



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)
COIMBATORE – 35

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (UG & PG)

Subject Code & Name: 19ITE305-Big Data& Analytics

| Hadoop | Data warehouse |
|--|--|
| An open-source software framework for the distributed storage and processing of huge datasets. | a central database of structured, ordered data. |
| It uses Distributed file system (HDFS) for data storage. | It uses a Relational database or structured storage system for data storage. |
| MapReduce programming model and ecosystem are used for data processing. | SQL-based queries are used for data processing. |
| Designed to scale horizontally. | Designed to scale vertically. |
| It can handle variety of data like structured, unstructured and semi structured data. | It can mainly handle structured data. |
| It offers high scalability and is capable of handling petabytes of data. | The scalability offered by a data ware house is limited depending on hardware resources. |
| The speed of processing data is very slow. | The data processing speed is faster in the data warehouse. |

| | |
|---|---|
| It is ideal for complex data transformations. | It has limited capability to handle complex data transformations. |
| It is affordable and has quite a lower cost. | It is highly expensive. |
| It provides direct access to raw data. | It provides aggregated data for analysis purposes. |
| It uses the "Schema-on-Read" Data schema. | It uses the "Schema-on-Write" Data schema. |
| It is mainly used for big data analysis and processing. | It is mainly used for reporting and business intelligence. |