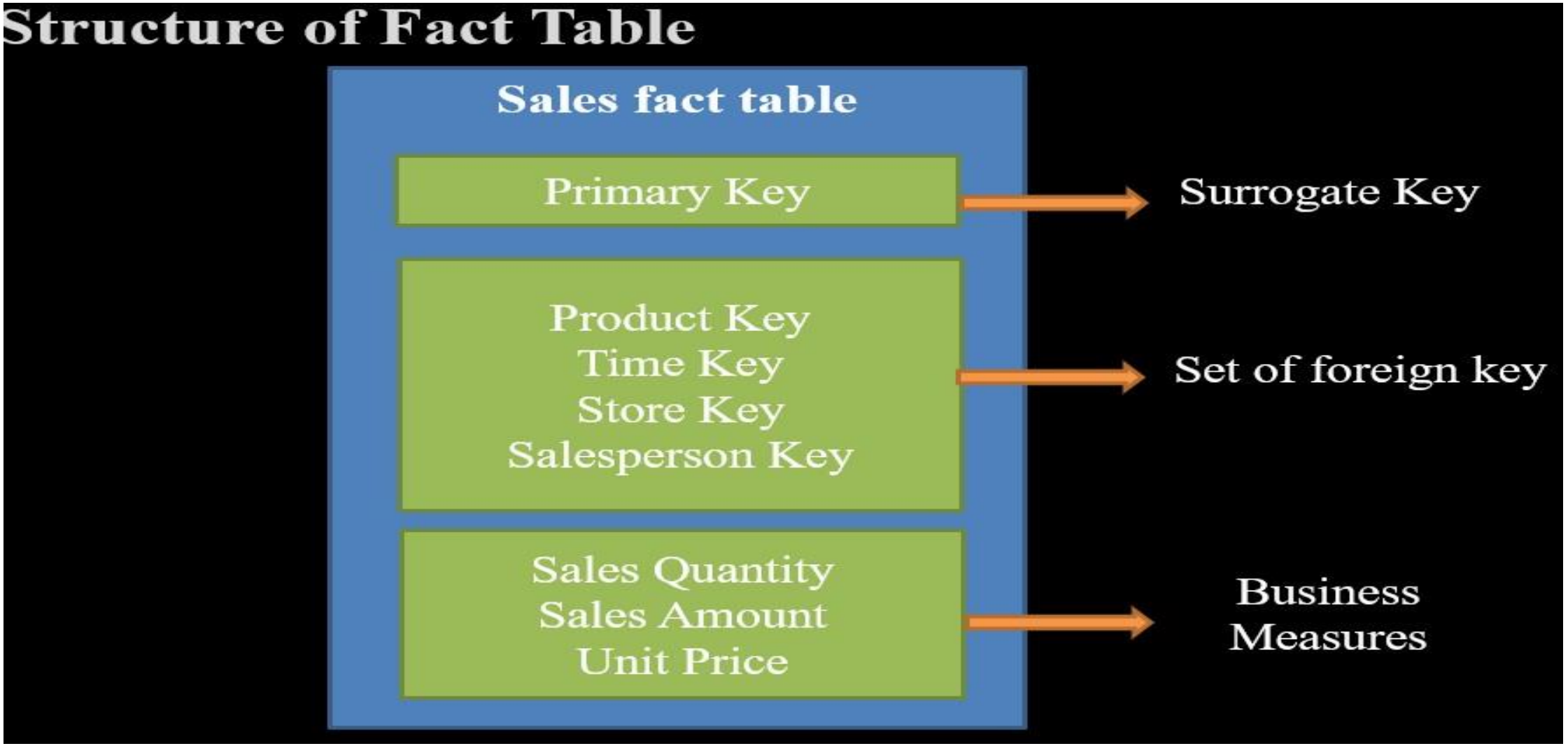
DATA WAREHOUSING AND MINING

T.E. CSE-DS, Sem V
Academic Year: 2023-24

