

SQOOP

Hadoop is also known as a new paradigm of handling data processing at monstrous scale. In any case, so many data processing pipelines and applications still make use of relational databases. A typical need, in cases like this, is to consider the importation of data from the relational database just like SQL Server, Oracle, etc. directly to Hadoop for the purpose of processing and then export the data being processed back from Hadoop straight to the RDBMS. The final data being processed in the RDBMS is utilized as one of an application or together with the data in the RDBMS for any analytic needs.

In the Hadoop environment, Sqoop is referred to as a tool of choice for this purpose. Apache Sqoop is another tool designed to efficiently transfer enormous data between Hadoop and the stores of structured data like the relational databases (for instance, SQL Server, Oracle). Sqoop benefits greatly from both worlds. It depends on the database for the definition of schema and utilizes Map Reduce for the importation and exportation of the data. With the use of Map Reduce, you have the ability of scaling to any number of machines and you will be able to run codes in parallel. Also, it offers fault tolerance, which is a vital piece present in the machines with hundreds or even thousands of machines.

<https://sqoop.apache.org/>