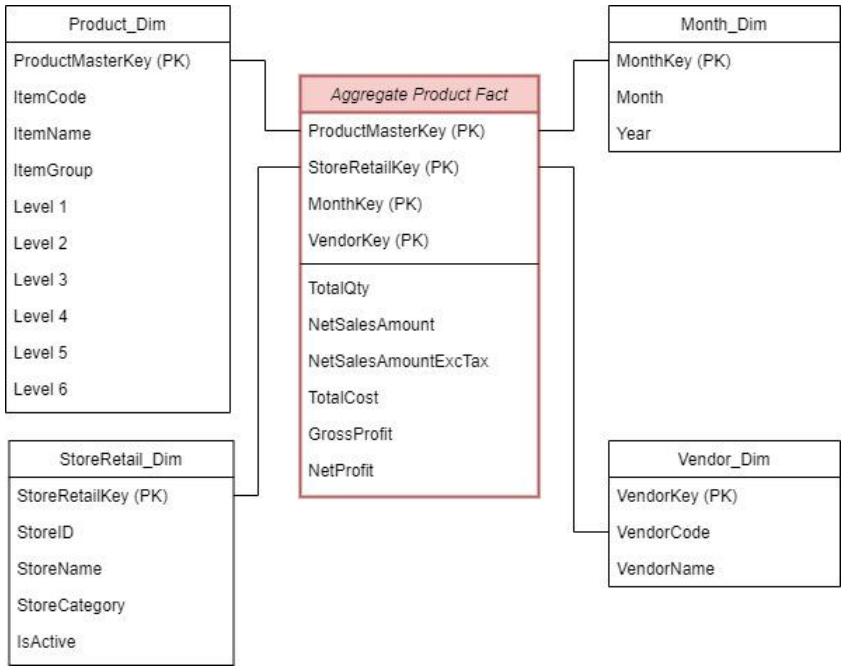
4 Measures

- TotalQty
- NetSalesAmount
- NetSalesAmountExcTax
- TotalCost
- NetProfit
- GrossProfit

1 Business Process
2 Granularity
3 Dimensions
4 Measures



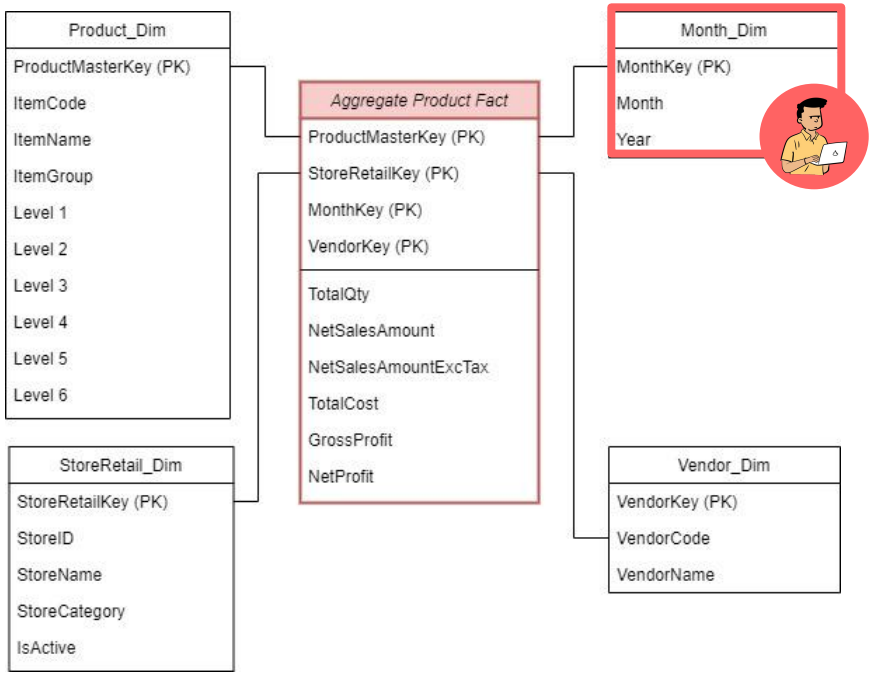
Aggregate Product Fact



1 What can "Aggregate Product Fact" tell us?

- What kind of product sell the highest in each month?
- What kind of product makes the most profit in each month?
- What kind of product makes the worst profit?
- What retail store makes the best profit in each month?
- What retail store makes the worst profit in each month?
- What vendor does we depend on the most?
- What item group makes the most profit?
- Does the retail store makes profit in each year?
- Can we make an annual report on product sales?
- Can we make a monthly report on product sales?

Aggregate Product Fact



2 Techniques

- **Aggregated Fact Table**

  This fact table is aggregated from **Product Fact** for faster query and report purpose.

  - **Lost Dimension**: TransactionType_Dim, Time_Dim, SalesLine_Dim
  - **Shrunken Dimension**: Date_dim

- **Aggregated Measure**

  The measure from the base fact table is obtained, but the value is sum by month, product, vendor, and store retail.

Member Fact

1 Business Process

Member Analysis

2 Granularity

1 row per 1 receipt per payment type

3 Dimensions

- Account_Dim
- Account-To-Member Bridge
- Member_Dim
- MemberDemographic_Dim
- Company_Dim
- PaymentType_Dim
- Date_Dim
- Time_Dim
- StoreDim

4 Measures

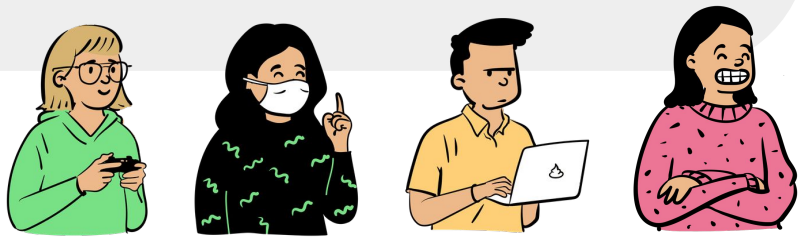
- PaymentAmount
- EarnPoint

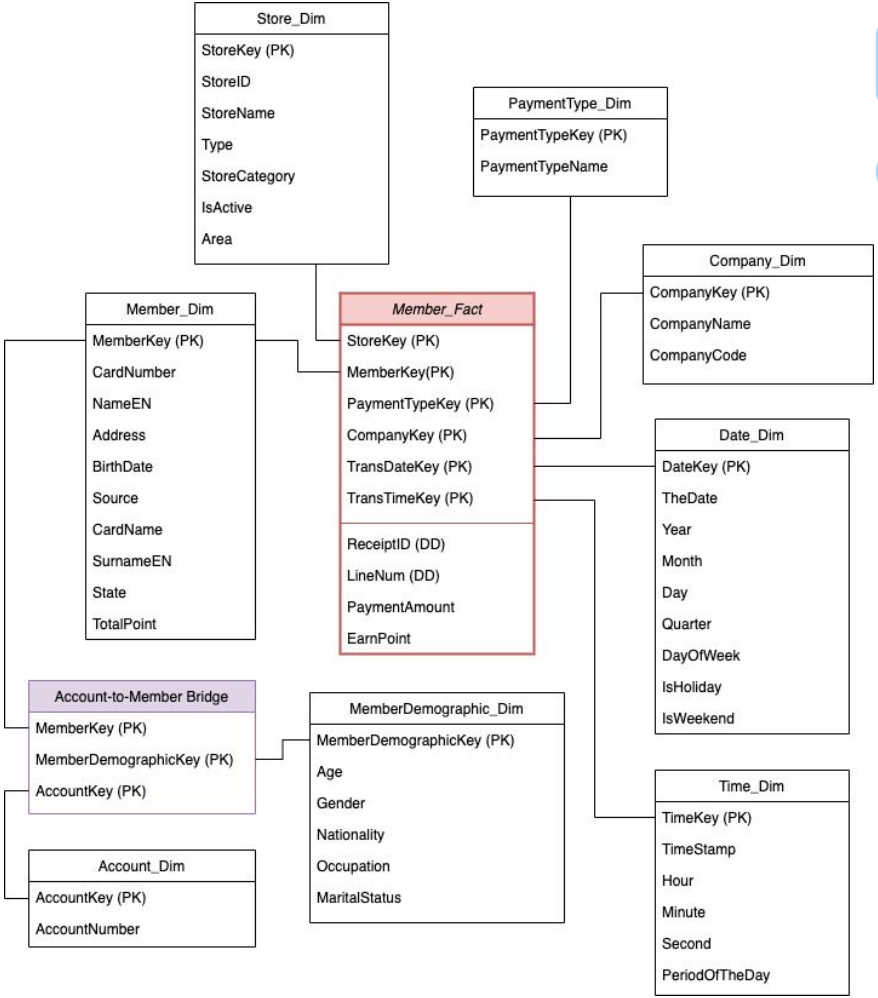
1 Business Process

2 Granularity

3 Dimensions

4 Measures

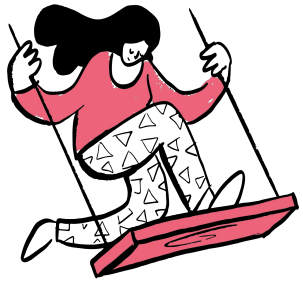


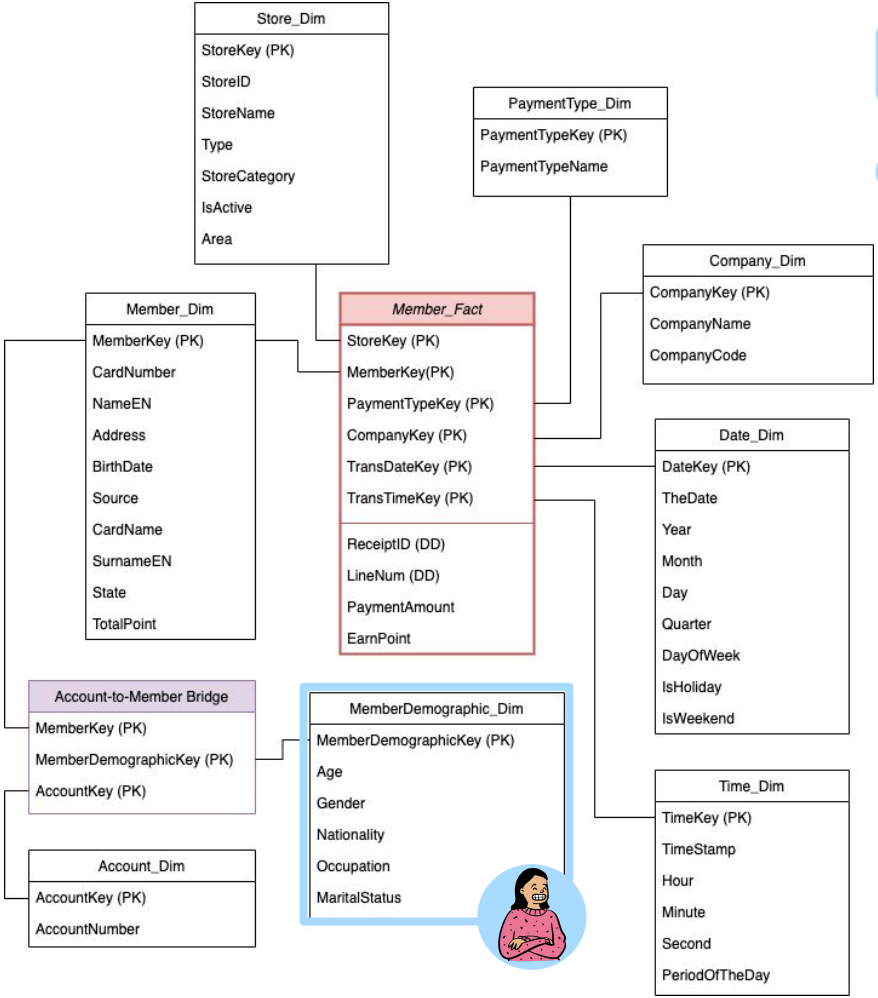


Member Fact

1 What can "Member Fact" tell us?

- Which types of member spend the most?
- Payment behavior of members, which type of payment (cash/credits/debits/points) using with which store or company?
- What type of occupation/nationality/gender/age spends with which store/company the most?
- Which range (age) of customer can/cannot register membership via application? → target that range of customer that cannot register easily and provide more approachable way to register in order to get more data from membership