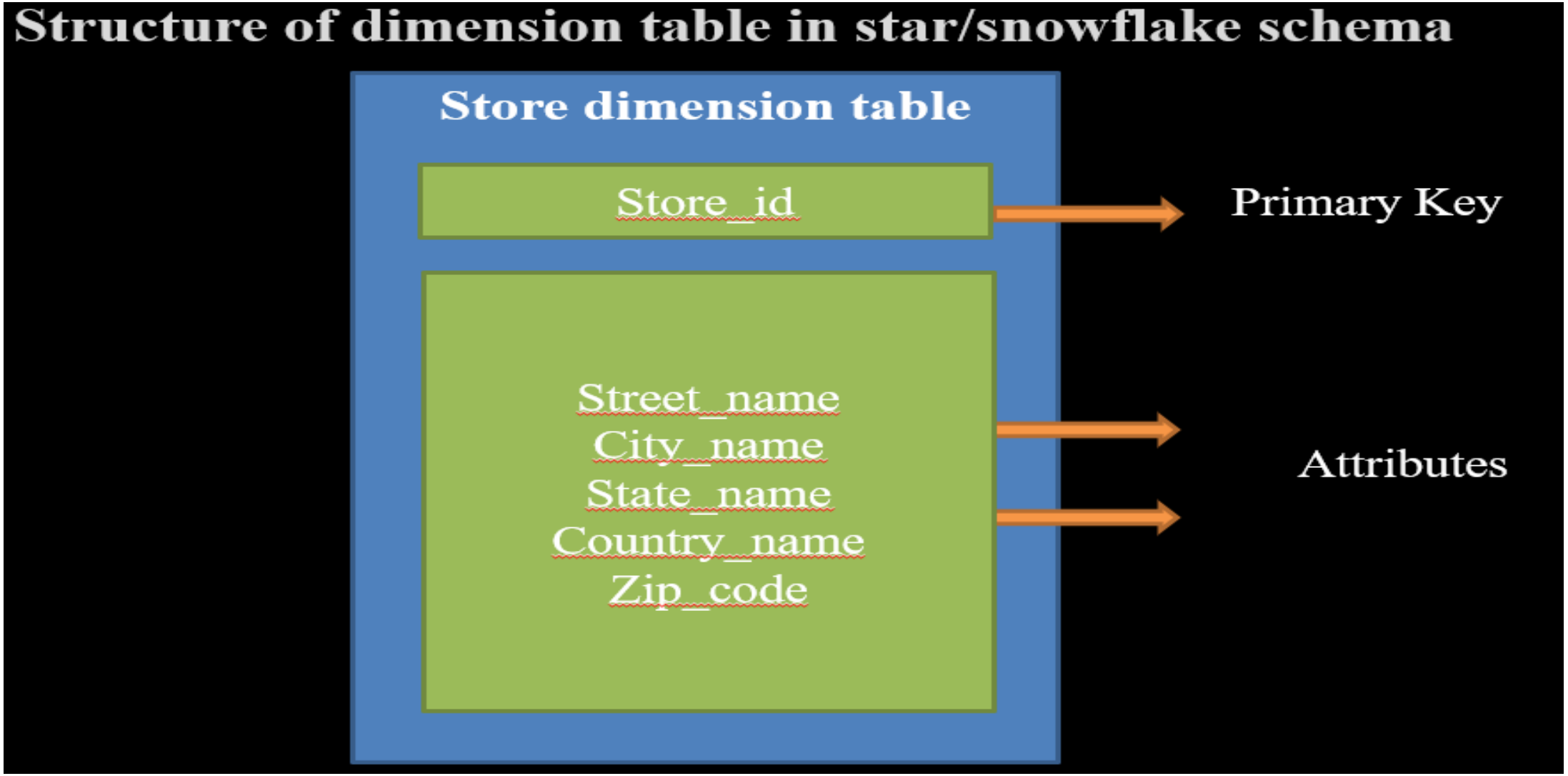
Data Warehousing Fundamentals: ER vs Dimensional Modelling, IPD, Star Schema

