



Vidya Vardhini's College of Engineering and Technology
Department of Artificial Intelligence & Data Science

AY: 2024 - 25

Class:	TE	Semester:	V
Course Code:	CSC504	Course Name:	DWM

Name of Student:	Scunit Sunil khavu
Roll No. :	19
Assignment No.:	1
Title of Assignment:	
Date of Submission:	
Date of Correction:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Completeness	5	3
Demonstrated Knowledge Legibility	3	2
Legibility	2	2
Total	10	8

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Completeness	5	3-4	1-2
Demonstrated Knowledge Legibility	3	2	1
Legibility	2	1	0

Checked by

Name of Faculty

: Neha Raut

Signature

:

Date

:

Q.1) Given a sales cube with dimensions such as Time, Product, and Geography, describe how you would perform the following operations: roll-up to see yearly sales, drill down to see daily sales for a specific month, slice to view sales for a specific product category, dice to view sales for a specific region and product category, and pivot to change the view from product-centric to region-centric.

Ans:-

Performing operations on a sales cube with dimensions such as Time, product & geography can be done using Online Analytical Processing (OLAP) technique.

① Roll-up:- A roll-up operation aggregates data from a lower level to a higher level in dimension hierarchy.

In this case, we're moving from a more granular time view to an yearly view.

It start with the Time dimension at its lowest level (daily sales). Aggregate the sales data by summing it up to the year level. Result will show total sales for each year, combining all products & regions.

② Drill Down:- Drill down is the opposite of rollup moving from a higher level to a lower level in a dimension hierarchy. Begin with the Time dimension at the month level. It select a specific month of interest. It this will display daily sales for all products and regions within that month.

③ Slice:- Slicing involves selecting 1 particular member of a dimension and viewing data for all other dimensions.

In the product dimension , select a specific product category . It view the resulting 2D slice showing sales across Time and Geography for only that product category.

④ Dice :- Dicing is similar to slicing but involves selecting specific members from multiple dimensions.

It select a specific region from the geographic Dimension.

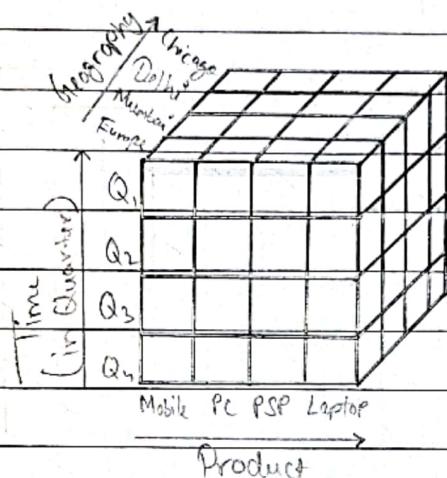
It choose a specific product category from the product dimension. It view the resulting subset of the cube , showing sales over Time for the chosen region and product category.

⑤ Pivot :- Pivoting involves rotating the cube to change the dimensional orientation of the result.

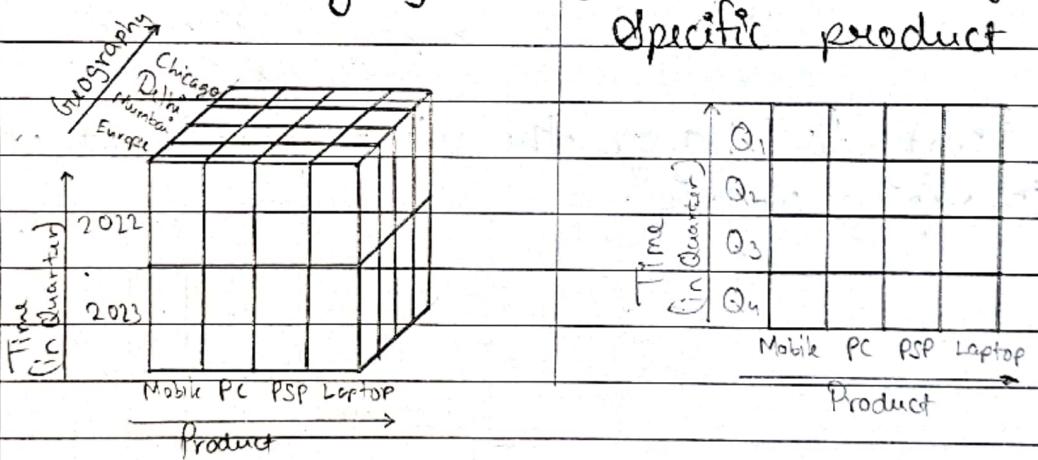
It start with a product - centric view where Product is on 1 axis and Time on another. It Rotate the cube to replace product with Geography on the axis. The result will show sales over Time for different regions , instead of ~~for~~ for different products.