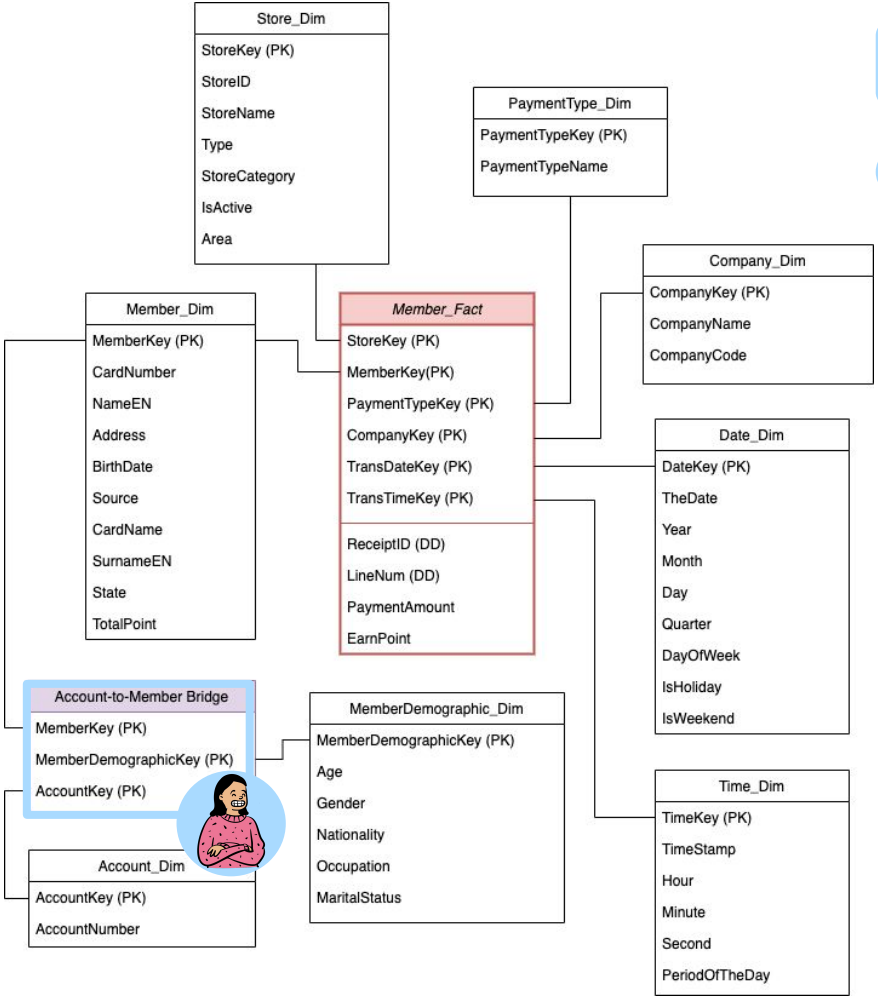




Member Fact

2 Techniques

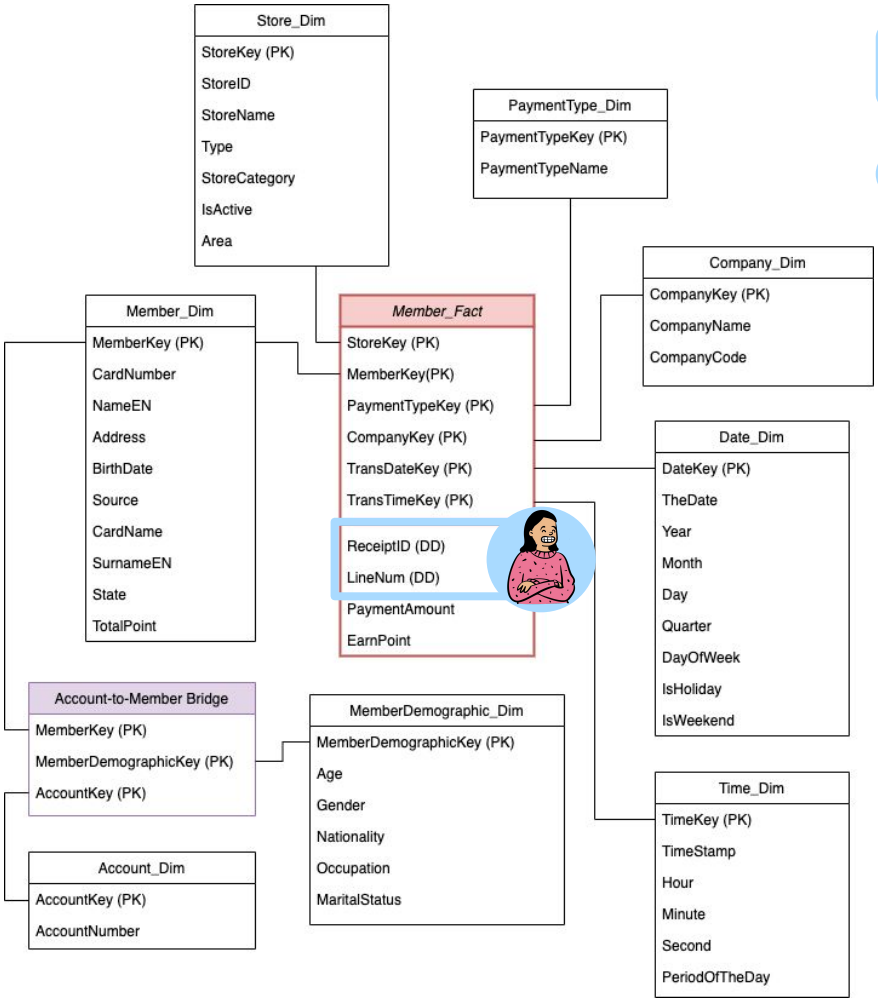
- **Mini-dimension:** MemberDemograhic_Dim
  - Contains frequently changing information that directly related to member



Member Fact

2 Techniques

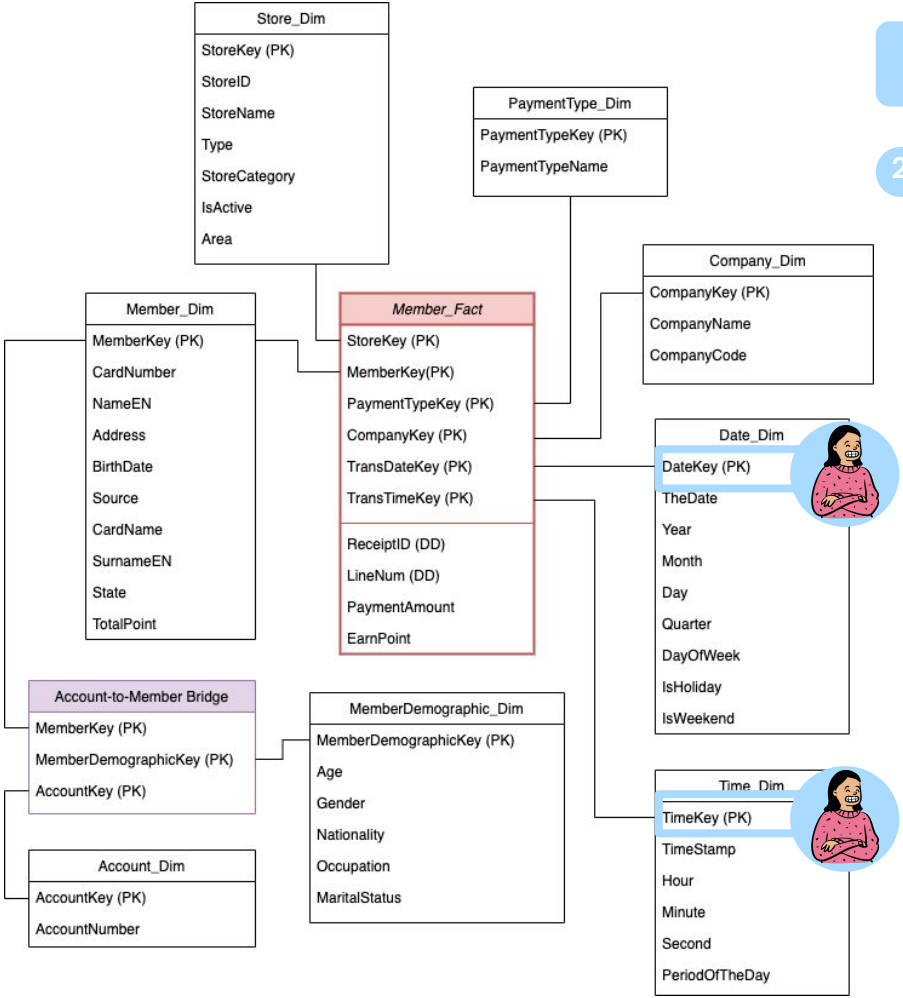
- **Mini-dimension:** MemberDemograhic_Dim
  - Contains frequently changing information that directly related to member
- **Multi-Value Dimension (MVD):** Account-to-Member Bridge Table
  - To associate each member with its account



Member Fact

2 Techniques

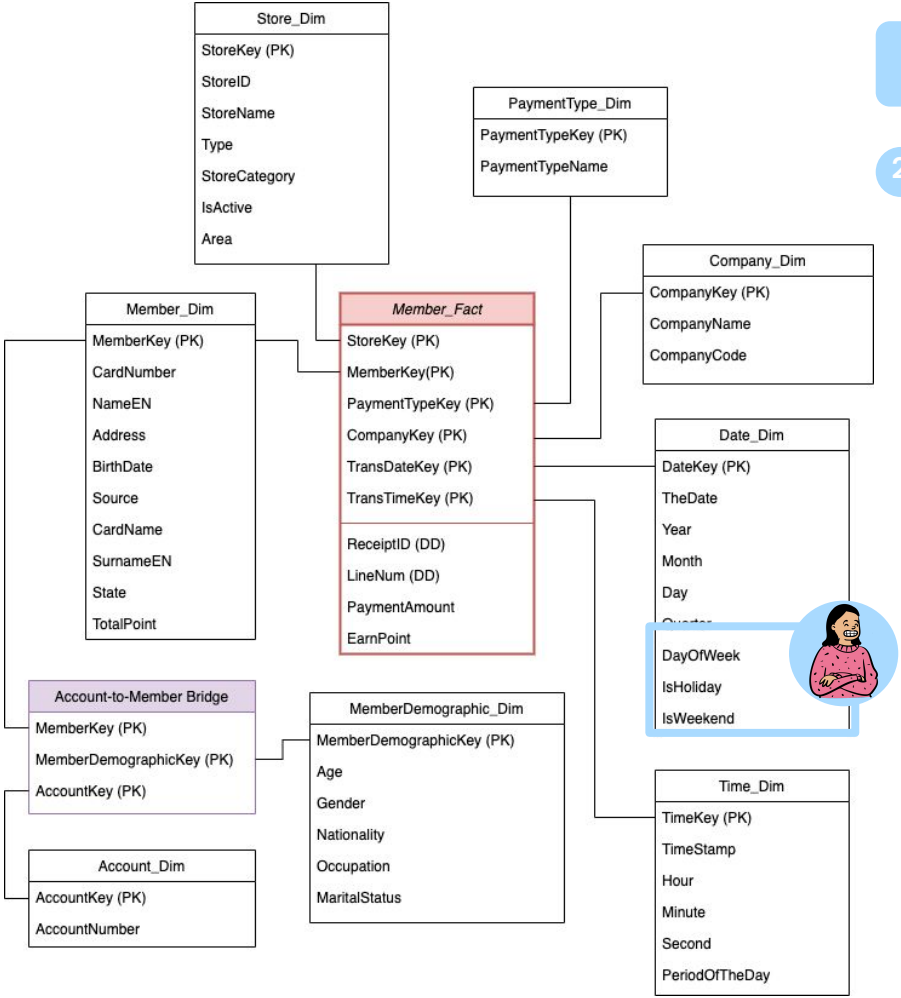
- **Multi-Value Dimension (MVD):** Account-to-Member Bridge Table
  - To associate each member with its account
- **Mini-dimension:** MemberDemograhic_Dim
  - Contains frequently changing information that directly related to member
- **Degenerate dimension (DD):** ReceiptID, LineNum
  - To refer back to the each transaction and LineNum



