Dimensional Modelling

Dimensional Modeling

✓Method of organizing data (in a data warehouse)

✓Fact

- Measurements like profit

✓Dimensions

- Context like category or time period

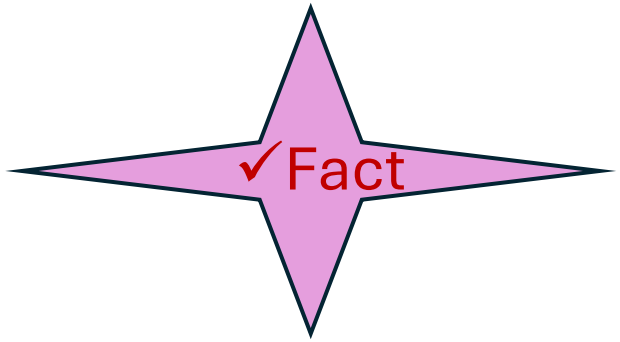
Profit by year

Profit by Category

Dimensional Modeling

✓Dimensions

✓Dimensions



✓Dimensions

✓Dimensions

Star Schema

Dimensional Modeling

Unique technique of structuring data

Commonly used in DWH

Optimized for faster data retrieval

Oriented around performance & usability

Designed for reporting/OLAP

Why Dimensional Modelling?

Need of Dimensional modelling

Unique technique of structuring data

Optimized for faster data retrieval

Oriented around performance & usability

id	date	product	category	Customer_id	name	profit
1	01-01-2025	Tomato	vegetables	2	Anil	10
2	01-01-2025	Chilly	vegetables	2	Anil	20
3	01-01-2025	Bhendi	vegetables	1	Lokesh	40
4	01-01-2025	Vapours	snacks	5	Sachin	17
5	01-01-2025	Farsan	snacks	3	Ajit	54
6	01-01-2025	Masala powder	Herbs	4	Maruti	87

Need of Dimensional modelling

Unique technique of structuring data

Optimized for faster data retrieval

Oriented around performance & usability

id	date	product	category	Customer_id	name	profit
1	01-01-2025	Tomato	vegetables	2	Anil	10
2	01-01-2025	Chilly	vegetables	2	Anil	20
3	01-01-2025	Bhendi	vegetables	1	Lokesh	40
4	01-01-2025	Vapours	snacks	5	Sachin	17
5	01-01-2025	Farsan	snacks	3	Ajit	54
6	01-01-2025	Masala powder	Herbs	4	Maruti	87

FK

id	date	product	category	Customer_id	profit
1	01-01-2025	Tomato	vegetables	2	10
2	01-01-2025	Chilly	vegetables	2	20
3	01-01-2025	Bhendi	vegetables	1	40
4	01-01-2025	Vapours	snacks	5	17
5	01-01-2025	Farsan	snacks	3	54
6	01-01-2025	Masala powder	Herbs	4	87

Profit Fact

PK

Customer_id	Customer Name	Address	Contact
1	Lokesh	Pune	32490349
2	Anil	Mumbai	43390349
3	Suresh	Nashik	54290349
4	Maruti	Nagpur	65190349
5	Sachin	Kolhapur	76090349
6	Rakesh	Nanded	86990349

Customer Dim

Need of Dimensional modelling

Unique technique of structuring data

Optimized for faster data retrieval

Oriented around performance & usability

Product_id	date	product	category	Customer_id	profit
1	01-01-2025	Tomato	vegetables	2	10
2	01-01-2025	Chilly	vegetables	2	20
3	01-01-2025	Bhendi	vegetables	1	40
4	01-01-2025	Vapours	snacks	5	17
5	01-01-2025	Farsan	snacks	3	54
6	01-01-2025	Masala powder	Herbs	4	87

FK

id	date	product_id	Customer_id	profit
1	01-01-2025	2	2	10
2	01-01-2025	5	2	20
3	01-01-2025	16	1	40
4	01-01-2025	23	5	17
5	01-01-2025	3	3	54
6	01-01-2025	1	4	87

Profit Fact

PK

product_id	Product	Category
1	product_1	Vegetables
2	product_2	snacks
3	product_3	Herbs
4	product_4	snacks
5	product_5	snacks
6	product_6	Herbs

