SQOOP

22 September 2020   20:52

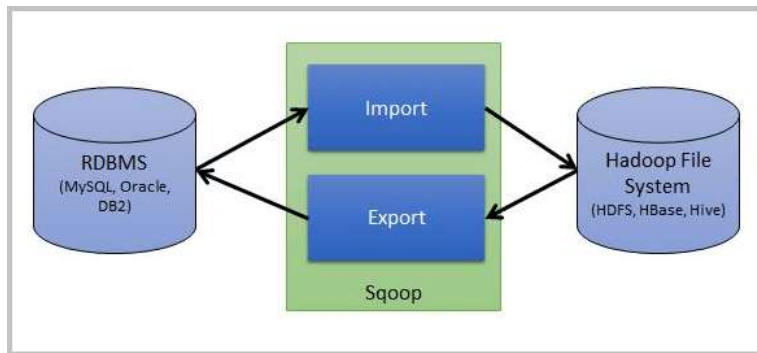**Data INGESTION Tools (to fetch data in to HDFS):**
1. Sqoop
2. Flume

Sqoop is meant to fetch structured data in HDFS from any RDBMS server (MySQL server in my case).

Features:
-> It is named Sqoop as SQL to Hadoop, Hadoop to SQL
-> It is a command line interpreter

-> Basically whenever need to process Big Data using MapReduce or HIVE, the data was supposed to be a part of HDFS only. However, most of the data that we had before was saved into some other databases and this data was supposed to be transferred into HDFS in bulk, thus we came across tools like Sqoop and Flume
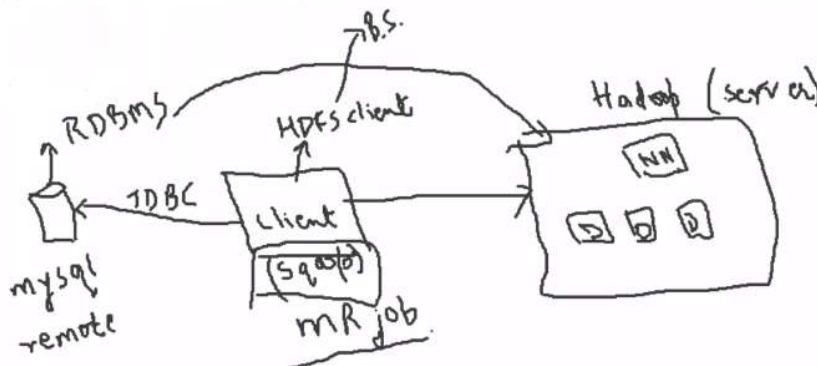

For Sqoop, we can export and import both(as shown in the diagram) and that too in a parallel manner

-> We can do all types of import: be it incremental import or complete import
-> It uses JDBC driver to connect to the database(one needs to download the jars and then put it in lib folder so that it can interact with the database)

To Install Sqoop:
sudo yum install sqoop

Sqoop Architecture:
https://medium.com/edureka/apache-sqoop-tutorial-431ed0af69ee