Member Fact

2 Techniques

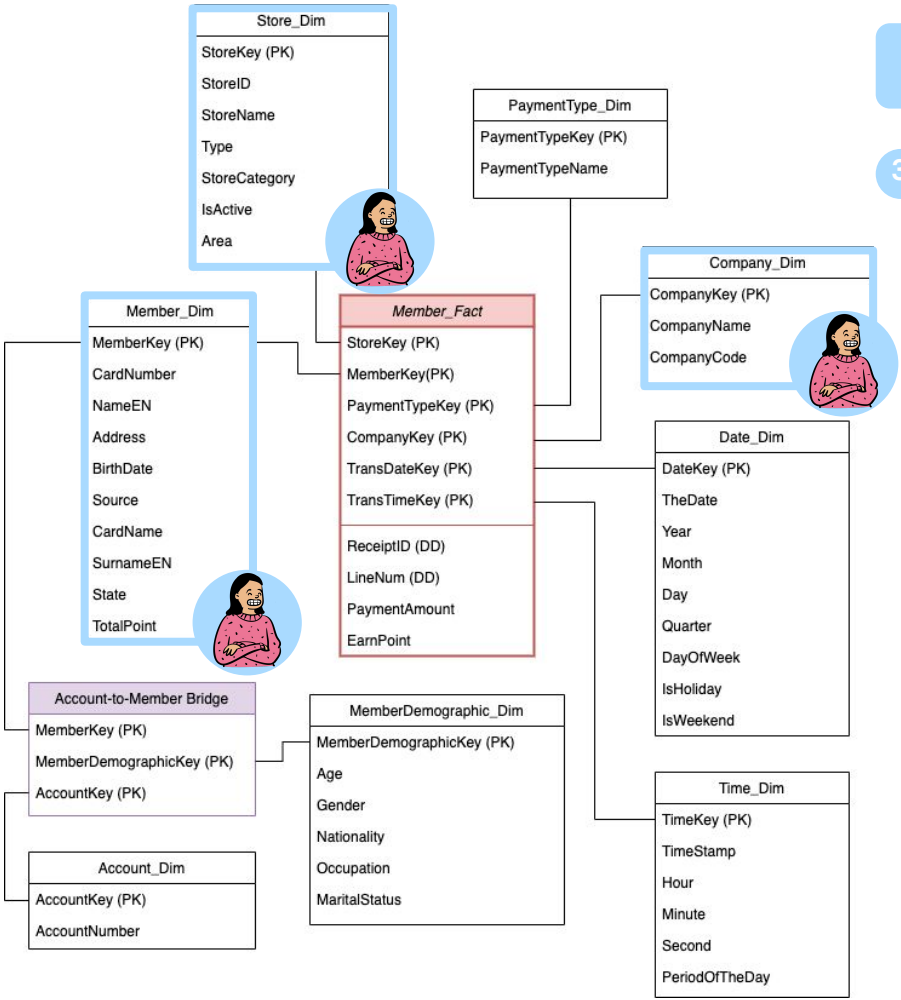
- **Multi-Value Dimension (MVD):** Account-to-Member Bridge Table
  - To associate each member with its account
- **Mini-dimension:** MemberDemograhic_Dim
  - Contains frequently changing information that directly related to member
- **Degenerate dimension (DD):** ReceiptID, LineNum
  - To refer back to the each transaction and LineNum
- **Smart Key** (Date_Dim, Time_Dim)



Member Fact

2 Techniques

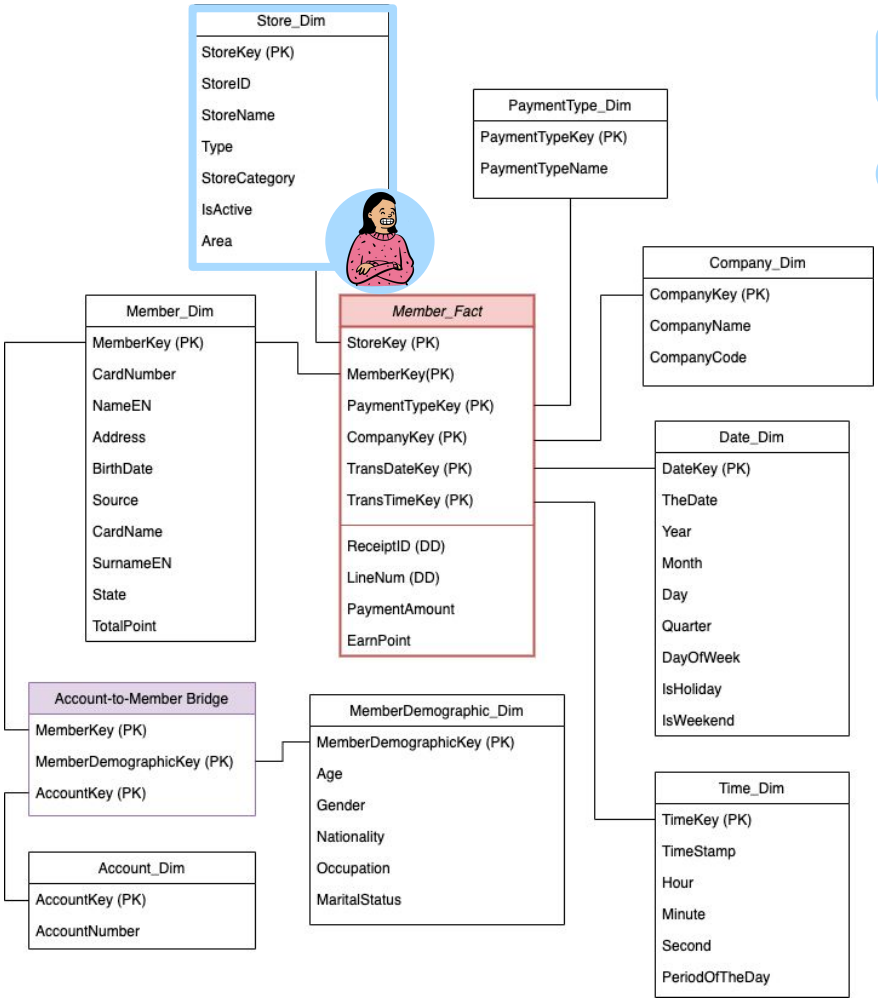
- **Multi-Value Dimension (MVD):** Account-to-Member Bridge Table
  - To associate each member with its account
- **Mini-dimension:** MemberDemograhic_Dim
  - Contains frequently changing information that directly related to member
- **Degenerate dimension (DD):** ReceiptID, LineNum
  - To refer back to the each transaction and LineNum
- **Smart Key** (Date_Dim, Time_Dim)
- **Textual Attributes** (Date_Dim: isHoliday, isWeekend)



Member Fact

3 Slowly Changing Dimensions

- **SCD1** (Company_Dim)
  - Overwrite company name
- **SCD1** (Store_Dim)
  - Overwrite store type
  - Overwrite IsActive
- **SCD1** (Member_Dim)
  - Overwrite member current TotalPoint

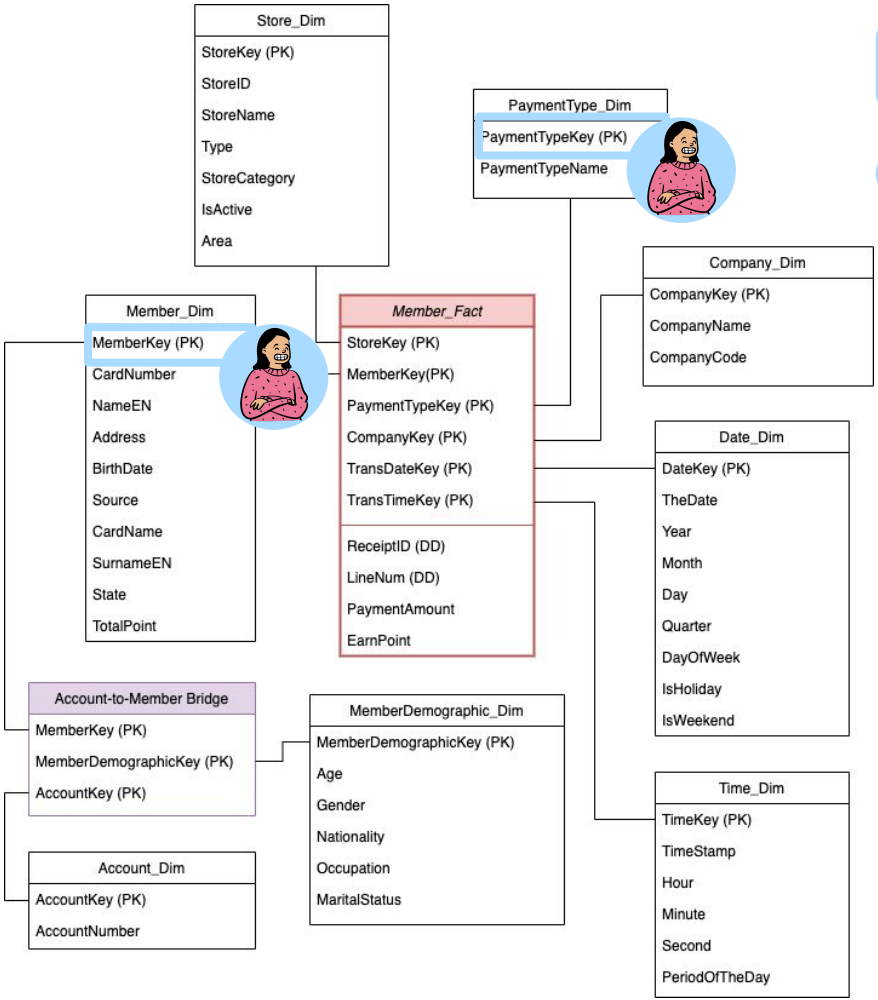


Member Fact

3 Slowly Changing Dimensions

- **SCD1** (Company_Dim)
  - Overwrite company name
- **SCD1** (Store_Dim)
  - Overwrite store type
  - Overwrite IsActive
- **SCD1** (Member_Dim)
  - Overwrite member current TotalPoint
- **SCD2** (Store_Dim)
  - Track of the store area
  - Track of the store name
  - Track of the store category

	StoreKey	StoreID	StoreName	Type	StoreCategory	IsActive	Area	tatus
4	4	2021	PM The Square	retail	PM	0	NULL	Current
5	5	2022	PM The Park	retail	PM	1	NULL	Current
6	6	2023	PM 4AM	retail	PM	0	NULL	Current
7	7	2024	PM ZPELL	retail	PM	1	NULL	Current
56	56	1	ก๋วยเตี๋ยวเรือยกพลฯ	ten...	NULL	0	Oasis	Current
57	57	10	Adidas	ten...	NULL	0	4PM	Current



Member Fact

3 Slowly Changing Dimensions

- **SCD1** (Company_Dim)
  - Overwrite company name
- **SCD1** (Store_Dim)
  - Overwrite store type
  - Overwrite IsActive
- **SCD1** (Member_Dim)
  - Overwrite member current TotalPoint
- **SCD2** (Store_Dim)
  - Track of the store area
  - Track of the store name
  - Track of the store category