These operations allows for flexible analysis of sales data across different dimensions , enabling businesses to gain insights into trends , performance and opportunities in their sales processes. Each operation provides a unique perspective , helping stakeholders make informed decisions based on the data.

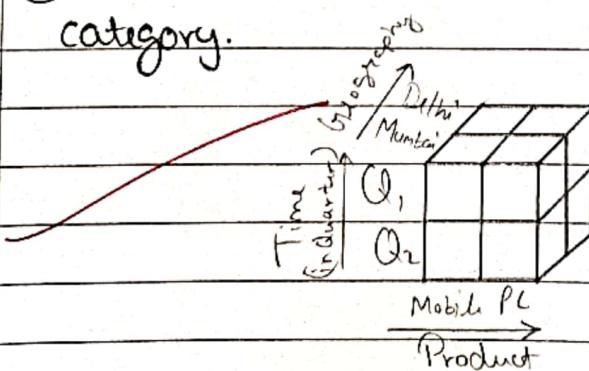
Sales cube with dimensions Time, product and geography.



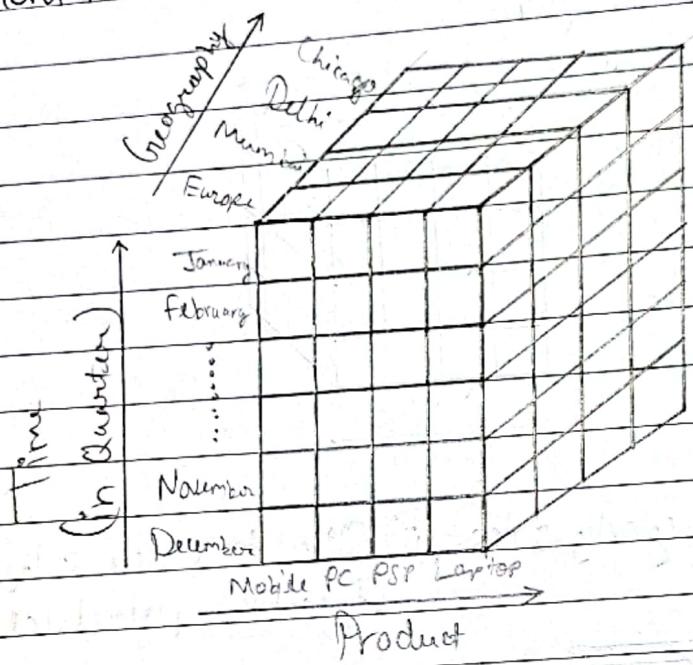
- (i) Roll-up to see yearly sales
- (ii) Slice to view sales for a specific product category



