Explain in Brief:

● The workflow of Oozie and its Benefits

Oozie is mainly used to manages the [hadoop jobs in HDFS](http://www.credosystemz.com/training-in-chennai/best-bigdata-training-in-chennai/" \t "_blank) and it combines the multiple jobs in particular order to achieve the big task. It is the open source framework and used to make multiple hadoop jobs. Oozie supports the jobs in mapreduce,hive and hdfs also. In Oozie job workflow based on Directed Acylic Graph and it contains two nodes for managing the jobs that nodes are action and control flow nodes.

Advantages of Oozie is it integrate with [hadoop stack](http://www.credosystemz.com/training-in-chennai/best-bigdata-training-in-chennai/" \t "_blank) and also support mapreduce and hdfs jobs. Oozie contains following three types of jobs

**1. Workflow jobs –** It used to represent the sequence of jobs executed.

**2. Coordinator Jobs –** It contains workflow jobs and it triggered by time

**3. Bundle Jobs –** It contains the workflow and coordinator jobs

Oozie Benefits:

* Complex workflow action dependencies: Oozie workflow comprises of actions and dependencies among them.
* Reduces Time-To-Market (TTM): The DAG specification enables users to specify the workflow.
* Frequency execution: Users can specify execution frequency and can wait for data arrival to trigger an action in the workflow.
* Native Hadoop stack integration: Oozie supports all types of Hadoop jobs.
* Oozie is validated against the Hadoop stack.
* Oozie is integrated with the Yahoo! Distribution of Hadoop with security and is a primary mechanism to manage a variety of complex data analysis.

**● The workflow of Sqoop and its Benefits**

Apache Sqoop is a tool in [Hadoop ecosystem](https://www.edureka.co/blog/hadoop-ecosystem) which is designed to transfer data between [HDFS](https://www.edureka.co/blog/hdfs-tutorial) (Hadoop storage) and relational database servers like mysql, Oracle RDB, SQLite, Teradata, Netezza, Postgres etc. Apache Sqoop imports data from relational databases to HDFS, and exports data from HDFS to relational databases. It efficiently transfers bulk data between Hadoop and external datastores such as enterprise data warehouses, relational databases, etc.

This is how Sqoop got its name – “SQL to Hadoop & Hadoop to SQL”.

A common workflow with Sqoop is one where data is brought in Hive so intermediate processing and transformation tasks can be done on Hadoop. Once processing is done data can be exported back to a database. This is one of many ways to perform "data warehouse offloading" where Hadoop is used for ETL purposes

Benefits of sqoop:

1. imports
2. exports
3. incremental imports
4. import all tables from a single database in one go