4 Remark

- **PaymentTypeKey**= 0 → *not specify payment type*
- **MemberKey** = 0 → *customer is not member*

Promotion Fact

1 Business Process

Promotion Analysis

2 Granularity

1 row per 1 promotion used in 1 item on the receipt

3 Dimensions

- Product_Dim
- Vendor_Dim
- StoreRetail_Dim
- Date_Dim
- Time_Dim
- DIscountMaster_Dim
- TransactionType_Dim
- Member_Dim
- SalesLine_Dim
- ReceiptID (DD)
- DiscountType (DD)

4 Measures

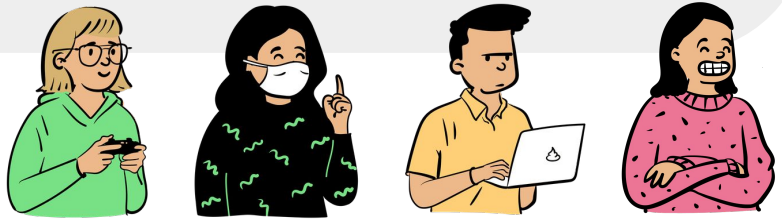
- QTY
- DiscTotalAmount
- DiscLineAmountTotal
- DiscLineAmountCompany
- DiscLineAmountVendor
- DiscLineAmountLoyalty
- DiscBillAmount

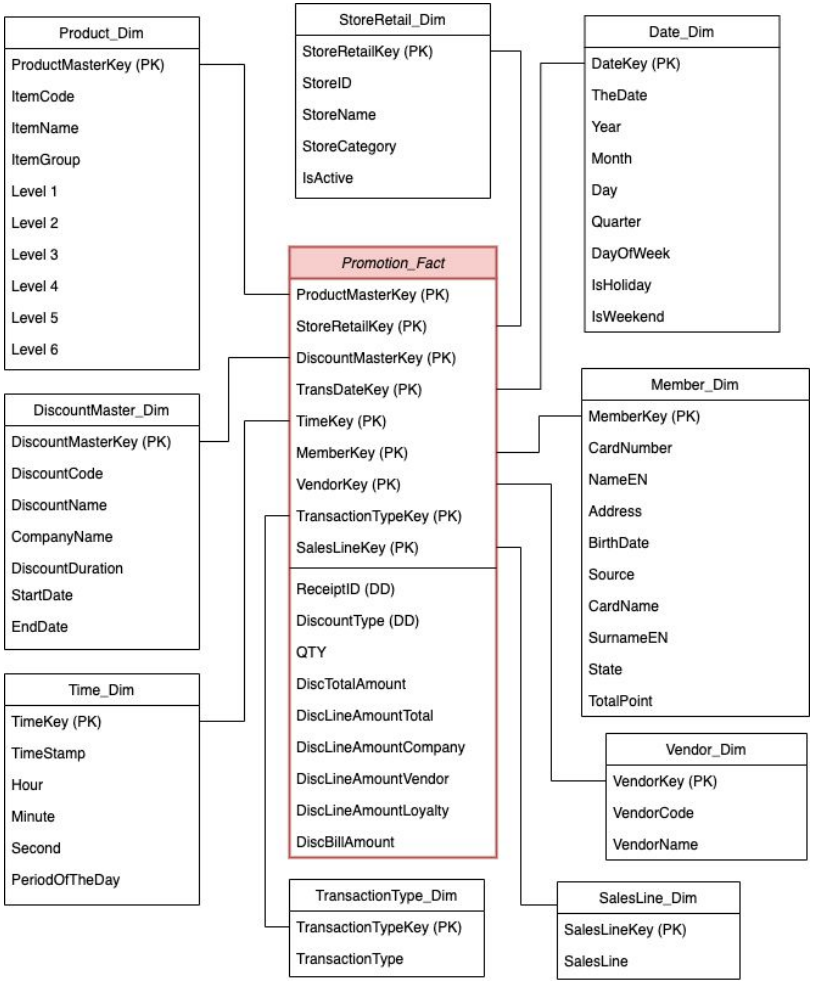
1 Business Process

2 Granularity

3 Dimensions

4 Measures

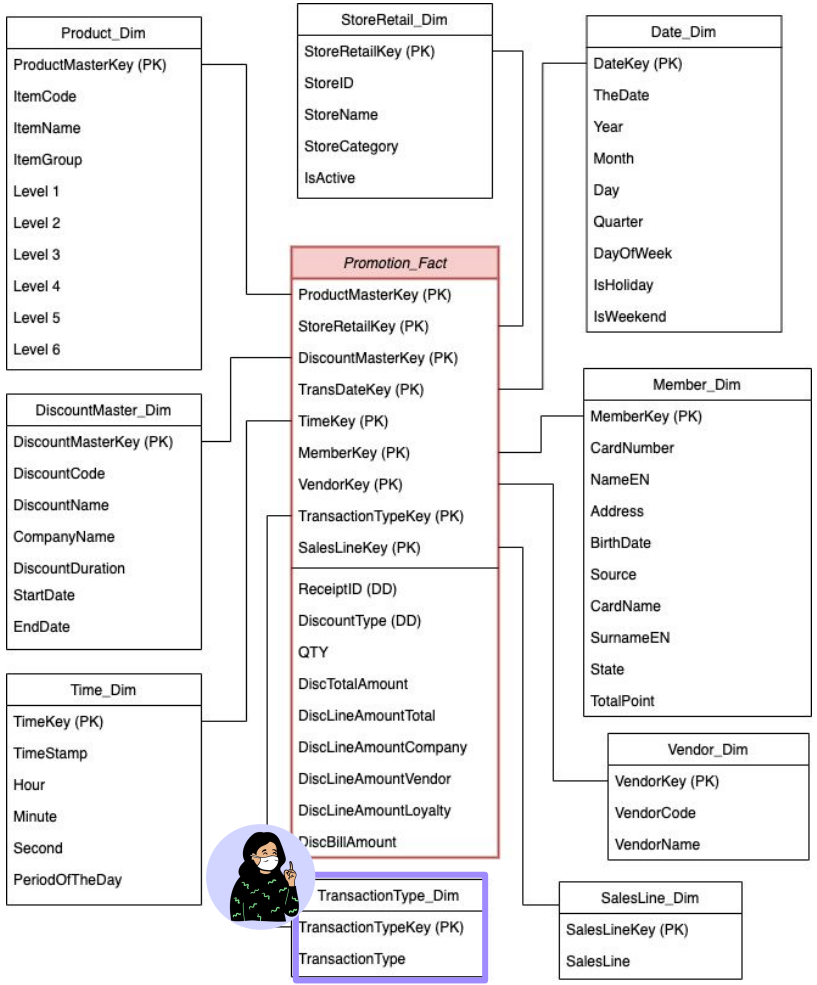




Promotion Fact

1 What can "Promotion Fact" tell us?

- What promotion is frequently use?
- On what day and time is the promotion redeem by the customer?
- Is the promotion become more popular during weekend or holiday?
- Which promotion on this product can increase sales the most?
- Which group of customers often use this promotion?
- Which retail store customer mostly redeem this promotion when they purchase the product?

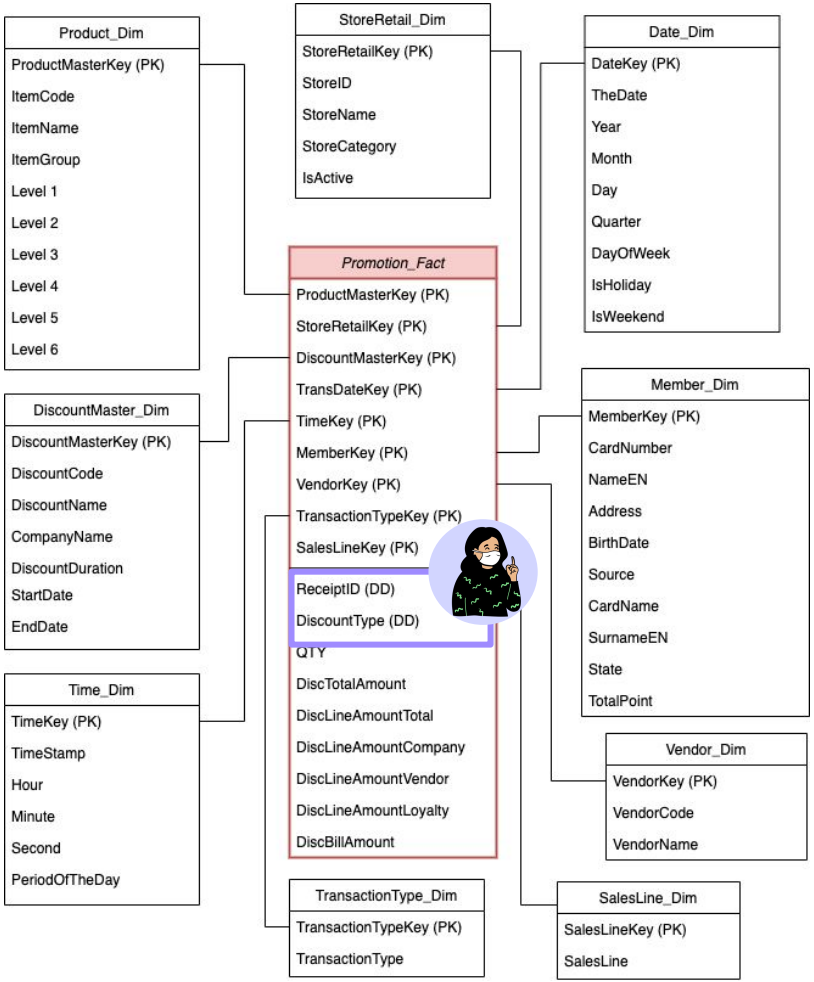


Promotion Fact

2 Techniques

- **Transaction Type Fact** (Promotion_Fact, TransactionType_Dim)

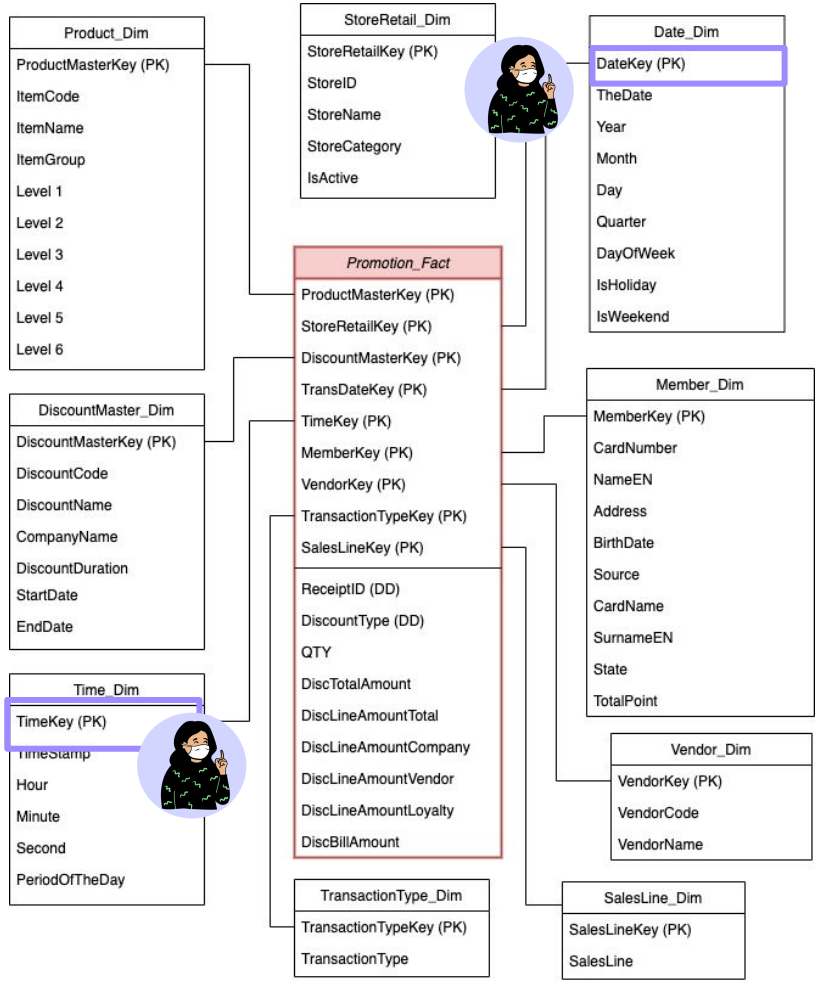
Type	Condition
Buy	(**Qty**) > 0, (**NetSalesAmount**) >0 , (**Price**) > 0
Full Discount	(**SalesAmount**) = (**DiscTotalAmount**), (**Price**) > 0
Giveaway	(**NetSalesAmount**) = 0, (**Price**) = 0, **Cost** = 0
Defected	**NetSalesAmount**<0, **Cost**>0
Return	Have (**ReturnTransactionId**)
Error	Other



