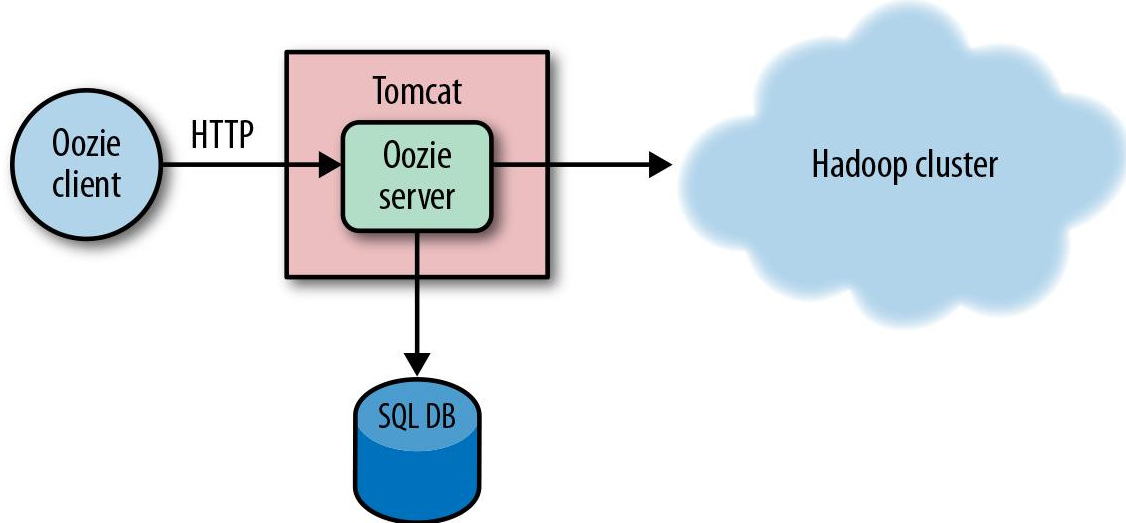
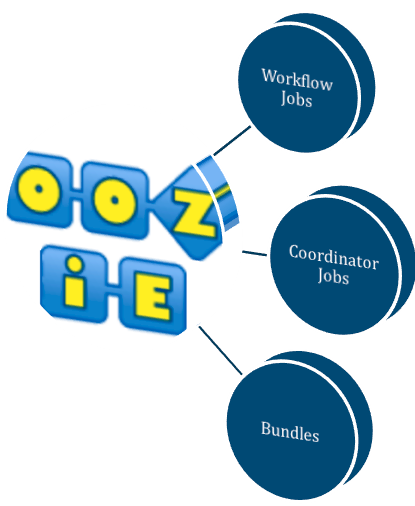
OOZIE workflow – Scheduling tool

**Architecture – Client/Server Models**



1. All the workflow jobs and metadata is present in SQL DB
2. Hue can be used to retrieve the job data from the SQL DB
3. Execute commands which will be sent to server

**Three types of Jobs**



* Workflow jobs – Simple pipeline/parallel job flow
* Coordinator Jobs – Jobs executes based on timing or event trigger
* Bundles – working with a set of coordinators

**Actions vs Control are the important components which are part of the workflow**



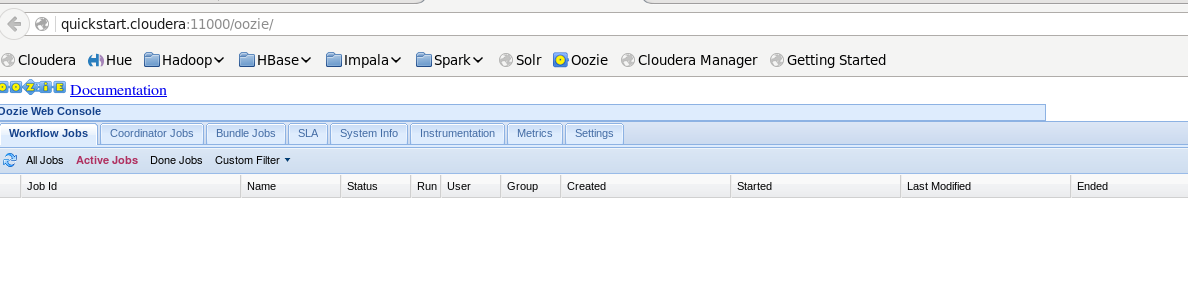
**How to configure a the jobs ?**

* Use the **xml notation** to configure the job or use **Hue** to execute the job

**Check if the OOZIE server is running**



**Oozie Web Client**



**Checking the OOZIE process list to make sure it is running**



* Bootstrap and the QuorumPeerMain (Zoo Keeper) is required for OOZIE to work

**Two mandatory important files for each job/action**

* **workflow.xml**
* **job.proerties**
  + If the job is a mapreduce job, it will have a JAR file
  + If shell script it will have a script file etc

**Structure of job.properties file**

The value of these below properties can be used in the workflow.xml

nameNode = hdfs://rootname

jobTracker = xyz.com:8088

script\_name\_external = hdfs\_path\_of\_script/external.hive

script\_name\_orc=hdfs\_path\_of\_script/orc.hive

script\_name\_copy=hdfs\_path\_of\_script/Copydata.hive

database = database\_name

**Structure of workflow.xml file**