Product Dim

Need of Dimensional modelling

Unique technique of structuring data

Optimized for faster data retrieval

Oriented around performance & usability

id	Date_id	product_id	Customer_id	profit
1	01-01-2025	2	2	10
2	01-01-2025	5	2	20
3	01-01-2025	16	1	40
4	01-01-2025	23	5	17
5	01-01-2025	3	3	54
6	01-01-2025	1	4	87

FK

id	date_id	product_id	Customer_id	profit
1	20250101	2	2	10
2	20250101	5	2	20
3	20250101	16	1	40
4	20250101	23	5	17
5	20250101	3	3	54
6	20250101	1	4	87

Profit Fact

PK

date_id	weekday	month
20220101	Monday	January
20220102	Tuesday	January
20220103	Wednesday	January
20220104	Thursday	January
20220105	Friday	January
20220106	Saturday	January

Date Dim

Need of Dimensional modelling

id	date_id	product_id	Customer_id	profit
1	20250101	2	2	10
2	20250101	5	2	20
3	20250101	16	1	40
4	20250101	23	5	17
5	20250101	3	3	54
6	20250101	1	4	87

Profit Fact

date_id	weekday	month
20220101	Monday	January
20220102	Tuesday	January
20220103	Wednesday	January
20220104	Thursday	January
20220105	Friday	January
20220106	Saturday	January

Date Dim

Customer_id	Customer Name	Address	Contact
1	Lokesh	Pune	32490349
2	Anil	Mumbai	43390349
3	Suresh	Nashik	54290349
4	Maruti	Nagpur	65190349
5	Sachin	Kolhapur	76090349
6	Rakesh	Nanded	86990349

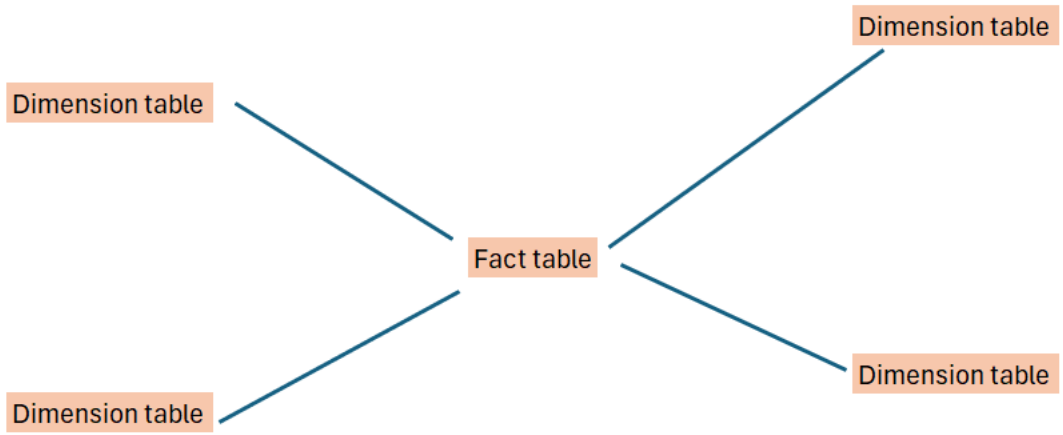
Customer Dim

product_id	Product	Category
1	product_1	Vegetables
2	product_2	snacks
3	product_3	Herbs
4	product_4	snacks
5	product_5	snacks
6	product_6	Herbs

