**1)Oozie workflow and its benefits**

Oozie is mainly used to manage the [**hadoop jobs in HDFS**](http://www.credosystemz.com/training-in-chennai/best-bigdata-training-in-chennai/) 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 map reduce, 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/) and also support map reduce and hdfs jobs. Oozie contains following three types of jobs

**1. Workflow jobs** – It used to represents 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

[**Types of Nodes in Apache Oozie:**](http://www.besthadooptraining.in/)

**Action Node** – It represents the workflow jobs and jobs program are written in java

**Control Flow Node** – It used to controls the workflow jobs between actions

**Start Node** – It used to starts the jobs execution

**End Node** – It used to stops the jobs execution

**Error Node** – If any error occurs while execution of job error node prints the error message

**2)Sqoop workflow and its benefits**

Apache Sqoop efficiently transfers bulk data between Apache Hadoop and structured datastores such as relational databases. Sqoop helps offload certain tasks (such as ETL processing) from the EDW to Hadoop for efficient execution at a much lower cost. Sqoop can also be used to extract data from Hadoop and export it into external structured datastores. Sqoop works with relational databases such as Teradata, Netezza, Oracle, MySQL, Postgres, and HSQLDB

**What Sqoop does**

Apache Sqoop does the following to integrate bulk data movement between Hadoop and structured datastores:

| **Function** | **Benefit** |
| --- | --- |
| Import sequential datasets from mainframe | Satisfies the growing need to move data from mainframe to HDFS​ |
| Import direct to ORCFiles | ​Improved compression and light-weight indexing for improved query performance |
| Data imports | Moves certain data from external stores and EDWs into Hadoop to optimize cost-effectiveness of combined data storage and processing |
| Parallel data transfer | For faster performance and optimal system utilization |
| Fast data copies | From external systems into Hadoop |
| Efficient data analysis | Improves efficiency of data analysis by combining structured data with unstructured data in a schema-on-read data lake |
| Load balancing | Mitigates excessive storage and processing loads to other systems |

YARN coordinates data ingest from Apache Sqoop and other services that deliver data into the Enterprise Hadoop cluster.