Promotion Fact

2 Techniques

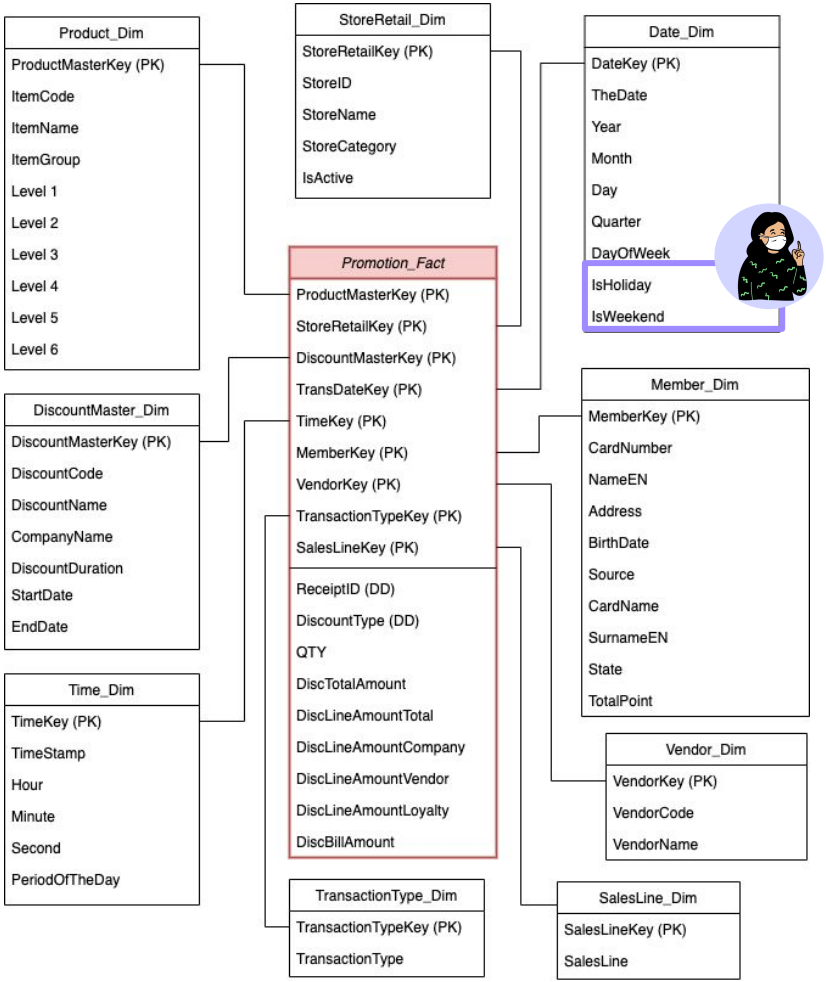
- **Transaction Type Fact** (Promotion_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID, DiscountType)



Promotion Fact

2 Techniques

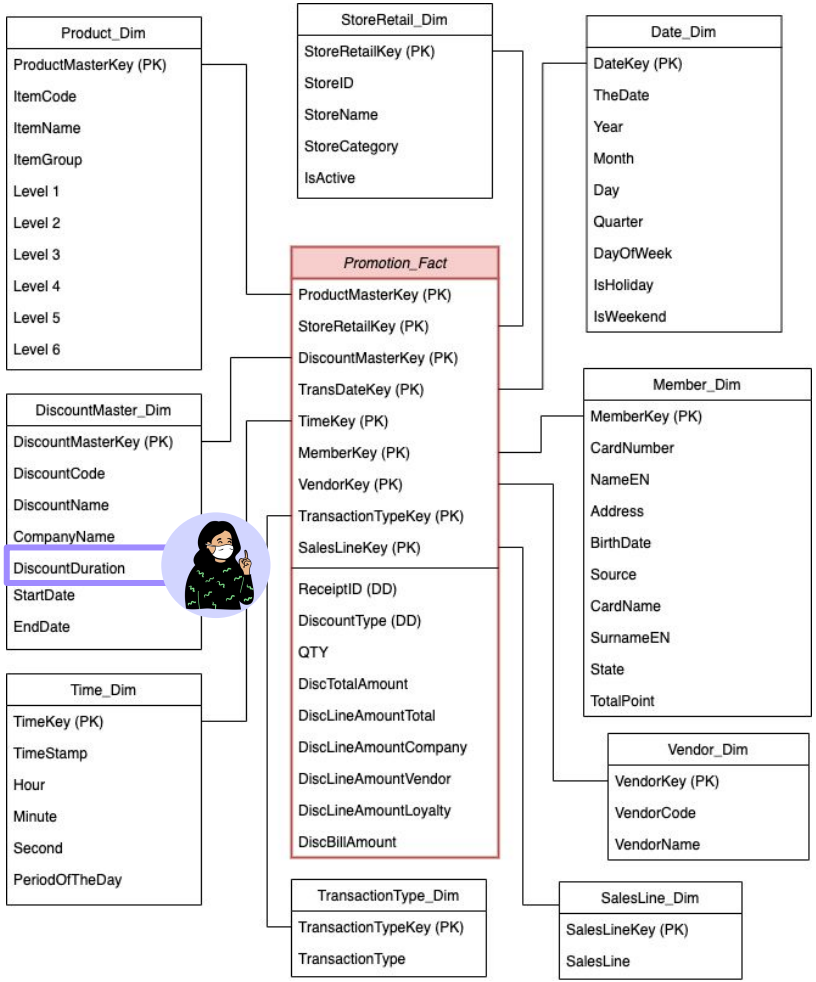
- **Transaction Type Fact** (Promotion_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID, DiscountType)
- **Smart Key** (Date_Dim, Time_Dim)



Promotion Fact

2 Techniques

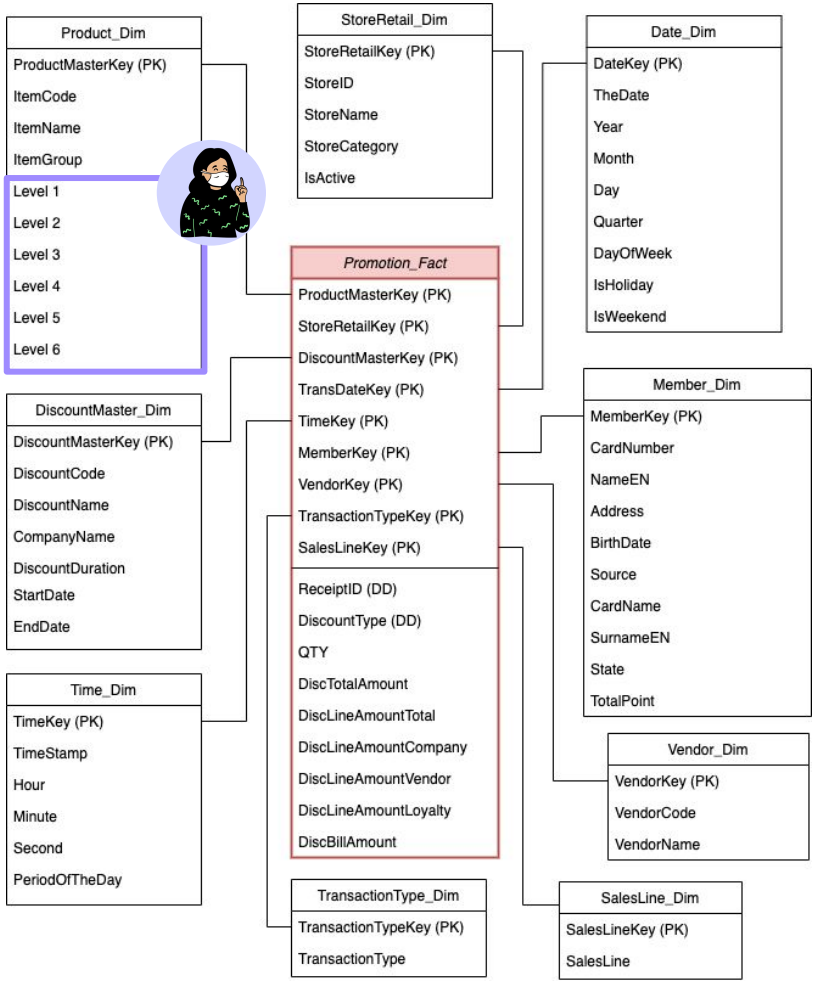
- **Transaction Type Fact** (Promotion_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID, DiscountType)
- **Smart Key** (Date_Dim, Time_Dim)
- **Textual Attributes** (Date_Dim: isHoliday, isWeekend)



Promotion Fact

2 Techniques

- **Transaction Type Fact** (Promotion_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID, DiscountType)
- **Smart Key** (Date_Dim, Time_Dim)
- **Textual Attributes** (Date_Dim: isHoliday, isWeekend)
- **Filtering/Grouping** (DiscountMaster_Dim: *DiscountDuration*)

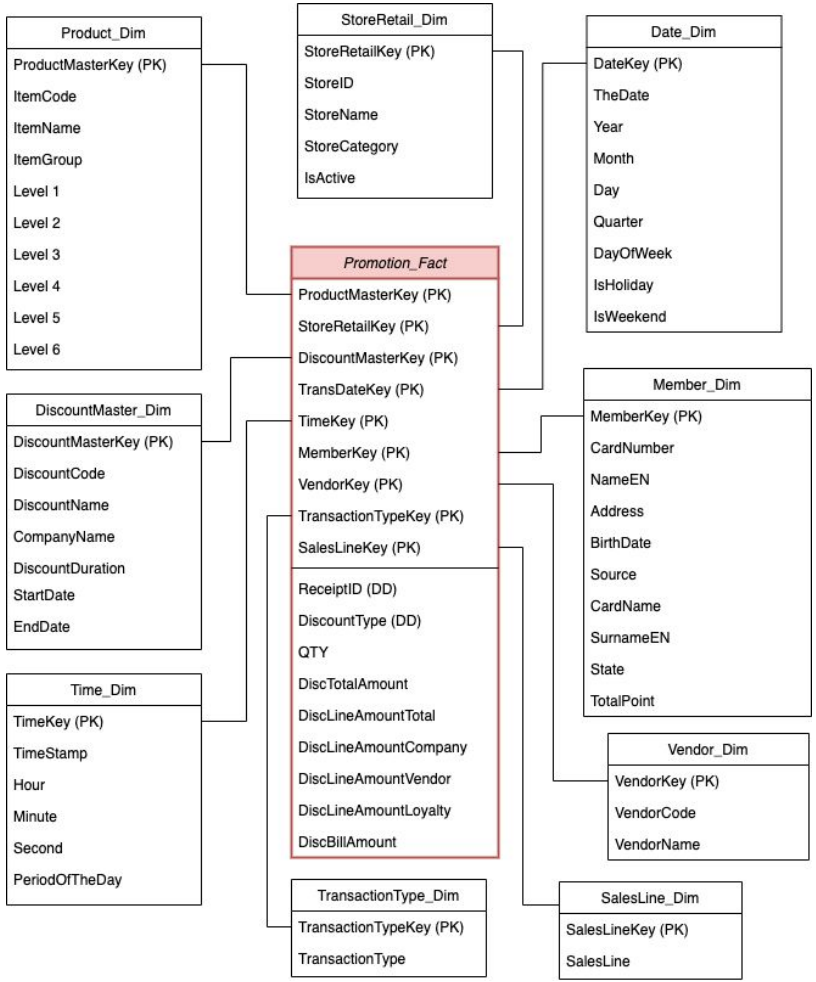


Promotion Fact

2 Techniques

- **Transaction Type Fact** (Promotion_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID, DiscountType)
- **Smart Key** (Date_Dim, Time_Dim)
- **Textual Attributes** (Date_Dim: isHoliday, isWeekend)
- **Filtering/Grouping** (DiscountMaster_Dim: *DiscountDuration*)
- **Hierarchy** (ProductMaster_Dim)