Product Dim

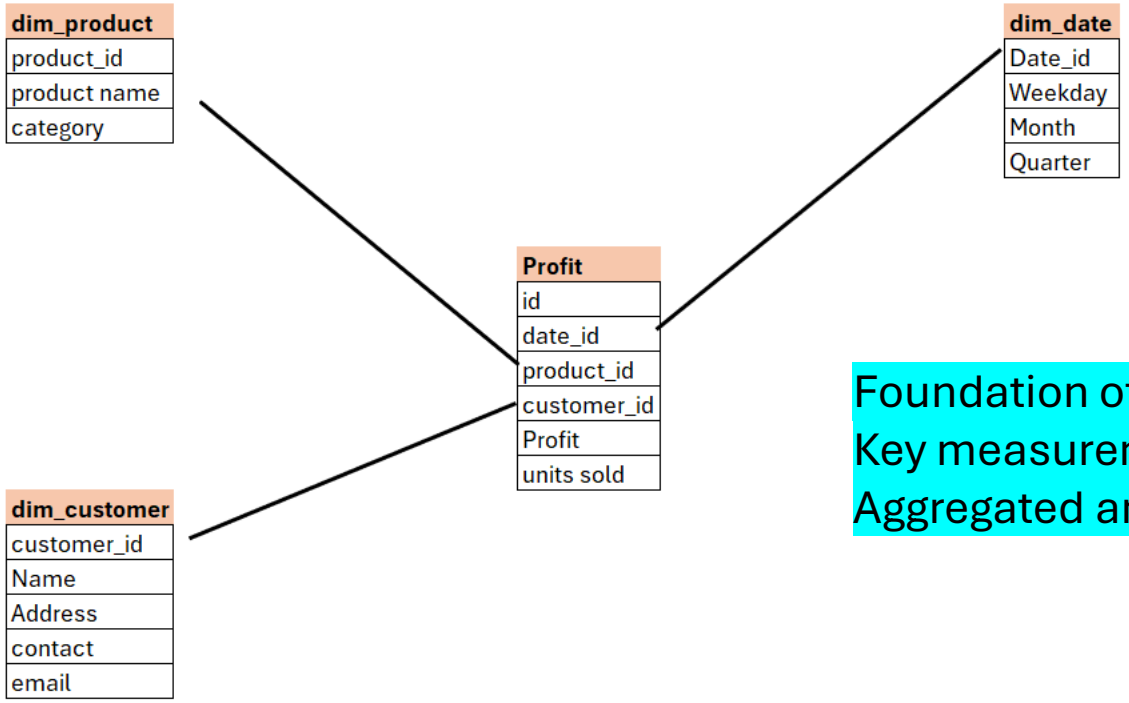
Facts

Facts – Star Schema



Star Schema

Facts



Foundation of DWH
Key measurements
Aggregated and analyzed

Aggregatable(numeric values)
Measurable vs descriptive
Event or transactional data
Date/time fact table

Star Schema

Facts Example

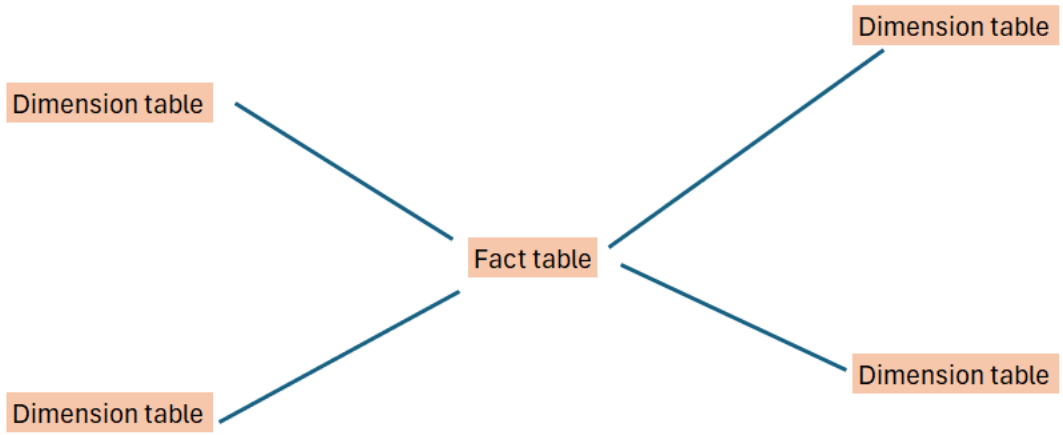
id	date_id	product_id	Customer_id	profit
1	20250101	2	2	10
2	20250101	5	2	20
3	20250101	16	1	40
4	20250101	23	5	17
5	20250101	3	3	54
6	20250101	1	4	87

