60. Subrato - D15AD

BDA - Experiment 8

Aim: Use Sqoop to load data from RDBMS (weblog/transaction data) and analyse it using Hive/Pig.

Theory: Hadoop ecosystem is a collection of open-source tools and frameworks designed to process, store and analyze large volumes of data in a distributed computing environment.

- **Hive:** Hive is like a translator for Hadoop. It allows you to write queries in a language similar to SQL (called HiveQL) and then translates those queries into MapReduce jobs that can be executed in a Hadoop Cluster.

It's great for data analysts who are familiar with SQL because they can use Hive to query and analyse data stored in Hadoop's distributed file system - HDFS.

Pig: Pig is a platform that simplifies the process of writing data transformation for Hadoop instead of writing complex Java code for MapReduce, you can use a simple scripting language called Pig Latin.

Sqoop: Sqoop is a tool for efficiently transferring data b/w Hadoop and relational databases like MySQL or Oracle. It helps to import data from databases into Hadoop or export data from Hadoop back to databases. It is essential when we have traditional databases that we want to analyse with Hadoop.

Conclusion: We have used Sqoop to understand and load data from RDBMS. We also understood its analysis using Hive & Pig.