

* Designing Business Data Warehouse*

OLTP and OLAP Systems:-

OLAP

i) OLAP stands for online analytical processing

(i.) OLAP involves historical processing of information

(ii.) OLAP systems are used by knowledge-workers such as executives, managers and analysts.

iv.) OLAP is used to analyse business.

v.) OLAP is based on star-schema, snowflake schema & fact constellation schema.

vi.) OLAP contains historical data

vii.) OLAP provides summarised & multidimensional view of data.

viii.) The no. of users is in hundreds

OLTP

i) OLTP stands for Online Transactional processing

ii.) OLTP involves day-to-day processing.

iii.) OLTP systems are used by clerks, DBA's or database professionals.

iv.) OLTP is used to run the business.

v.) OLTP is based on Entity-Relationship (ER) Model.

vi.) OLTP contains current data.

vii.) OLTP provides detailed & flat relational view of data.

viii.) The no. of users is in thousands.

OLAP

OLTP

The no. of records accessed is in millions.

The database size is from 100GB to 100TB

i) OLAP access latency is negligible.

ix.) The no. of records accessed is in tens.

x.) The database size is from 100MB to 100 GB.

xi.) OLTP provides high performance.

Designing business Information DW :-

i) Principles of dimensional modeling.

ii) Data cube

iii) Data cube operations.

iv) Data cube schemas.

* 1st Model: E-R Model:-

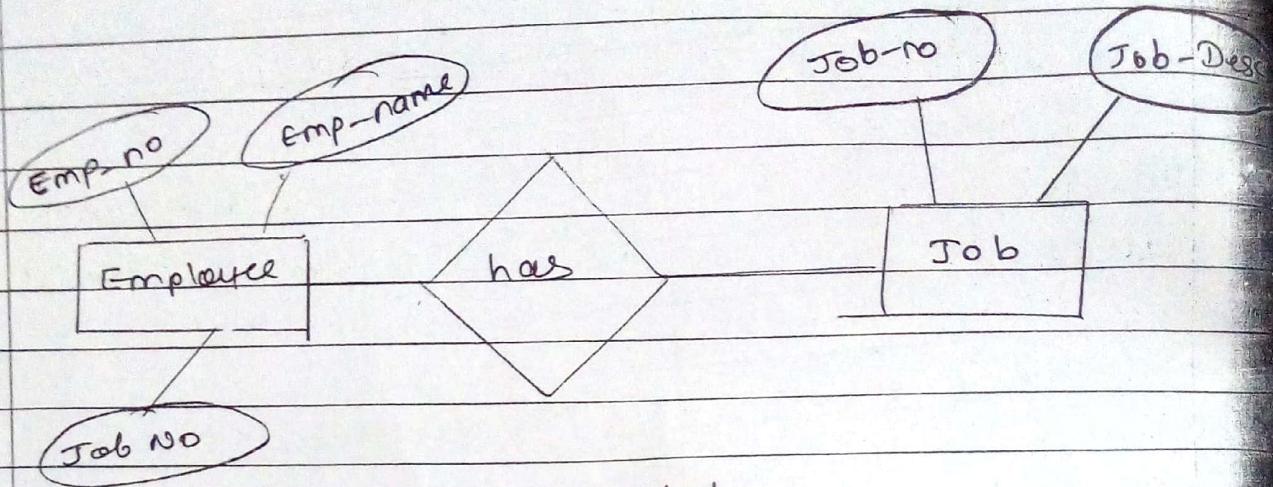
↳ ER stands for Entity - Relationship Data model.

↳ Basic component of ER model:-

i) Entity: It is real world object.

ii) Key: Uniquely defines entity.

iii) Relationship: It is used to describe association that exists between entities