Lecture 5

Structure of Fact Table



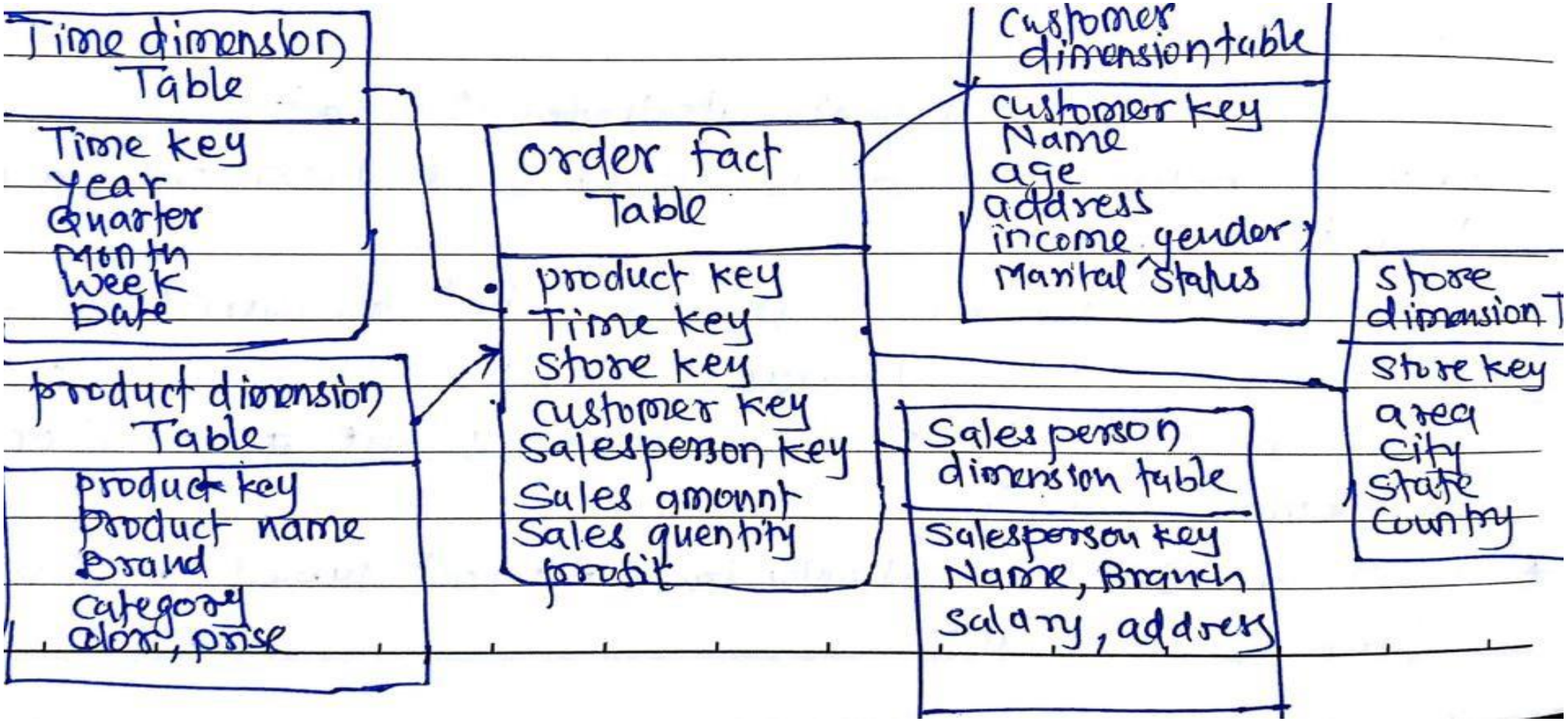
Structure of Dimension Table



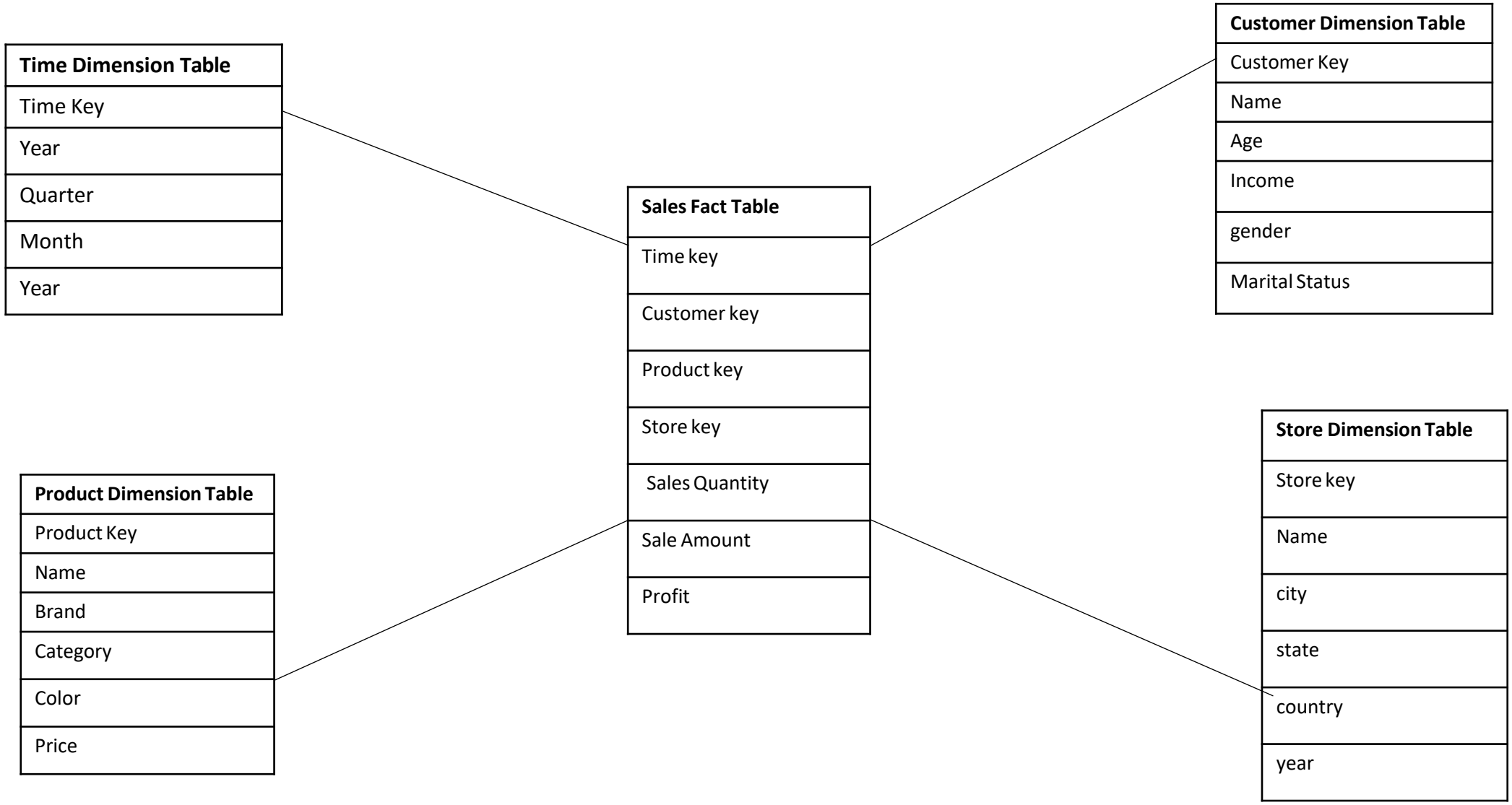
The Star Schema

1. It is called a star schema because its diagram represent a star like structure
2. Every dimension table has a direct relationship with the fact table in the middle thereby allowing every dimension table with its attribute to have an equal chance of participation in a query to analyse the attributes in the fact table
3. The fact tables contain primary information in the DW and the dimension table contain information about the entries for a particular attribute in the fact table
4. Each dimension table is joined to the fact table using a primary key to foreign key join, but the dimension tables are not joined to each other

