NM-Big Data
Homework 7

1) What is a data warehouse? List the types of Data warehouse architectures.

A data warehouse is the electronic storage of an organization's historical data for the purpose of data analytics. In other words, a data warehouse contains a wide variety of data that supports the decision-making process in an organization.

There are mainly 3 types of data warehouse architectures:
- Single-tier architecture
- Two-tier architecture
- Three-tier architecture

2) What does OLAP stand for?

OLAP stands for OnLine Analytical Processing System.

It is a system which collects, manages, and processes multidimensional data for analysis and management.

3) What does OLTP stand for?

OLTP stands for OnLine Transaction Processing.

It is a system which modifies the data whenever it is received, to a large number of concurrent users.It is used to record every transaction.

4) What is a star schema?

Star schema is a schema used in data warehousing where a single fact table references a number of dimension tables. In a star schema, "keys" from all the dimension tables flow into the fact table. This entity-relationship diagram resembles a star, hence it is named a Star schema.

5) What is a snowflake schema?

Just like the star schema, a single fact table references a number of other dimension tables in snowflake scheme. Here however, these dimension tables are further normalized into multiple related tables. As these tables are further snowflake into smaller tables, this schema is called a snowflake schema.

6) Define fact-less fact.

Factless fact is a fact table that does not contain any value. Such a table only contains keys from different dimension tables.

7) **What do you understand by dimensional modeling?**

Dimensional model is a methodology that consists of "dimensions" and "fact tables". Fact tables are used to store various transactional measurements from "dimension tables" that qualify the data.

8) **What is a data mart?**

Data mart is a subset of organizational data.

In other words, it is a collection of data specific to a particular group within an organization.