

Assignment 10.1

Problem Statement 1 :-

Explain in Brief:

- The workflow of Oozie and its Benefits
- The workflow of Sqoop and its Benefits

Solution:-

- **Workflow of Oozie and its Benefits:-**

As discussed with Amit and Acadgild Support Team, this topic is not yet covered by Mentor. I will submit this portion of the assignment when it is completed by the Mentor.

- **Workflow of Sqoop and its Benefits:-**

Sqoop allows easy import and export of data from structured data stores such as relational databases, enterprise data warehouses, and NoSQL systems. Using Sqoop, we can provision the data from external system on to HDFS, and populate tables in Hive and HBase. Sqoop integrates with Oozie, allowing you to schedule and automate import and export tasks. Sqoop uses a connector based architecture which supports plugins that provide connectivity to new external systems.

The following command is used to import all data from a table called company from a MySQL database using Single Mapper:-

```
sqoop import --connect jdbc:mysql://localhost/db1 --username 'acadgild' -P --table 'company' --target-dir '/sqoopout' -m 1;
```

The import is done in two steps:-

- In the first Step Sqoop introspects the database to gather the necessary metadata for the data being imported.
- The second step is a map-only Hadoop job that Sqoop submits to the cluster.

Benefits of Sqoop:-

Ease of Use - Sqoop lets connectors to be configured in one place, which can be managed by the admin role and run by the operator role. This centralized architecture helps in better deployment of Big Data analytics and solutions.

Ease of Extension - The connectors of Sqoop are not restricted to just the JDBC model. It has the competencies to extend and define its own vocabulary without having the need to mention a table name.

Security - The fact that Sqoop operates as server based application that secures access to external systems and does not allow code generation, makes its security to go by.

Submitted By:-

Hardik Kaushik