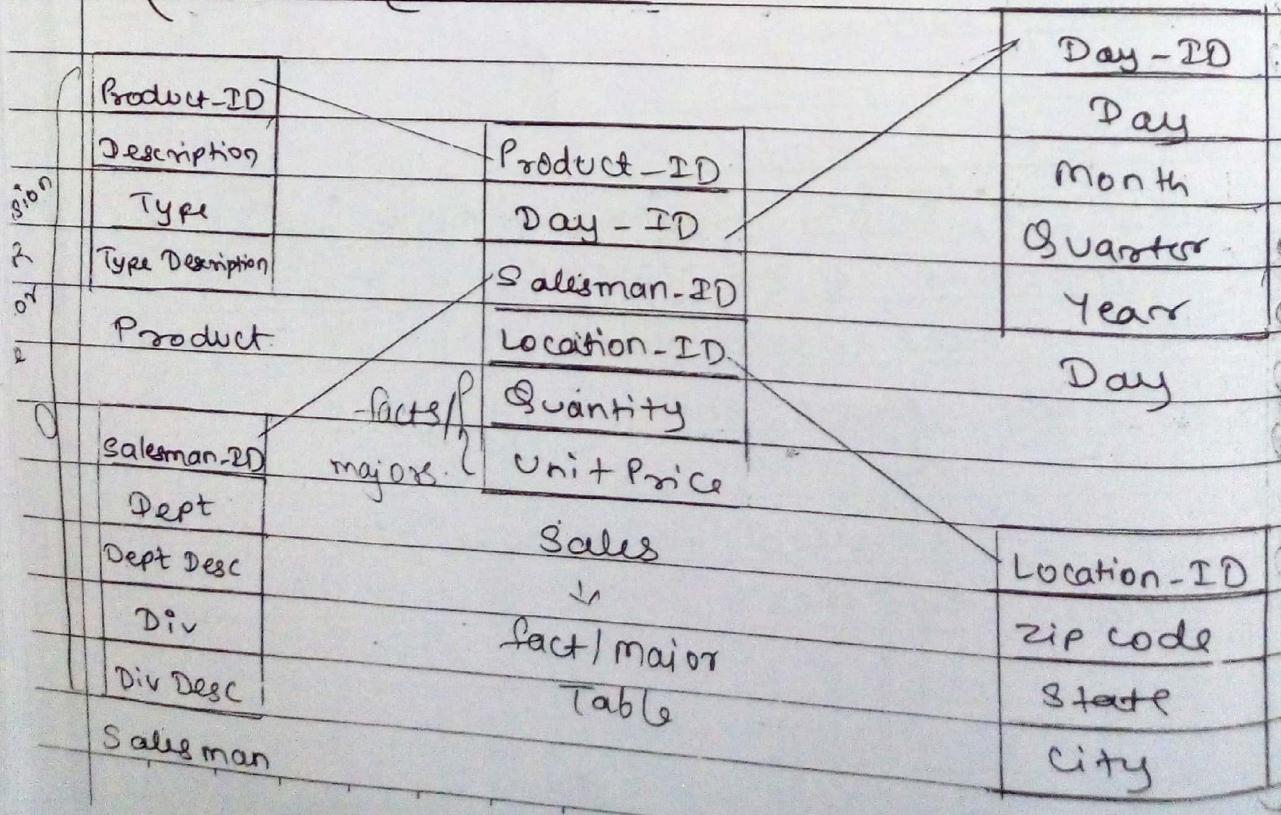
E-R model

Multidimensional Model:-

- It uses multidimensional schemas to ~~represent~~^{new} multidimensional data
- Multidimensional schema's include :-

 - 1.) Star - Schema
 - 2.) Snowflake schema
 - 3.) Fact constellation schema.

1.) Star Schema:-



A star schema shows data as collection of 8 types:-

i.) facts

ii) dimension.

Unlike relational schema which is flat, a star schema is a graphical view of the data. At the center of the star, data being examined the facts are shown in fact table, called major tables.]

[On the outside of the fact table each dimension is shown separately in dimension table called minor table.

Each facts points to one tuple in each of the dimension

Actual data are stored in fact table and tends to be quiet large.

Descriptive information about the dimension stored in the dimension table tends to be smaller.

Fact include quantity and price. (all numeric values)

2) Snowflake Schema:-

Location-id	zipcode	State	city
L1	Z1	A	X
L2	Z2	B	Y
L3	Z1	A	X
L4	Z1	A	X

Zip code:

Zipcode	State	City
z1	A	x
z2	B	y

Location:

L1	z1
L2	z2
L3	z1
L4	z1

Product ID		Day-ID
Description		Day
Type		month
Type Description		Quarter
Product	Product-ID	Year
	Day-ID	Day
	Salesman-ID	
	Location-ID	
	Quantity	
	Unit Price	

Salesman-ID	Sales	Location-ID
Dept		zip code
Dept Desc		Location
Div		\$
Div Desc		

Salesman

zip code
State
City

zip loc

- In snowflake, all or one dimension table may be normalized.
- Eliminates redundancy - save space
- Required when dimension table is wide.
- Complex query is required.