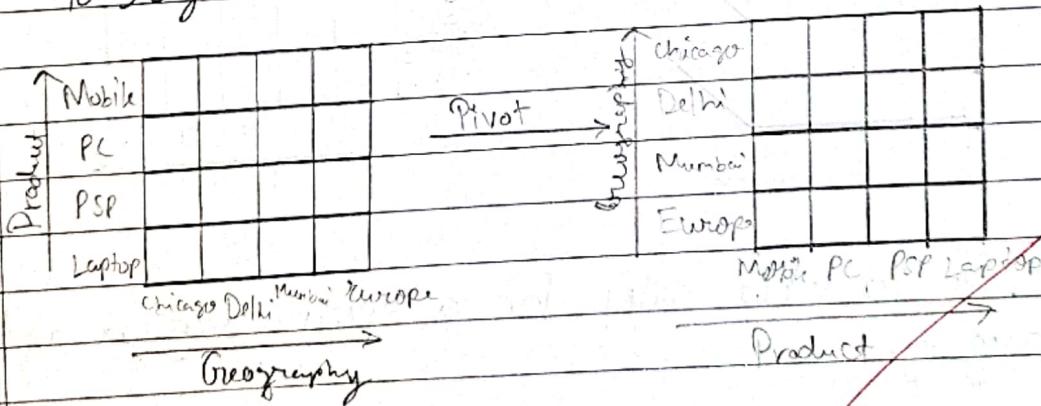
- (iii) Dice to view sales for a specific region and product category.



iv) Drill down to see daily sales for a specific month



v) Pivot to change the view from Product centric to region centric.



Q.2)

build a star and snowflake schema from online book store.
 Define the fact table, dimension table, and any subdimension tables.
 Explain the relationship between them.

Ans:-

Author

Dimension

Author_id(PK)

Author_name

country

Book dimension

Book_id (PK)

Title

Author_id(fk)

Price

Publication year

Publisher

Store Dimension

Store_id(PK)

Store_name

Store_location

Store_manager

Sales fact table

Sales_id (PK)

Book_id (fk)

Date

Customer

Dimension

Customer_id(PK)

Customer_id(fk)

Dimension

first_name

Store_id (fk)

Date

last_name

date_id(fk)

Day of week

birth_date

quantity_sold

Month

gender

total_amount

quarter

Address

discount

Year

Dimension

Profit

holiday_flag

address_id(PK)

email_id

street

Phone_number

city

address_id(fk)

