

**Solution-**

**Apache Oozie-**

Apache Oozie is a real time scheduler and workflow engine that blends well with large production environments. It is a server based workflow engine

**Apache Oozie Workflow nodes-**

**Control Flow**

• Start/end/kill

• Decision

• Fork/join

**Actions**

• Map-reduce

• Pig

• Hdfs

• Sub-workflow

• Java-run custom java code

**Oozie workflow nodes-**

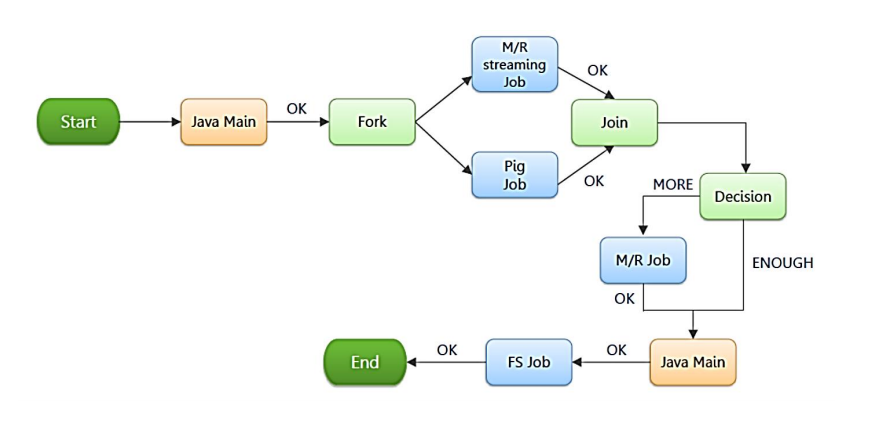
To run oozie workflows, two files are needed.

1. workflow.xml (stored in HDFS)

• It contains the structure of workflow.

1. job.properties (stored in local)

• It contains the configuration properties.



Oozie is a workflow scheduler system to manage Apache Hadoop jobs.

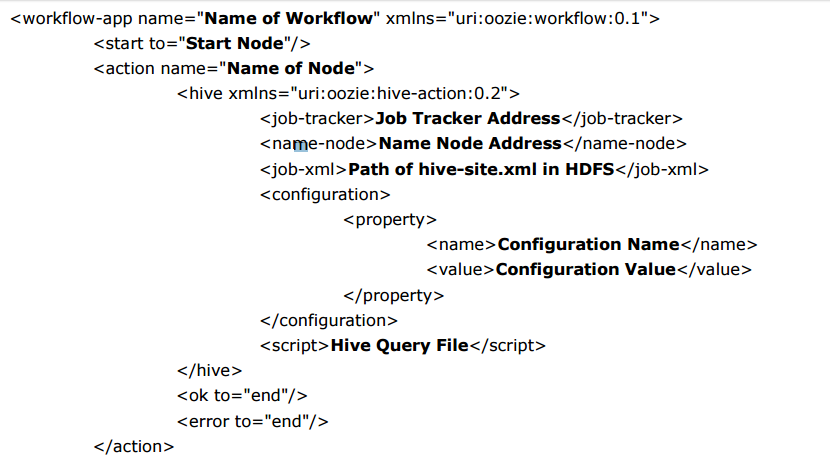
Oozie Workflow jobs are Directed Acyclical Graphs (DAGs) of actions.

Oozie Coordinator jobs are recurrent Oozie Workflow jobs triggered by time (frequency) and data availability.

Oozie is integrated with the rest of the Hadoop stack supporting several types of Hadoop jobs out of the box (such as Java map-reduce, Streaming map-reduce, Pig, Hive, Sqoop and Distcp) as well as system specific jobs (such as Java programs and shell scripts).

Oozie is a scalable, reliable and extensible system.

Developers interested in getting more involved with Oozie may join the mailing lists, report bugs, retrieve code from the version control system, and make contributions.

****