STUDY BUDDIES

Page 51 / 16

Snowflake schema is the extension of star schema.

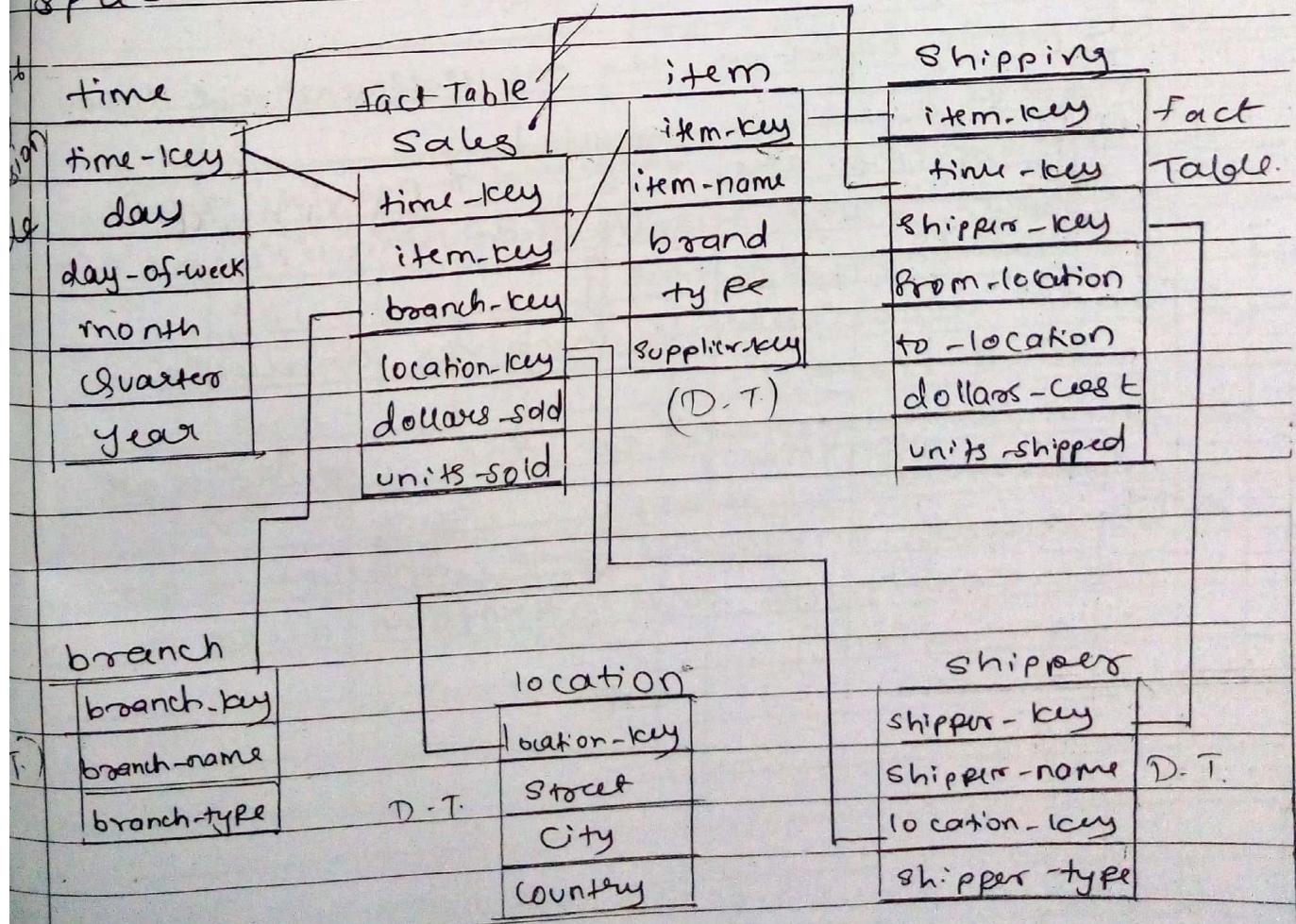
It facilitates more complex data view.

It is partially normalized version of star schema.

