MANDAPALLI HEMA

ASSIGNMENT-1

Data Engineering:

Designing of model,Building of model and scaling system that organize data(collection of information) for analytics(measuring of things).

ETL(Extract,Transform,Load):

It is used to collect the data from various sources. It transform

the data and that will store in the database.

Data Classification:

DATA

BIG DATA

It will stored in the big data

* Summirized data

COOKED DATA

* The data which is processed.
* Schema is applied
* The data which doesn’t have any parameters.
* It just as coming from the data source.
* No schema applied

RAW DATA

PROCESSED DATA

Big data properties:

Volume: how much data we have.

Velocity: how fast the data is getting.

Variety: how different the data is.

Variency:how the data is reliable.

Batch processing:

The processing of data which are stored and the particular data will measured in analytics then it go for insight.

Stream processing:

Stream means continuous amount of flowing. The sequence of events in time.

Map reduce:

Big will work on map reduce.organising of data and values it combine with the matching key until we have the final key value.

Examples: apache spark, hadoop

Data storage:

1. Relational Database(SQL)
2. Document Database(NO SQL)

Data warehouse:

It is a storing of data, Datamarts. The place where we place a entire data. It is a subject oriented,integrated,timevariant,non volatile collection of data in support of management system.

Features of data warehouse:

* Subject-oriented: it is focus on modelling and analysis of data for decision makers and not on transcation processing.
* Integrated: data warehouse is integration of different typeof data sources.
* Time variant:data collected in a data warehouse identifed with a particular period:
* Non-volatile:data is not erased when the new data is added to it.

Purpose of data warehouse:

to perform queires and analysis and it contains large amount of historical data.To centralize large amount of data from multiple sources.

Database architecture:

It discribes how a database management system will be integrated with our application. Database could centralized or decentralized.

Operational data store:

usually stores and processes data in real time.it is commonly used in the transaction processing applications and it is a central database.

OLTP VS WAREHOUSE APPLICATIONS:

OLTP: it allows us to use it to read,update,and delete the data it contain.

WAREHOUSE APPLICATIONS:banking,airline industry,insurance sector,telecommunications

DATA MARTS VS DATA WAREHOUSES:

A data warehouse stores the data in structured formatand data and A data mart is a data warehouse that serves the needs of a specific business unit.

Data marts:

A data storage system that contains information specific to an organization.

Data Warehouse life cycle:Requirement analysis, design,construction,testing,deployment,operation,maintenance and retirement.