18CE1009
A1A3
Shivam Kendre

Date
Page 1

Q.1]

1] Data warehouse :-

A data warehouse is subject oriented integrated, time-variant and non volatile collection of data in support of management's decision making process

it was defined by bill inmon in 1990

characteristics

1) Subject oriented :-

Data that gives information about a particular subject instead of about a company's outgoing operations

2) Integraited :-

Data that is gathered into data warehouse from a variety of sources and integ merged into a coherent whole.

3) Time variant :-

All data in the data warehouse is identified with a particular time period

4) Non Volatile :-

Data is stable in data warehouse. but data is never removed. This enables management to gain consistent picture of business

Q. 1]

2]

OLTP	OLAP
1) it is transaction oriented	1) it is subject oriented
2) many users	2) few users
3) Real time information	3) historical information
4) it has highly volatile data	4) it has non volatile data
5) Single purpose model - supports operational system	5) Multiple models - support informational system
6) Validate model against business function analysis	6) Validate model against reporting requirements
7) Natural or surrogate key	7) Surrogate keys
8) Technical meta data depends on business requirements	8) Technical metadata depends on data mapping results
9) Data redundancy is bad	9) Data redundancy is good

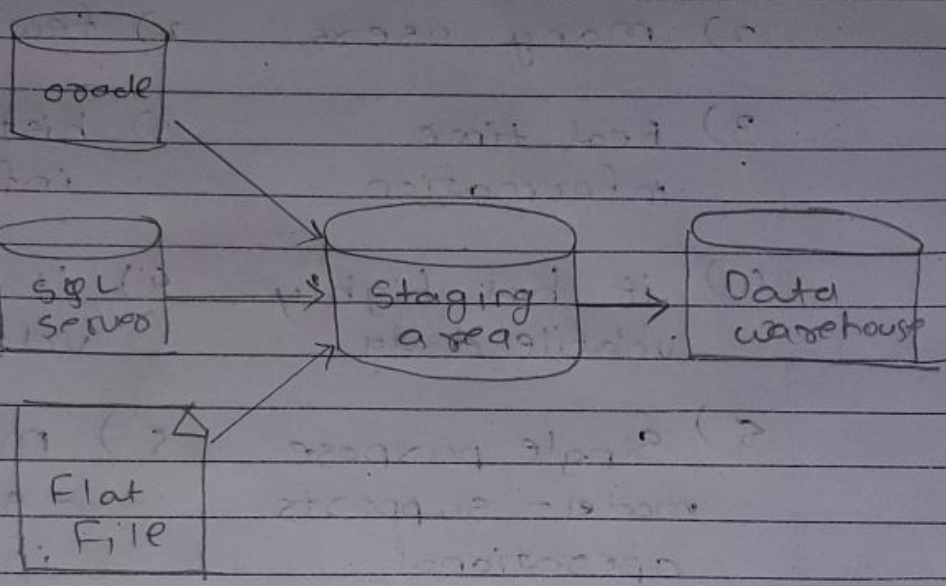
18CE1009
A1A3
Shivam Kendre

Date Page 3

Q.2]

2] a] ETL process in data warehouse

ETL is 3-step process



1) Extraction :-

In this step, data is extracted from the source system into the staging area. Transformations if any are done in staging area so that performance of source system is not degraded. Also, if corrupted data is copied directly from the source into Data warehouse database, roll back will be a challenge. staging area gives an oppurtunity to validate extracted data before it moves into the data warehouse.

2) Transformation :-

Data extracted from source server is raw and not usable in its original form. Therefore it needs to be cleaned. mapped and transformed. In fact, this is the key step where ETL process add value and changes data.

it is one of the important ETL concepts where you apply a set of functions on extracted data. Data that does not require any transformation is called as direct move or pass through data.

In transformation step, you can perform customized operation on data. for instance, if the user wants sum-of-sales revenue which is not in database

3) loading :-

loading data into the target datawarehouse database is last step of ETL process. In typical data warehouse, huge volume of data needs to be loaded in relatively short period. Hence this should b optimized for performance