**ASSIGNMENT 10.1**

PROBLEM STATEMENT:

**1) Explain the workflow of Oziee and its Benefits.**

Solutions:



Oziee workflow nodes:

a) Control flow:

i) Start/end/kill.

ii) Decision.

iii) Fork/join.

b) Actions:

i) Map-reduce

ii) Pig

iii) Hdfs

iv) Sub-workflow

v) Java-run custom java code

To run the oziee workflow, two files are needed:

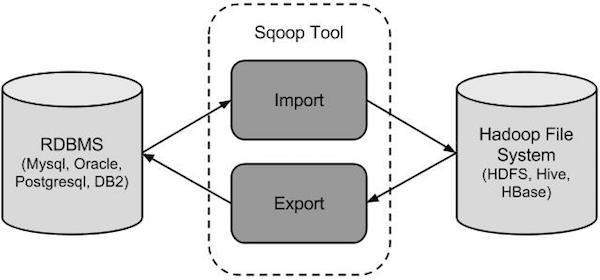
i) Workflow.xml (stored in Hdfs)

It contains the structure of the workflow.

ii) job.properties (stored in Local)

It contains the configuration properties.

**2) Explain the workflow of Sqoop and its benefits.**



**Sqoop Import**

The import tool imports individual tables from RDBMS to HDFS. Each row in a table is treated as a record in HDFS. All records are stored as text data in text files or as binary data in Avro and Sequence files.

**Sqoop Export**

The export tool exports a set of files from HDFS back to an RDBMS. The files given as input to Sqoop contain records, which are called as rows in table. Those are read and parsed into a set of records and delimited with user-specified delimiter.

**Benefits of Sqoop:**

1. Apache Sqoop automates most of this process, relying on the database to describe the schema for the data to be imported.

2. Sqoop uses MapReduce to import and export the data, which provides parallel operation as well as fault tolerance.

3. With Apache Sqoop, you can import data from a relational database system or a mainframe into HDFS.

4. When we execute any Apache Sqoop command it converts into a Java jar file which executes and inside the jar you can see the Map and Reduce execution.