ProductMasterKey	ItemCode	ItemName	ItemGroupName	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
1	1000205	CAMPUS NOTEBOOK@KO900000005756UYO : 165	Not Specified	ALL	LIFESTYLES	FASHION	BAGS	WALLETS & SMALL GOODS	WALLETS & SMALL GOODS
2	1000213	Clear Book A4@Kokuyo : 95	Not Specified	ALL	LIFESTYLES	STATIONERY	PAPER	BOOK/NOTE BOOK	BOOK/NOTE BOOK



Promotion Fact

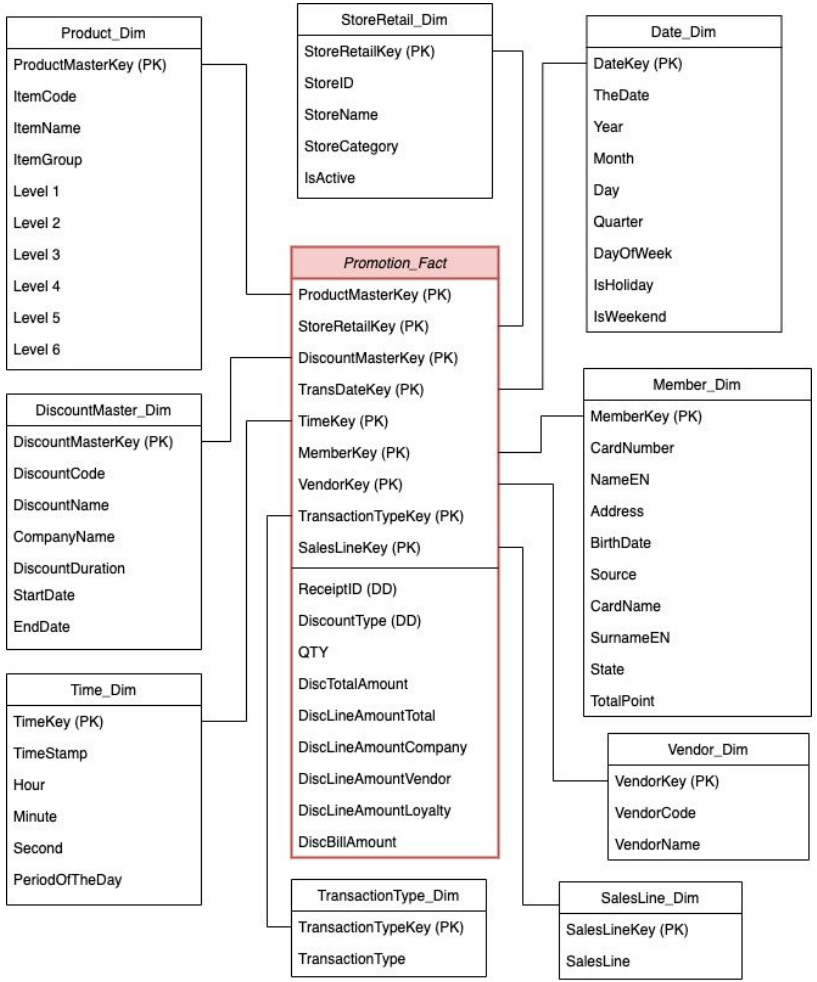
2 Techniques

- **Transaction Type Fact** (Promotion_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID, DiscountType)
- **Smart Key** (Date_Dim, Time_Dim)
- **Textual Attributes** (Date_Dim: isHoliday, isWeekend)
- **Filtering/Grouping** (DiscountMaster_Dim: *DiscountDuration*)
- **Hierarchy** (ProductMaster_Dim)

3 Slowly Changing Dimension

- *DiscountMaster_Dim*
  - ***SCD1***: *Overwrite DiscountName*
  - ***SDC2***: *Track of the Promotion End Date Extension*
  - ***SCD2***: *Track the Relaunch of Promotion*

DiscountMasterKey	DiscountCode	DiscountName	CompanyName	DiscountDuration	StartDate	EndDate	Status
18	CP200000008	PM Yellow Bazaar 250 THB	PINK POMELO,Ltd.	20	2017-06-20	2017-07-09	Current



Promotion Fact

2 Techniques

- **Transaction Type Fact** (Promotion_Fact, TransactionType_Dim)
- **Degenerated Dimension** (ReceiptID, DiscountType)
- **Smart Key** (Date_Dim, Time_Dim)
- **Textual Attributes** (Date_Dim: isHoliday, isWeekend)
- **Filtering/Grouping** (DiscountMaster_Dim: *DiscountDuration*)
- **Hierarchy** (ProductMaster_Dim)

3 Slowly Changing Dimension

- *DiscountMaster_Dim*
  - ***SCD1***: *Track of DiscountName*
  - ***SDC2***: *Track of the Promotion End Date Extension*
  - ***SCD2***: *Track the Relaunch of Promotion*

DiscountMasterKey	DiscountCode	DiscountName	CompanyName	DiscountDuration	StartDate	EndDate	Status
18	CP200000008	PM Yellow Bazaar 250 THB	PINK POMELO,Ltd.	20	2017-06-20	2017-07-09	Current

4 Remark

- **DiscountMasterKey** = 0 → *no discount*
- **MemberKey** = 0 → *customer is not member*

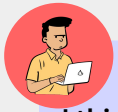
Cube

Cube

Cube should be for large data - to save time!

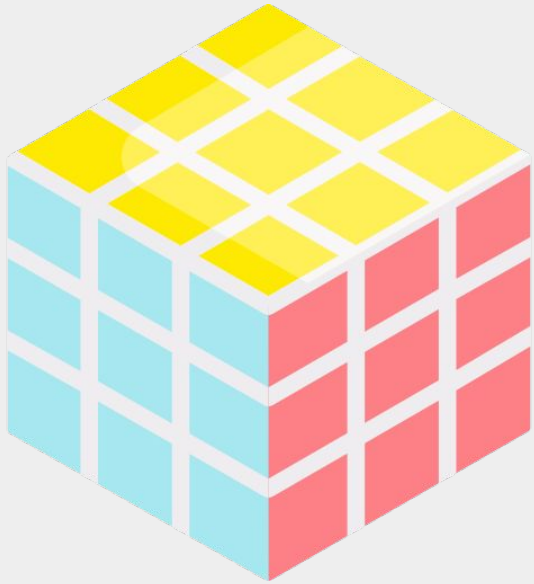
Promotion fact is pretty big...

Which aspect will the CEO want to see frequently?



I think product sale is an important point in business.

Use cube to understand data that is too complex

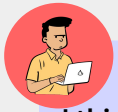


Cube

Cube should be for large data - to save time!

Promotion fact is pretty big...

Which aspect will the CEO want to see frequently?



I think product sale is an important point in business.

Use cube to understand data that is too complex

Product Cube

Promotion Cube

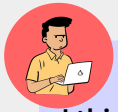
Store Cube

Cube

Cube should be for large data - to save time!

Promotion fact is pretty big...

Which aspect will the CEO want to see frequently?



I think product sale is an important point in business.

Use cube to understand data that is too complex

Product Cube

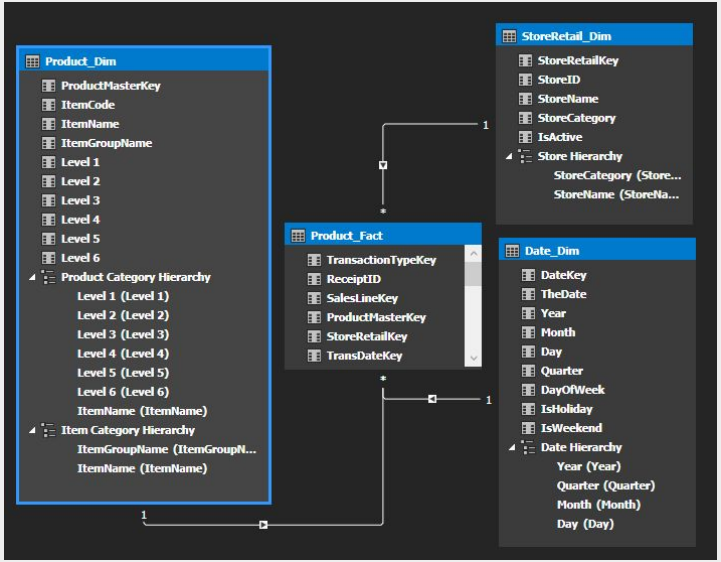
Promotion Cube

Store Cube

Product Cube

1 Why we chose this cube?

- To determine the net profit by Date, Time, Product, and Store
- To determine the total quantity sold by Date, Time, Product, and Store
- To determine the seasonal trends of store retail transactions



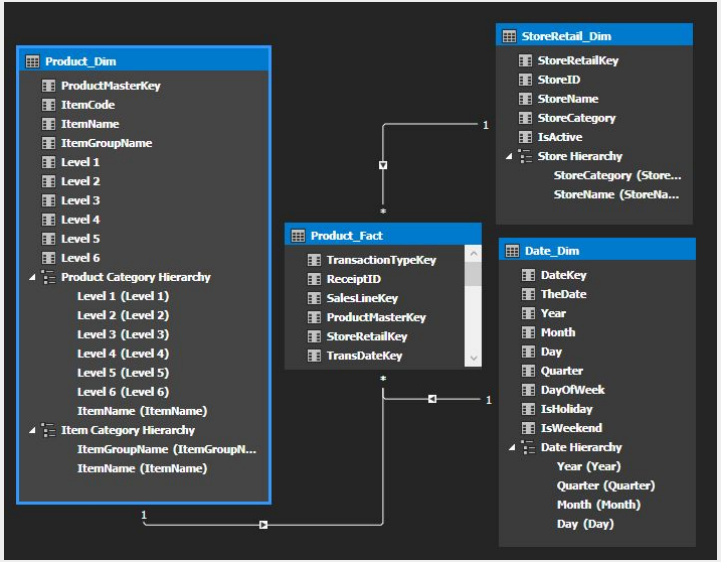
Product Cube

1 Why we chose this cube?

- To determine the net profit by Date, Time, Product, and Store
- To determine the total quantity sold by Date, Time, Product, and Store
- To determine the seasonal trends of store retail transactions