Unlike star schema, the dimension tables in a snowflake schema are normalized.

location table in snow

Due to normalization in snowflake schema, the redundancy is reduced and it becomes easy to maintain and save the storage space.



→ Fact Constellation Schema :-

- A fact constellation has multiple fact table
- It is also known as galaxy schema
- The diagram shows 2 fact tables, Sales and Shipping.
- The Sales fact table is same as that in star schema (item-key, time-key, Shippers-key, Freight)
- The shipping table has five dimensions and two measures.
- Two fact tables share the dimension tables.
Eg: Time, item & location tables are shared between Sales & Shipping fact table.

* O / A P :- Online Analytical processing (server)

- OLAP is based on the multidimensional data model]
- OLAP systems are targeted to provide more complex query results than traditional database systems.
- OLAP applications involves analysis of actual data
- OLAP is performed on data warehouse or data marts.
- It allows managers and analysts to get fast, consistent and interactive access to information.

Types of OLAP :-

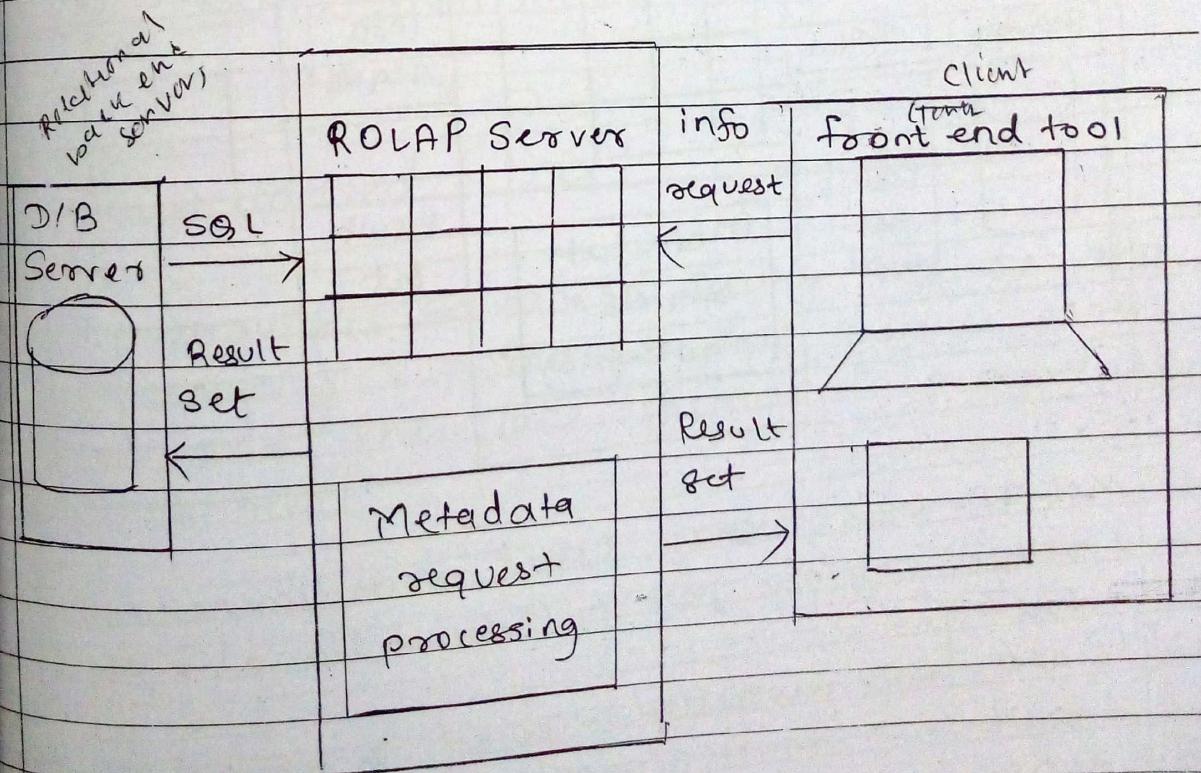
ROLAP (Relational OLAP)

MOLAP (Multidimensional OLAP)

HOLAP (Hybrid OLAP)

Specialized SQL Server.

ROLAP :- Relational Online Analytical Processing.

