1. The workflow of Oozie and its Benefits

Apache Oozie is a workflow scheduler for Hadoop. It is a system which runs workflow of dependent jobs. Here, users are permitted to create **Directed Acyclic Graphs** of workflows, which can be run in parallel and sequentially in Hadoop.

It consists of two parts:

* **Workflow engine** : Responsibility of a workflow engine is to store and run workflows composed of Hadoop jobs e.g., MapReduce, Pig, Hive.
* **Coordinator engine**: It runs workflow jobs based on predefined schedules and availability of data.
* **Oozie Bundle** − These can be referred to as a package of multiple coordinator and workflow jobs.

Oozie is scalable and can manage timely execution of thousands of workflows (each consisting of dozens of jobs) in a Hadoop cluster.

Oozie is very much flexible, as well. One can easily start, stop, suspend and rerun jobs. Oozie makes it very easy to rerun failed workflows.

● The workflow of Sqoop and its Benefits