2 Components

- StoreRetail_Dim
- Date_Dim
- Product_Dim



Product Cube

1 Why we chose this cube?

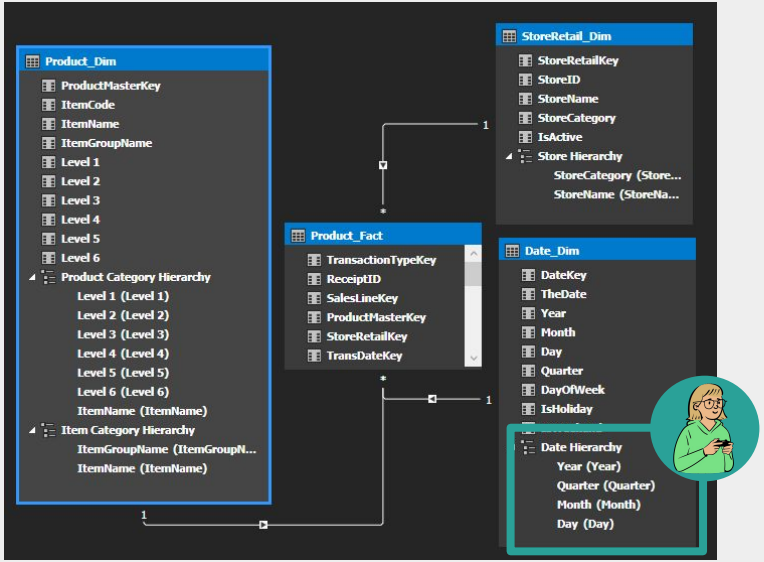
- To determine the net profit by Date, Time, Product, and Store
- To determine the total quantity sold by Date, Time, Product, and Store
- To determine the seasonal trends of store retail transactions

2 Components

- StoreRetail_Dim
- Date_Dim
- Product_Dim

3 Hierarchy

- Date Hierarchy



Product Cube

1 Why we chose this cube?

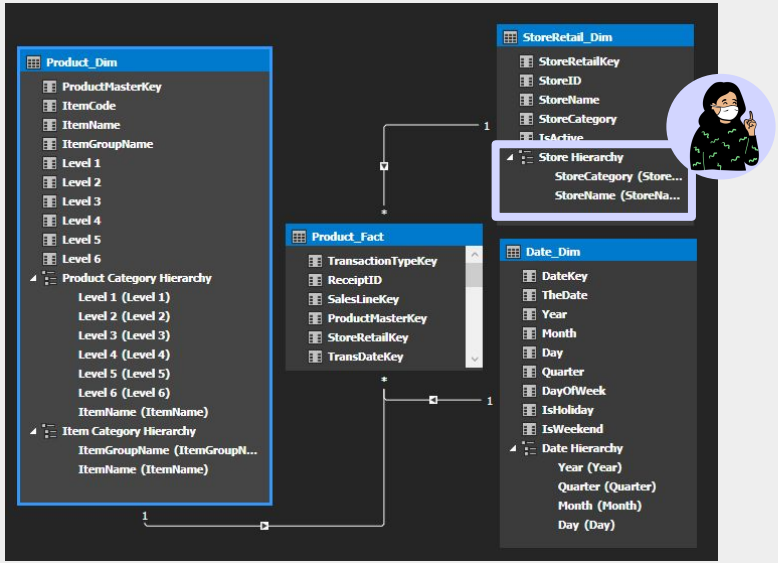
- To determine the net profit by Date, Time, Product, and Store
- To determine the total quantity sold by Date, Time, Product, and Store
- To determine the seasonal trends of store retail transactions

2 Components

- StoreRetail_Dim
- Date_Dim
- Product_Dim

3 Hierarchy

- Date Hierarchy
- Store Hierarchy



Product Cube

1 Why we chose this cube?

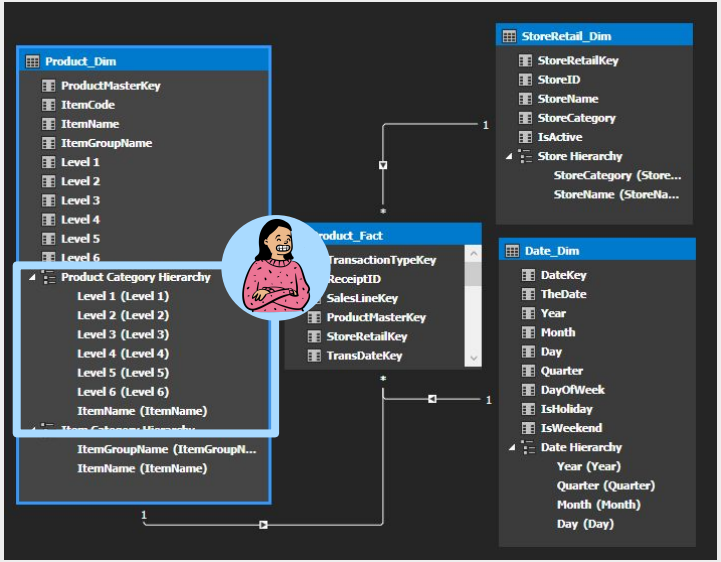
- To determine the net profit by Date, Time, Product, and Store
- To determine the total quantity sold by Date, Time, Product, and Store
- To determine the seasonal trends of store retail transactions

2 Components

- StoreRetail_Dim
- Date_Dim
- Product_Dim

3 Hierarchy

- Date Hierarchy
- Store Hierarchy
- Product Category Hierarchy



Product Cube

1 Why we chose this cube?

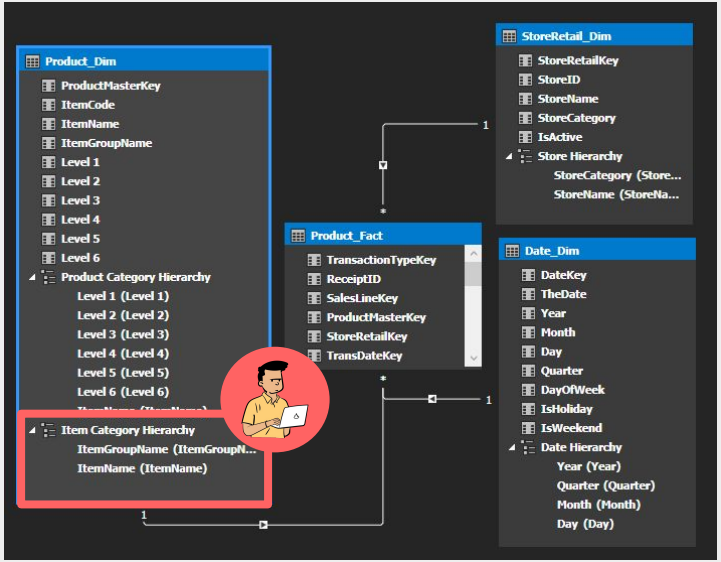
- To determine the net profit by Date, Time, Product, and Store
- To determine the total quantity sold by Date, Time, Product, and Store
- To determine the seasonal trends of store retail transactions

2 Components

- StoreRetail_Dim
- Date_Dim
- Product_Dim

3 Hierarchy

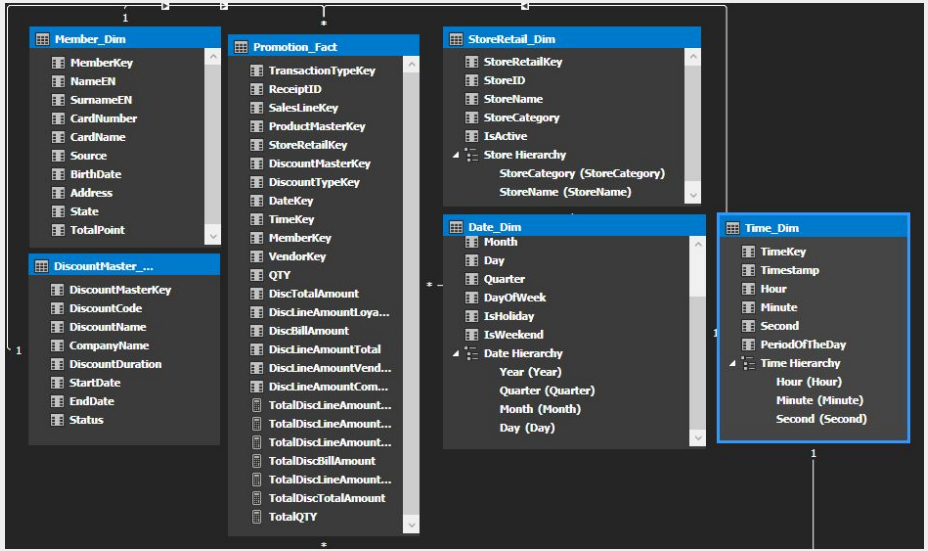
- Date Hierarchy
- Store Hierarchy
- Product Category Hierarchy
- Item Category Hierarchy



Promotion Cube

1 Why we chose this cube?

- To analyse all sorts of information about promotion in retail store
- We can further establish trends and analyse promotion performance



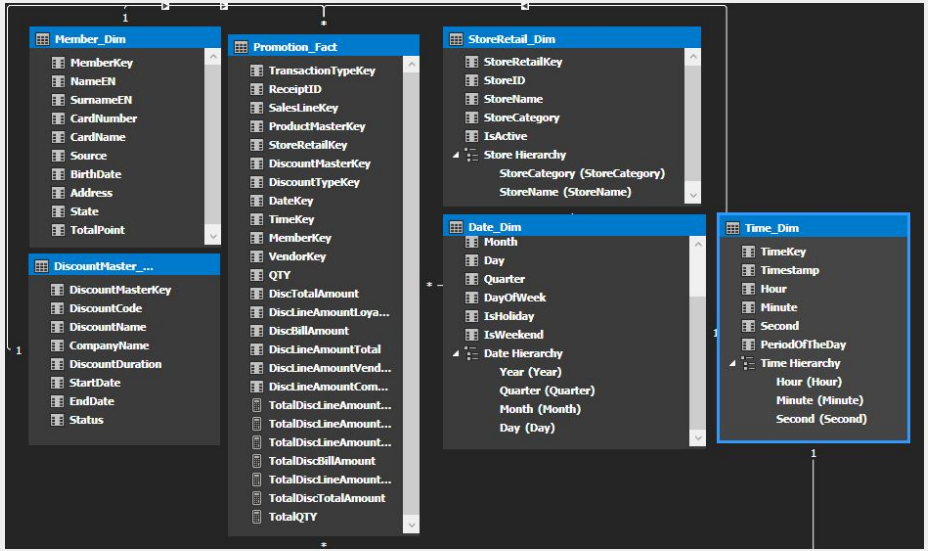
Promotion Cube

1 Why we chose this cube?

- To analyse all sorts of information about promotion in retail store
- We can further establish trends and analyse promotion performance

2 Components

- Member_Dim
- DiscountMaster_Dim
- StoreRetail_Dim
- Date_Dim
- Time_Dim



Promotion Cube

1 Why we chose this cube?

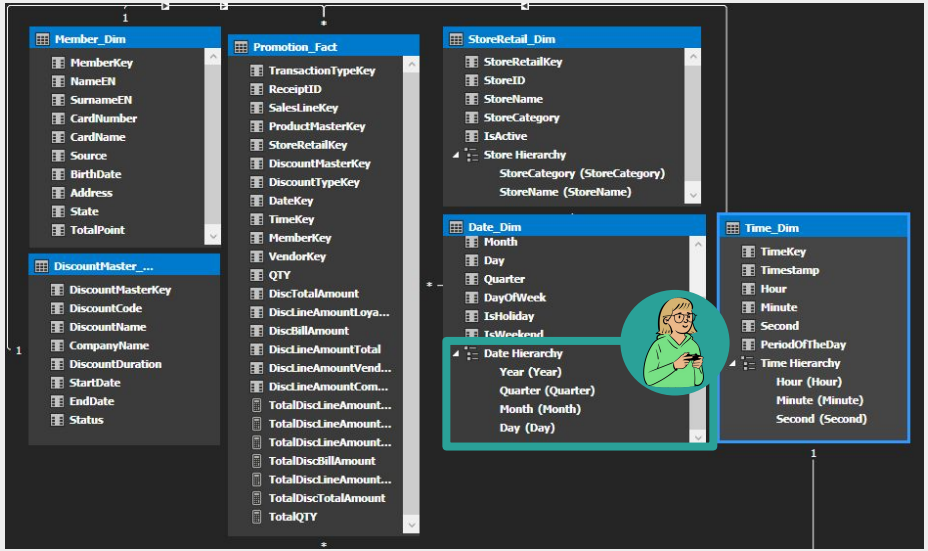
- To analyse all sorts of information about promotion in retail store
- We can further establish trends and analyse promotion performance