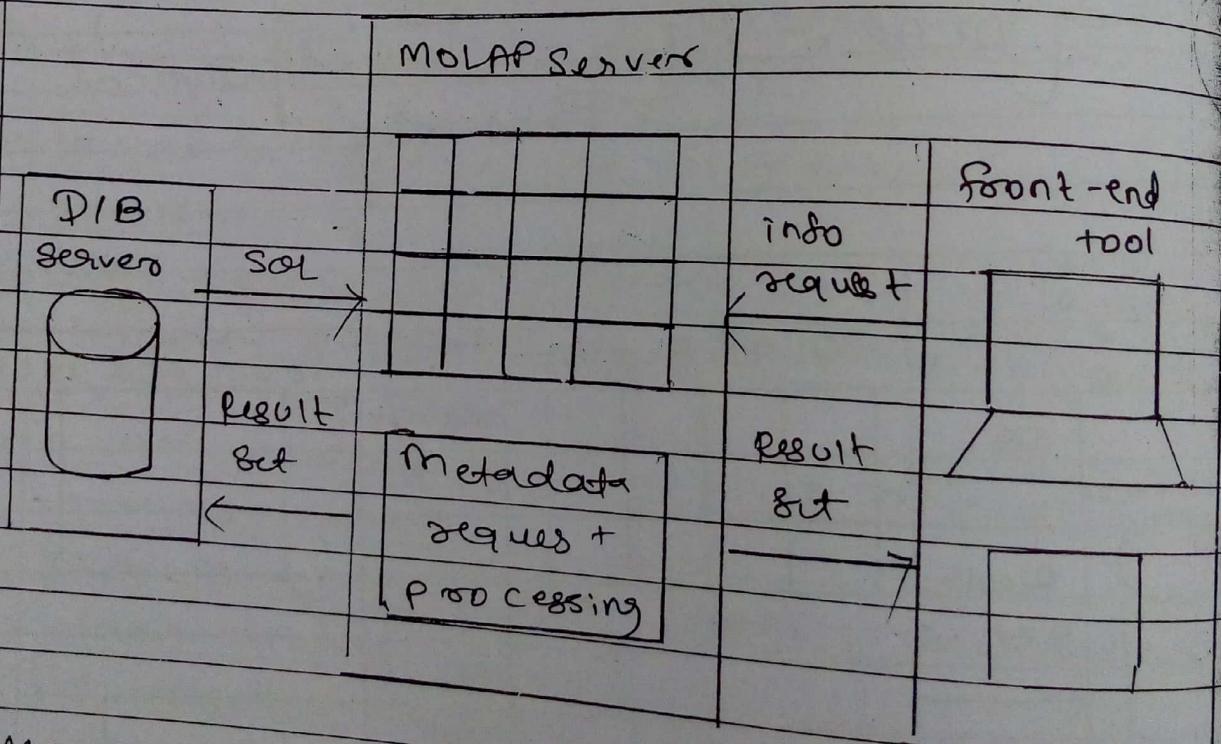


* ROLAP servers are placed between relational

back-end server and client front-end tool.

- 2.) To store and manage warehouse data, ROLAP uses relational databases.
- 3.) OLAP uses relational
- 3.) Data are stored in a relational database a rorap server (middleware) creates the multidimensional view of data for the user
- 4.) It is less complex and less efficient

2) MOLAP (Multidimensional OLAP)



- 1.) MOLAP uses array-based multidimensional storage engines for multidimensional data]
- 2.) Data are modelled, viewed, & physically stored in a multidimensional database. (MDDB are implemented by specialized tools)

Software system supporting multidimensional data]
MOLAP works on multidimensional data.]
With MOLAP, information retrieval is fast.
It is more efficient compared to
ROLAP.

- 3) Hybrid OLAP : (ROLAP + MOLAP) MOLAP
- i) offers higher Scalability of ROLAP & faster computation of -
 - It combines best features of ROLAP and MOLAP
 - Queries are stated in multidimensional form
 - Not frequently updated data stored as MDD
 - Frequently updated data stored as QDB
 - HOLAP offers higher scalability of ROLAP & faster computation MOLAP.