Fact table

- PK, FK & Facts
- grain – most atomic facts
- Different types of facts

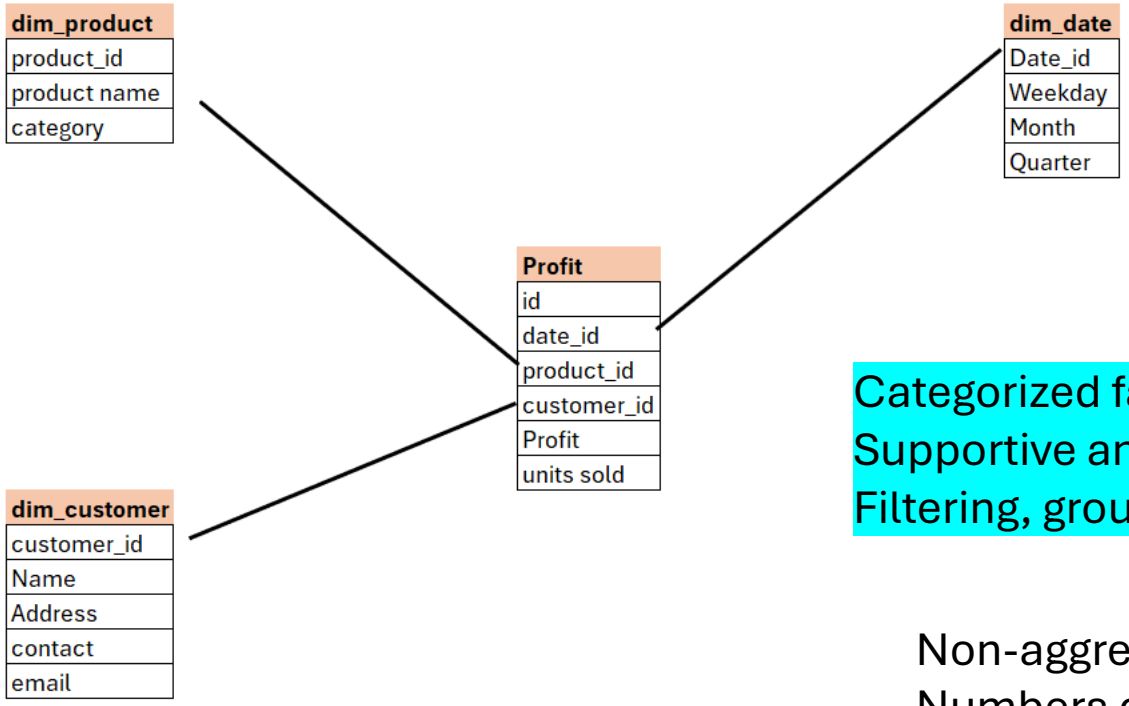
Dimensions

Dimensions – Star Schema



Star Schema

Dimensions



Categorized facts
Supportive and descriptive
Filtering, grouping & labelling

Non-aggregatable
Numbers can be dimension but not aggregatable
static

Star Schema

Dimension Example

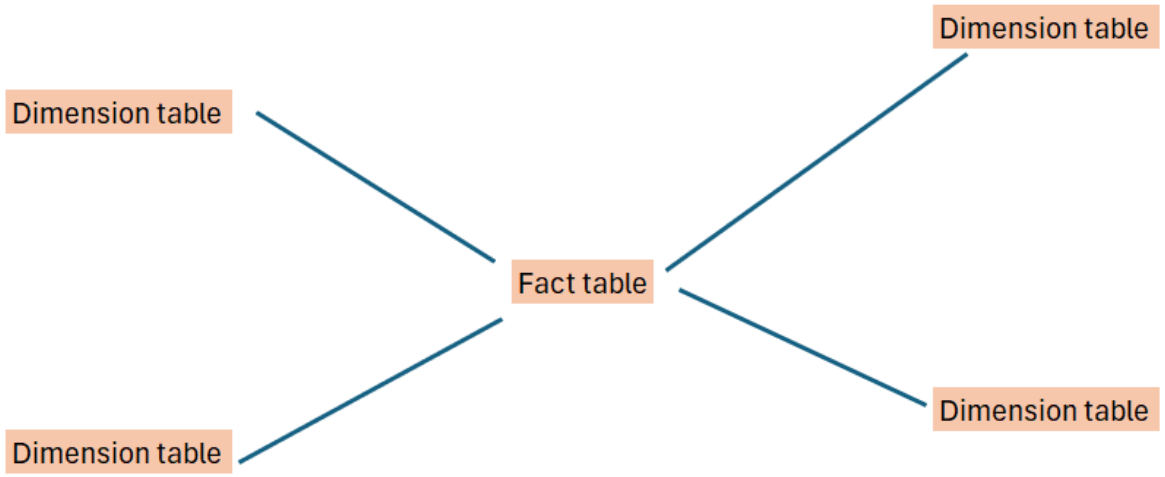
Customer Name	Address	Contact
Lokesh	Pune	32490349
Anil	Mumbai	43390349
Suresh	Nashik	54290349
Maruti	Nagpur	65190349
Sachin	Kolhapur	76090349
Rakesh	Nanded	86990349