Star Schema


Time dimension Table
Time key
year
Quarter
month
week
Date
product dimension Table
product key
product name
Brand
category
color, price
order fact Table
product key
Time key
store key
customer key
Salesperson key
Sales amount
Sales quantity
profit
Customer dimension table
customer key
Name
age
address
income gender
Marital status
Salesperson dimension table
Salesperson key
Name, Branch
Salary, address
store dimension T
Store key
area
City
State
Country

Star Schema Generated from IPD – Retail Sales



How does a query execute

1. The star schema is simply a relational model with one to many relationship between fact and dimension tables.
2. It is a denormalized relational model.
3. When a query is executed against the star schema the result of the query is produced by combining or joining one or more dimension tables with the fact table.
4. How much profit in dollars did the salesperson David make on 2nd January 2020 by selling trousers to Jenny at Delhi store?

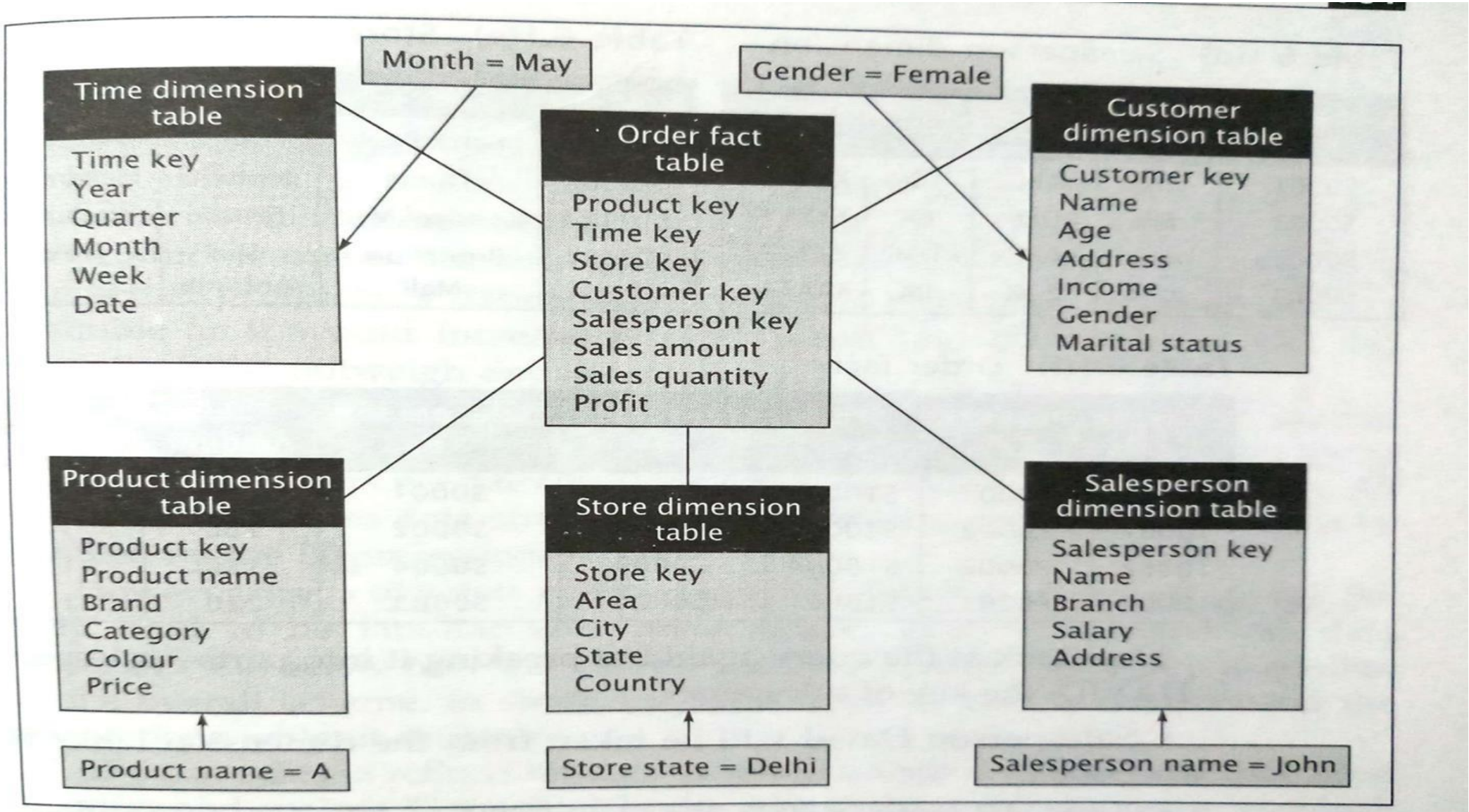


Figure 6.3 Representation of a star schema

Table 6.1(a) Customer dimension

Customer Key	Name	Age	Address	Income	Gender	Marital Status
C0001	Mary	25	XXXXXX	35K	Female	Single
C0002	Joe	32	YYYYYY	40K	Male	Married
C0003	Ken	30	ZZZZZZ	27K	Male	Single
C0004	Jenny	43	AAAAAA	43K	Female	Married

Table 6.1(b) Time dimension

Time Key	Year	Quarter	Month	Week	Date
T0001	2006	First	January	First	01
T0002	2006	First	January	First	02
T0003	2006	First	January	First	03
T0004	2006	First	January	First	04

Table 6.1(c) Product dimension

Product Key	Product Name	Brand	Category	Colour
P0001	Shirts	John Player	Formals	White
P0002	T-Shirts	Lee	Casuals	Maroon
P0003	Trousers	Peter England	Formals	Black
P0004	Jeans	Levis	Casuals	Blue

Table 6.1(d) Salesperson dimension