4) Specialised SQL Server

- It provides advanced query language and query processing support for SQL queries over star and snowflake schemas in a read only environment

Data Cube

Time	Location = "Gurgaon"			Location = "New Delhi"			Location = "Mumbai"					
	Item	mouse	mobile	modem	Item	mouse	mobile	modem	Item	mouse	mobile	modem
Q1	788	987	765	786	85	987	986	567	875			
Q2	678	654	987	659	786	436	980	876	908			
Q3	899	875	190	983	909	237	987	100	108			
Q4	787	969	908	537	567	836	837	926	987			

Location			cities	Mumbai	986	567	875	908	108	237	987
Time	mouse	mobile									
Q1	788	987	765								
Q2	678	654	987								
Q3	899	875	190								
Q4	787	969	908								

item(types)

* Description for Data Cube:

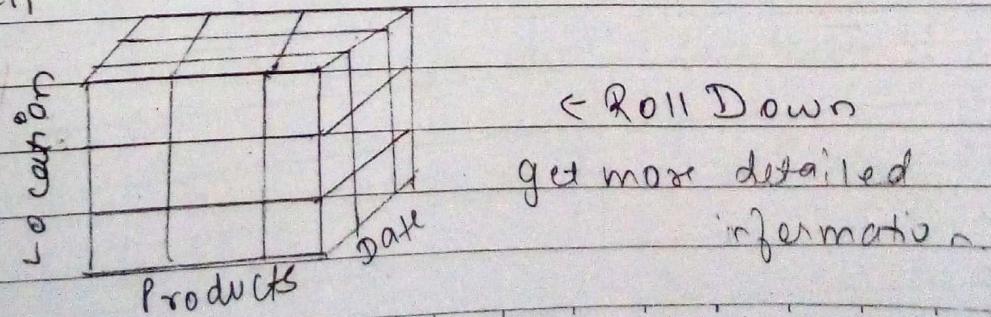
- A data cube represent the data in multiple dimensions. It is defined by dimensions & facts.
- The dimensions are the entities w.r.t. which enterprise preserves the record.

OLAP Operations:

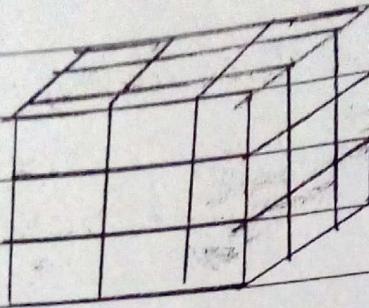
Date 3/2/16
Page 57
STUDY BUDDIES

- 1) Slice: This slice is performed by selecting one dimension, look at a subcube to get more specific information.
 - That means looking at the portion of the cube.
- 2) Dice: Dice look at a subcube by selecting on 2 or more dimensions.
 - It is performed by slice on one dimension and then rotating a cube to select on a 2nd dimension.
 - In figure the slice is rotated from all cells from of one product to all cells of one location, to get the dice.
- 3) Roll up: It allows the user to ask questions that move up an aggregation hierarchy.
 - In a rollup, instead of looking at a single fact, we look at all the facts.
Eg: look at overall total sales for the company.
go in detailed.
- 4) Roll Down: (Drill Down): This function allows the user to get more detailed fact information by navigating lower in the aggregation hierarchy.
Eg: look at quantities sold within a specific area of each of the cities.

(a) Single cell



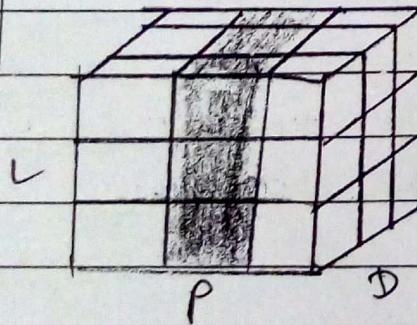
b.) Multiple cells



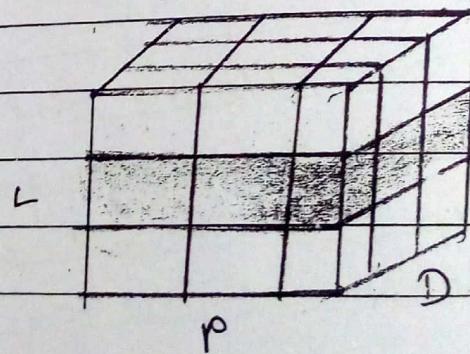
- Roll up

(looking at all
3 acts.)

c.) Slice



d.) Dice



Looking at
one portion of
the cube

Look at subcube by
selecting 2 or more
dimensions.