Fact table

- PK, dimension, (FK)
- People, products, places, time
- Different types of dimensions

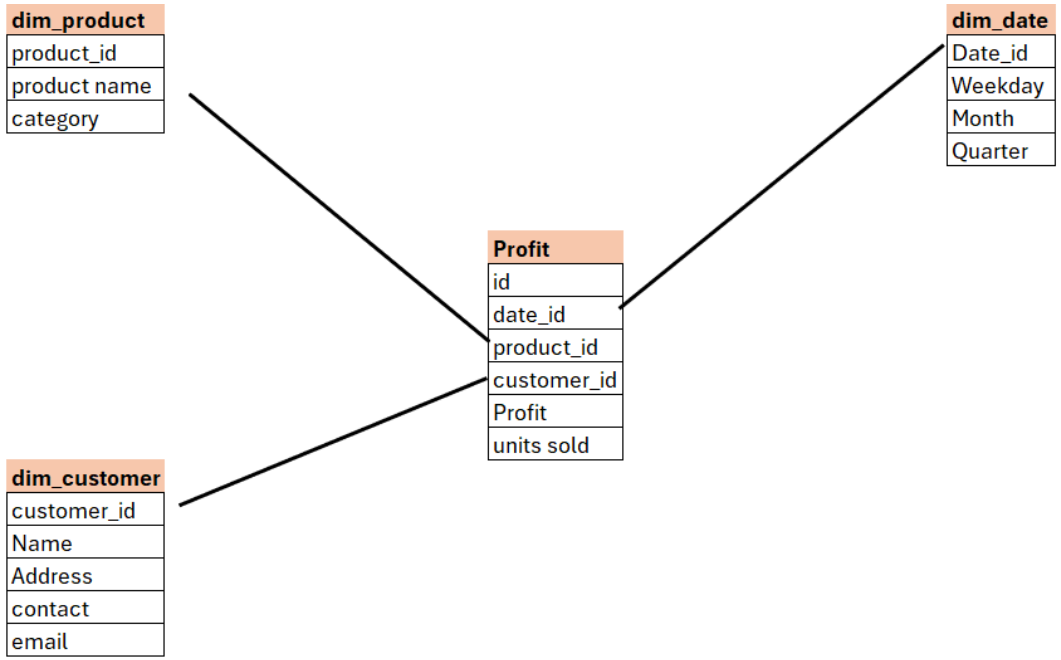
Star Schema

Star Schema



Star Schema

Star Schema example



Star Schema - Example

Normalization

Technique to avoid data redundancy
Minimizes storage
Performance (Write/update)
Many tables
Many joins

FK

id	date_id	product_id	Customer_id	profit
1	20250101	2	2	10
2	20250101	5	2	20
3	20250101	16	1	40
4	20250101	23	5	17
5	20250101	3	3	54
6	20250101	1	4	87

Fact table

1:n

PK

product_id	Product	Category	Sub Category
1	Chilli	Herbs	Spices
2	Garlic	Fruits & Vegetables	Vegetable
3	Banana	Fruits & Vegetables	fruits
4	Chocolate	Sweets& Snacks	Sweets
5	Chips	Sweets& Snacks	Snacks
6	Masala powder	Herbs	Spices

Dimension table

Denormalized

There is data redundancy
Optimized to get data out
Query performance(read)

Star Schema

- Most common schema in Data Mart
- Simplest form
- Work best for specific goals
- Usability + Performance for specific (read) use case

Snowflake Schema

Snowflake Schema - Example

FK

id	date_id	product_id	Customer_id	profit
1	20250101	2	2	10
2	20250101	5	2	20
3	20250101	16	1	40
4	20250101	23	5	17
5	20250101	3	3	54
6	20250101	1	4	87