Salesperson Key	Name	Branch	Salary	Address
S0001	James	ABC	10K	XXXXXX
S0002	Bill	DEF	8K	YYYYYY
S0003	David	ABC	10K	ZZZZZZ
S0004	John	PQR	9K	AAAAAA

Table 6.1(e) Store dimension

Store Key	Area	City	State	Country
ST0001	Karol Bagh	Delhi	New Delhi	India
ST0002	Thane	Mumbai	Maharashtra	India
ST0003	Kondael Mall	Seattle	Washington	USA
ST0004	Bronchure Mall	San Antonio	Texas	USA

Table 6.1(f) Order fact

Time Key	Product Key	Store Key	Customer Key	Salesperson Key	Sales Amount	Sales Quantity	Profit
T0001	P0001	ST0003	C0001	S0001	200	20	30
T0001	P0002	ST0001	C0002	S0002	180	17	35
T0002	P0002	ST0004	C0003	S0004	150	15	27
T0002	P0003	ST0001	C0004	S0003	220	18	20

Advantages of Star Schema

- Simplicity – far easy to understand.
- Every query can be executed in same fashion
- Simple joins required – no intermediatory tables are required to be navigated.
- Analytic Flexibility : Facts can be accessed and analysed across multiple dimensions. The user can perform drill down and roll up.
- Easy to reconfigure: Dimensional attributes and fact elements can be added easily without affecting the other table.

Disadvantages of Star Schema

- Star Schema offers moderate performance
- Not suitable for storing detailed data.

Exercise

Generate Star Schema for the Information package Diagram created for Hotel Occupancy case discussed in last lecture.

Draw Star Schema for retail chain sales store from the IPD created in last lecture.

Exercise

GlobalBook Inc is a large distributor with domestic and international distribution channels. The company orders from publishers and distributes publications to all the leading booksellers. You want to build a date warehouse to analyse shipments that are made from the company’s warehouse. Prepare an IPD and a star schema for the same.