**Mallu Sravana Sandhya Assessment-1**

**Data Engineering**

Designing ,building and scaling systems that organize the data for analytics

**Classification of data**

Raw data: data which comes from the data sources

Processed data: the data got when schema is applied to raw data

Cooked data: summerized data

All these data is stored in bigdata/data warehouses

**Bigdata properties**

Volume:- how much data you have

Velocity:- how fast data is getting to you

Variety:- how different your data is

Veracity:- how reliable your data is

**Data Processing methods**

Batch processing:-data is stored in storage then go for analytics and then go for insight

data

insight

Analytics

storage

Steam processing:- process data on fly as it comes on

Eg: streaming a live video on you tube

**Data warehousing**

Data warehouse is a collection of integrated structural and unstructural data during the processing of any application in operational Systems to support management decision making

**Purpose of Data Warehouse**

There are four features for datawarehouse

Subject oriented

* Data is organised focusing mainly on subject instead of any application
* It focuses on modeling and analysis of data for decision makers,not on daily operations that are performing

Integrated

* Datawarehouse is build by integrating multiple data files like relational databases,flat files etc
* Maintains consistency in naming, attribute measures etc among different data sources

Time varient

* It stores data for longer time than operational systems
* It maintains historical data over many years(5-10 years)

Nonvolatile

* No updates are allowed
* Once data is entered into datawarehouse we can not remove
* Represents company history

**Datawarehouse architecture**

accessing

Staging

Business data

Operational data

External data

datawarehouse

datamart

datamart

datamart

**Decision Support System**

It helps the businesses and organizations to make decisions

It resolves everyday business problems

It makes the analysis using structured data and makes business successful

**Operational data store**

Here the data collected fro data warehouse is processing by performing some operations and the data is stored here.

|  |  |
| --- | --- |
| **OLTP applications** | **Warehouse applications** |
| For Daily transactions  For Instant update | To store Historic data  To Add more data |

**Data marts**

The data fetched from the datawarehouses are organized by applying some operations or function which will support the decision makers, it is very useful for the management.

|  |  |
| --- | --- |
| **Datamarts** | **Data Warehouses** |
| Contains Structured data  Less data compared to warehouses  Consists of processed and cooked data  Maintains updated data  Eg: DMarts | Contains both stuctured and unstructred data  Vast amount data is present  Consists of raw data,processed data,cooked data  Maintains historic data  Eg: Storage houses of products |

**Datawarehouse lifecycle**

Requirements- data that should be stored in warehouse should be collected

Modeling- organize the data that was collected and build some connections between them.

ELT(Extract,Load, Transfer)- here data is extracted from the data sources and transfer them to different source and then load it into the data warehouse using different tools

OLAP- online analytical processing the data in the warehouses is analyzed using some operation/functions and data is processed in the form of tables,spread sheets,etc.

UI development- making the data visible for the client which is helpful to use data and make decisions

Maintenance- the datawarehouse should be maintained by updating the data that is used by the clients

Test- organisations test the warehouses to ensure that all requirements of business are meet