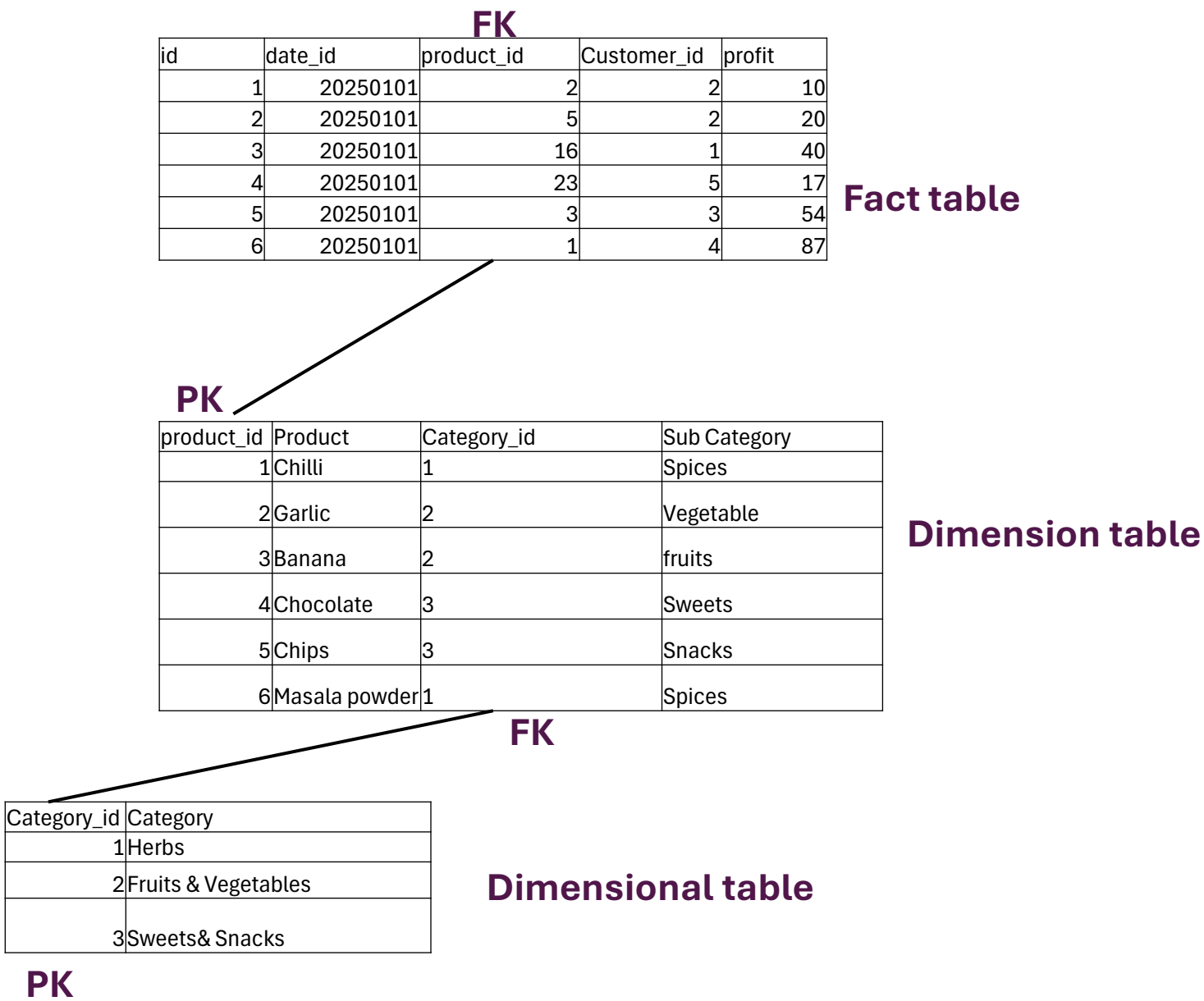
Fact table

1:n

PK

product_id	Product	Category	Sub Category
1	Chilli	Herbs	Spices
2	Garlic	Fruits & Vegetables	Vegetable
3	Banana	Fruits & Vegetables	fruits
4	Chocolate	Sweets& Snacks	Sweets
5	Chips	Sweets& Snacks	Snacks
6	Masala powder	Herbs	Spices

Snowflake Schema - Example



Snowflake Schema

Advantages

Less Space (storage Cost)

No redundant (less) data
- Easier to maintain/update,
- Less risk of corrupted data

Disadvantages

More complex

More joins (more complex queries)

Less performance data marts