

Assignment 10.1

1. Workflow of Oozie and its benefits

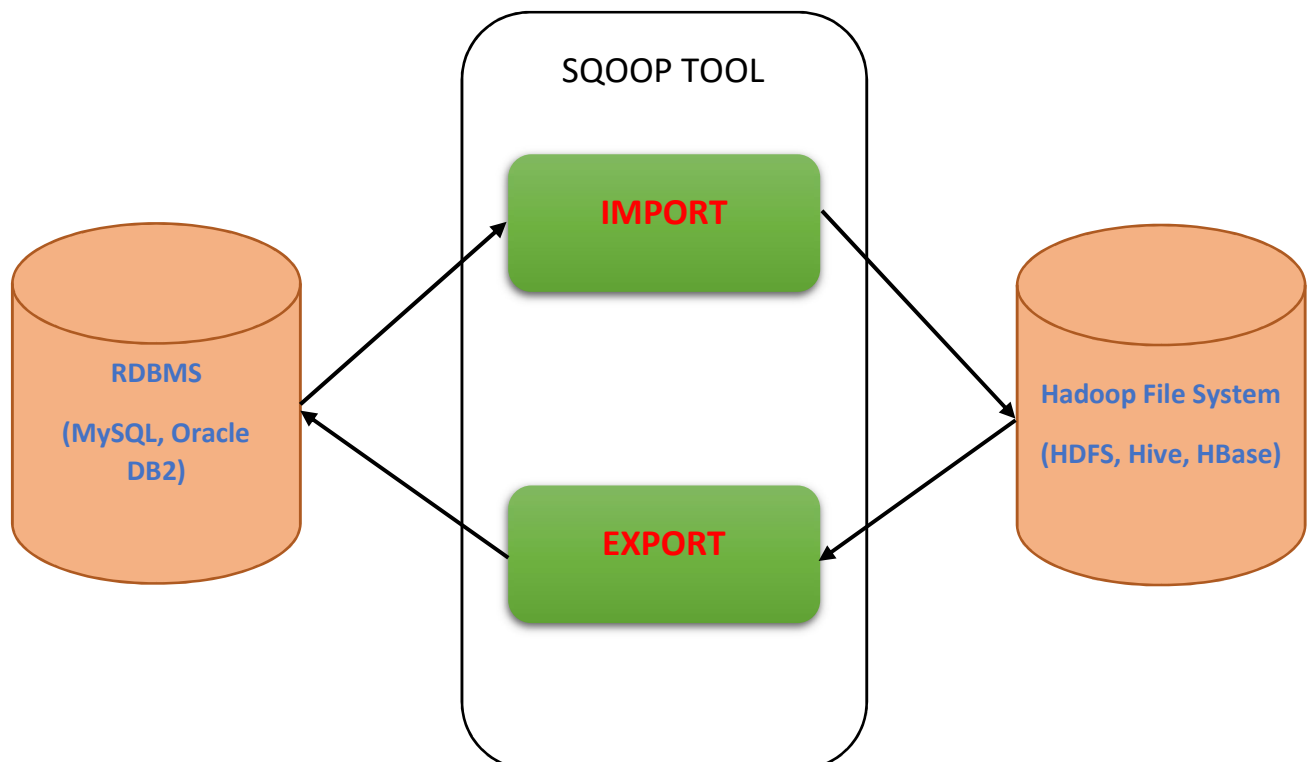
Hello Team, mentor has not yet completed the topic on Oozie, an email has been sent to supports@acadgild.com , notifying the same.

2. Workflow of Sqoop and its Benefits.

Apache Sqoop efficiently transfers bulk data between Apache Hadoop and structured datastores such as relational database. Sqoop helps offload certain tasks from EDW to Hadoop for efficient execution at a much lower cost.

Sqoop can also be use to extract data from Hadoop and export it into external structured datastores.

Sqoop works with relation databases such as Teradata, Netezza, Oracle, MySQL, Postgres and HSQLDB.



Few Functionalities of Sqoop are:

- Import sequential datasets from mainframe
- Import direct to ORC Files
- Data imports
- Parallel data transfer
- Fast data copies
- Efficient data analysis
- Load Balancing.

Benefits of Sqoop are:

- Satisfies the growing need to move data from mainframe to HDFS.
- Improved compression and light-weight indexing for improved query performance.
- Move certain data from external stores and EDW's into Hadoop to optimize cost-effectiveness of combined data storage and processing.
- Parallel data transfer for faster performance and optimal system utilization.
- Fast data copying from external systems into Hadoop.
- Improves efficiency of data analysis by combining structured data with unstructured data in a schema-on-read data lake.
- Mitigates excessive storage and processing loads to other systems.