state

country

zip_code

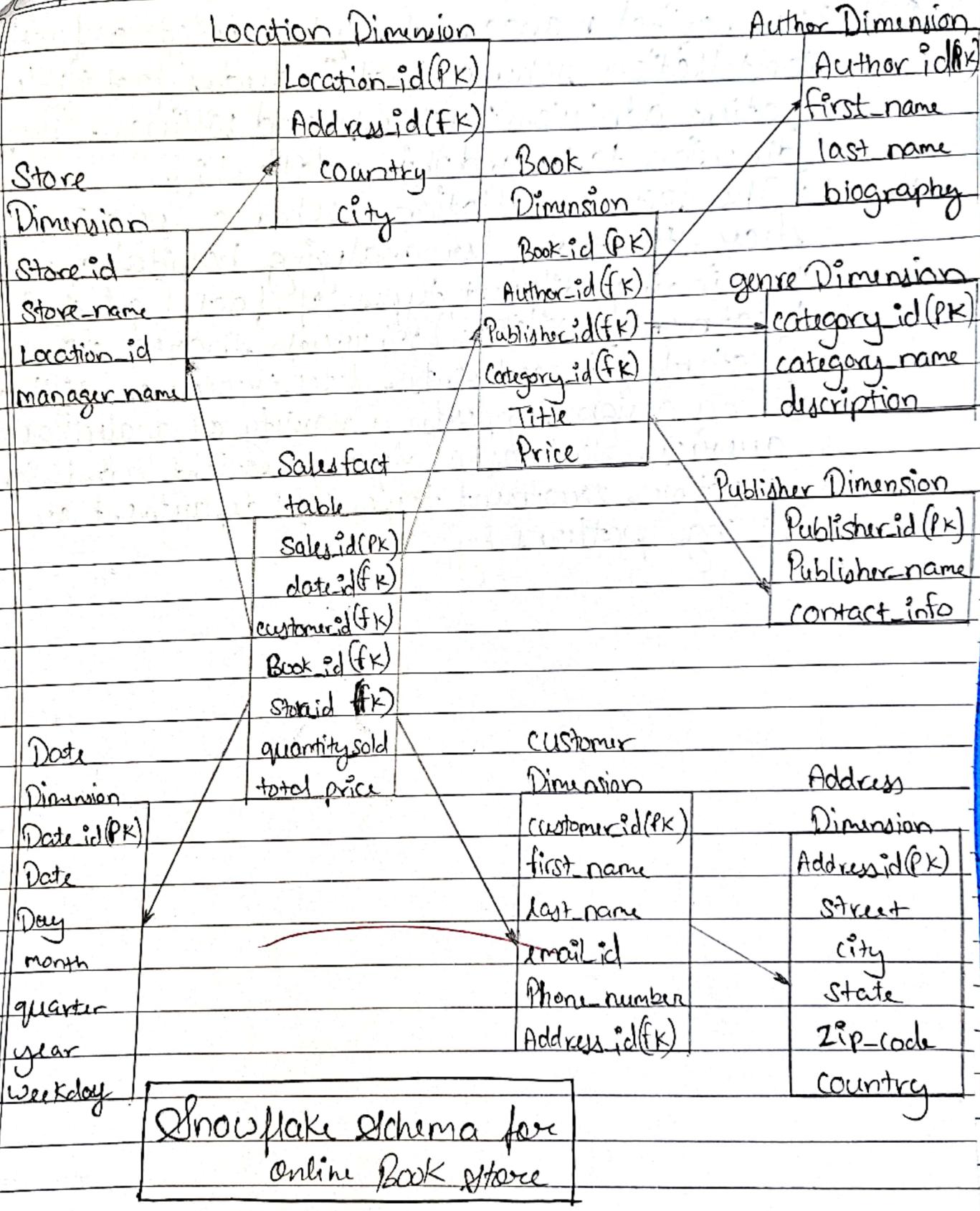
Star Schema

- ① Sales fact table is at the center of the schema, linking to the various dimensions.
- ② Book Dimensions in schema connects to Author and Publisher sub-dimension.
- ③ Store Dimension links to address sub-dimension
- ④ Store Dimension links to manager sub-dimension.
- ⑤ Date Dimension is stand alone, providing time-based attributes for analysis.

Snowflake Schema for online Book Store :-

- ① The Sales fact table is the central table that connects various dimension table.
- ② Each sale is linked to a specific date, customer, book and store.
- ③ The customer dimension includes customer details and links to the address dimension for further normalization.
- ④ The date dimension helps analyze sales over time.
- ⑤ The book dimension provides book-specific details and is linked to the location dimension, which is further normalized by linking to the address dimension.

The Star Schema is more straight forward and denormalized, making easier to query whereas the Snowflake Schema is normalized reducing redundancy but requiring more complex queries.



Q.3) For a hospital management system design a fact constellation schema that includes fact tables for patient admissions treatments and billings. Define the dimension tables and their relationship.

Ans -

- ① The fact constellation schema provides a clear structure for analyzing hospital operations by linking different types of facts (admissions, treatments, billing) through shared dimensions (patients, departments, dates, treatments, billing types.)
- ② This design supports a variety of analytical queries, such as analysing patient admissions overtime, treatment types by department or billing patterns by date.

Patient Dimension Table

Patient_id(PK)

Name

Age

Address

Contact_info

Doctor Dimension

Table

Doctor_id(PK)

Name

Specialization

Contact_info

Department_id(FK)

Treatment

Dimension

Table

Treatment_id(PK)

Treatment_name

cost

Billing Fact Table

Patient_id(FK)

Billing Date(FK)

Payment method

-Id (FK)

Amount

Insurance

Coverage

Patient admission

fact table

Patient_id(FK)

Admission Date(FK)

Discharge Date(FK)

Department_Id(FK)

Doctor_Id(FK)

Treatment

fact table

Patient_Id(FK)

Treatment Date(FK)

Treatment_Id(FK)

Department_id(FK)

cost

Duration

Success rate

Payment method

Dimension Table

Payment_method

-Id (PK)

Method

Name

Date Dimension

Table

Date_id(PK)

Date

Day

month

Quarter

year

Department

Dimension

Table

Department_Id(PK)

Name

Location

Date

Dimension

Table

Date_id(PK)

Date

Day

month

quarter

year