DAY - 8 HDFS Features

14 May 2024 11:08

HDFS Features and Goals

- The Hadoop Distributed File System (HDFS) is a distributed file system.
- It is a core part of Hadoop which is used for data storage.
- It is designed to run on commodity hardware.
- Unlike other distributed file system, HDFS is highly fault-tolerant and can be deployed on low-cost hardware.
- It can easily handle the application that contains large data sets.

Let's see some of the important features and goals of HDFS.

Features of HDFS

- **Highly Scalable -** HDFS is highly scalable as it can scale hundreds of nodes in a single cluster.
- **Replication -** Due to some unfavorable conditions, the node containing the data may be loss. So, to overcome such problems, HDFS always maintains the copy of data on a different machine.
- **Fault tolerance -** In HDFS, the fault tolerance signifies the robustness of the system in the event of failure. The HDFS is highly fault-tolerant that if any machine fails, the other machine containing the copy of that data automatically become active.
- **Distributed data storage -** This is one of the most important features of HDFS that makes Hadoop very powerful. Here, data is divided into multiple blocks and stored into nodes.
- **Portable -** HDFS is designed in such a way that it can easily portable from platform to another.

Goals of HDFS

- **Handling the hardware failure -** The HDFS contains multiple server machines. Anyhow, if any machine fails, the HDFS goal is to recover it quickly.
- **Streaming data access -** The HDFS applications usually run on the general-

purpose file system. This application requires streaming access to their data sets.

- **Coherence Model -** The application that runs on HDFS require to follow the write-once-ready-many approach. So, a file once created need not to be changed. However, it can be appended and truncate.