2 Components

- Member_Dim
- DiscountMaster_Dim
- StoreRetail_Dim
- Date_Dim
- Time_Dim

3 Hierarchy

- Date Hierarchy



Promotion Cube

1 Why we chose this cube?

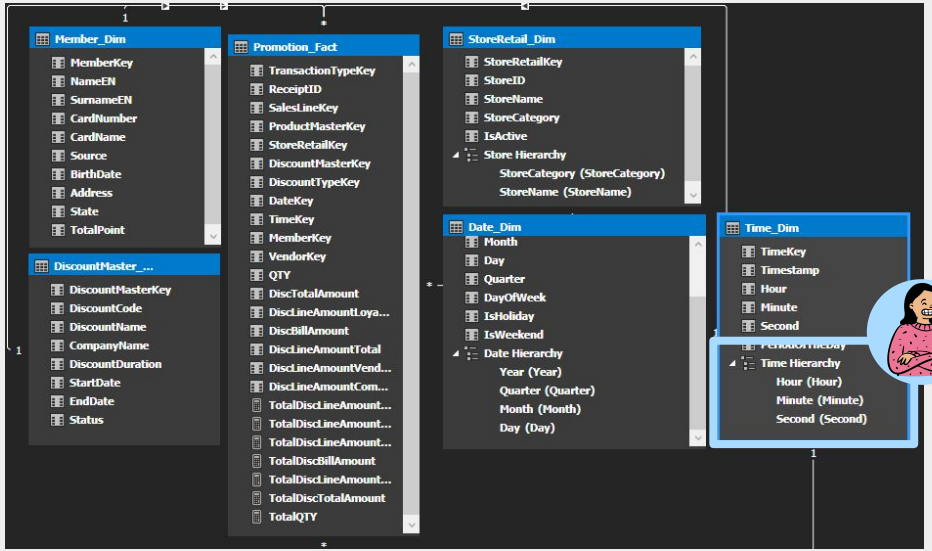
- To analyse all sorts of information about promotion in retail store
- We can further establish trends and analyse promotion performance

2 Components

- Member_Dim
- DiscountMaster_Dim
- StoreRetail_Dim
- Date_Dim
- Time_Dim

3 Hierarchy

- Date Hierarchy
- Time Hierarchy



Promotion Cube

1 Why we chose this cube?

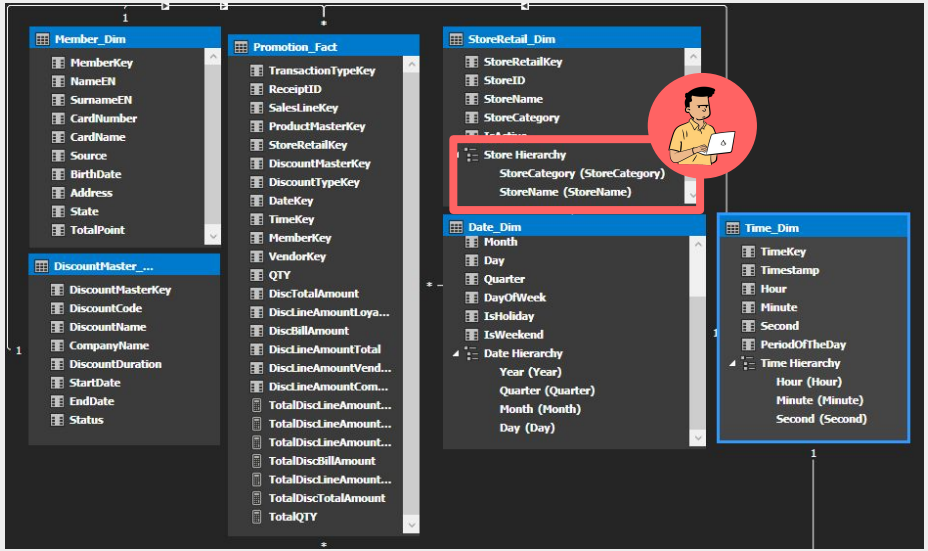
- To analyse all sorts of information about promotion in retail store
- We can further establish trends and analyse promotion performance

2 Components

- Member_Dim
- DiscountMaster_Dim
- StoreRetail_Dim
- Date_Dim
- Time_Dim

3 Hierarchy

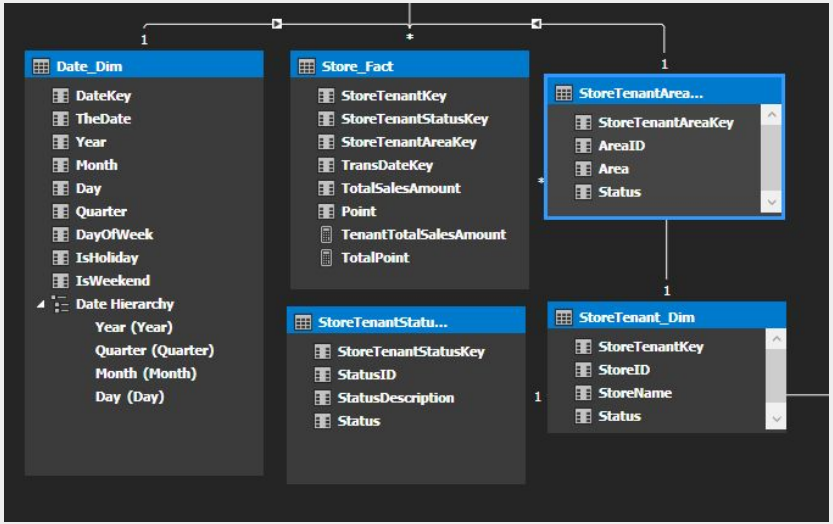
- Date Hierarchy
- Time Hierarchy
- Store Hierarchy



Store Cube

1 Why we chose this cube?

- To emphasize the importance of analysis tenant store as it can have effect with the retail store
- To analyse tenant store for its sales and benefit in using points



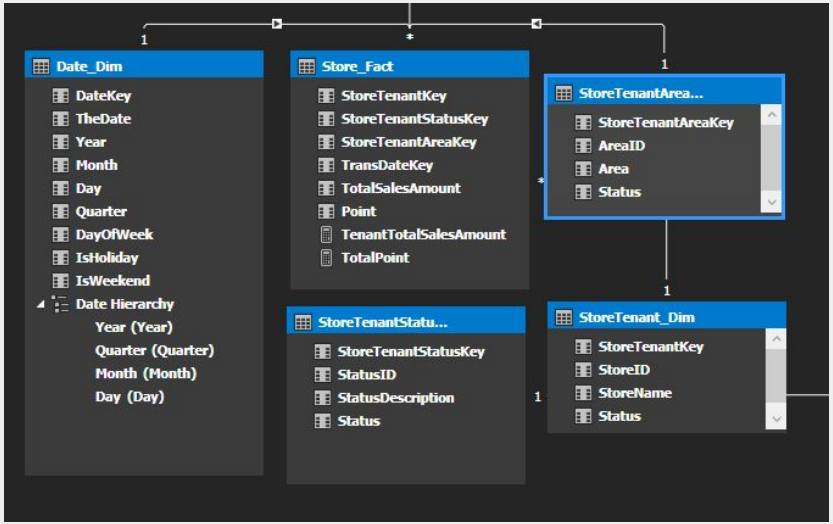
Store Cube

1 Why we chose this cube?

- To emphasize the importance of analysis tenant store as it can have effect with the retail store
- To analyse tenant store for its sales and benefit in using points

2 Components

- StoreTenant_Dim
- StoreTenantArea_Dim
- StoreTenantStatus_Dim
- Date_Dim



Store Cube

1 Why we chose this cube?

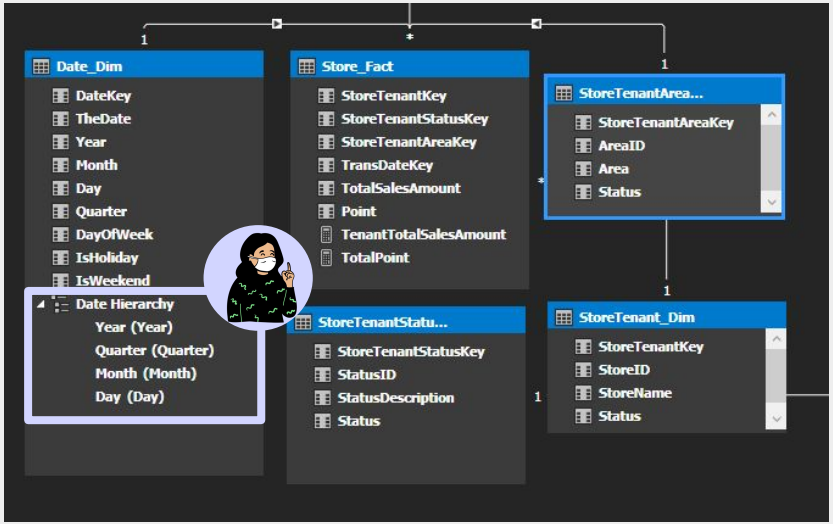
- To emphasize the importance of analysis tenant store as it can have effect with the retail store
- To analyse tenant store for its sales and benefit in using points

2 Components

- StoreTenant_Dim
- StoreTenantArea_Dim
- StoreTenantStatus_Dim
- Date_Dim

3 Hierarchy

- Date Hierarchy



We're DONE!

Thank you for listening.

1. Data Inspect - Each Table in the source
    a. What we have cleaned
    b. What we have mocked
    c. What we choose to analyse
2. 4-Step Design
    a. Bus
    b. Each fact → dim
3. ~~Design Technique~~
4. SCD
5. Cube

1. Data Inspect - Each Table in the source

1. ~~SALES_TRANSACTION → Add TransactionType (data error), Brand no mock~~
2. SALES_PAYMENT_TRANSACTION
3. ~~SALES_DISCOUNT_TRANSACTION~~
4. ~~SALES_TRANSACTION_SUMMARY~~
5. ~~MEMBER_PROFILE → null to nonspecific value, modify synonym data to one specific value (TH, THA → TH)~~
6. ~~CARD_DESC~~
7. ~~SHOP_TENANT~~
8. ~~SHOP_RETAIL - add isActive~~
9. ~~DISCOUNT_MASTER~~
10. ~~PRODUCT_MASTER~~
11. ~~PRODUCT_HIERARCHY~~

~~Bus Matrix~~

- *Member_Fact* **1 row/ 1 member receipt /payment type**
  - Account_Dim
  - Account-To-Member Bridge
  - Company_Dim
  - Member_Dim
  - MemberDemographic_Dim
  - PaymentType_Dim
  - Date_Dim
  - Time_Dim
  - StoreDim
- *Store_Fact* **1 row/ 1 summary store/ day**
  - StoreTenant_Dim
  - StoreTenantArea_Dim
  - StoreTenantStatus_Dim
  - Date_Dim
- *Product_Fact* **1 row/ 1 item in 1 receipt**
  - Product_Dim
  - Date_Dim
  - Time_Dim
  - SalesLine_Dim
  - Vendor_Dim
  - StoreRetail_Dim
  - TransactionType_Dim
- *AggregateProduct_Fact* **1 row/ 1 product sale in each store / month**
  - Month_Dim
  - Vendor_Dim
  - Product_Dim
  - StoreRetail_Dim
- *Promotion_Fact* **1 row/ 1 item on receipt**
  - DiscountMaster_Dim
  - Date_Dim
  - Time_Dim
  - Product_Dim
  - SalesLine_Dim
  - Venndor_Dim
  - Member_Dim
  - StoreRetail_Dim
  - TransactionType_Dim

- Bridge
- TransactionType
- Date, Time
  - Smart Key
- Ammy Store
- DD