Running an Oozie Workflow

Exercise directory: ~/workspace/oozie-labs

Oozie job folders:

lab1-java-mapreduce

lab2-sort-wordcount

**In this exercise, you will inspect and run Oozie workflows.**

1. Start the Oozie server from Ambari in Firefox.
2. Change directories to the exercise directory:

$ cd ~/workspace/oozie-labs

1. Inspect the contents of the job.properties and workflow.xml files in the lab1-java-mapreduce/job folder. You will see that this is the standard WordCount job.

In the job.properties file, take note of the job’s base directory (lab1-java-mapreduce), and the input and output directories relative to that. (These are HDFS directories.)

1. We have provided a simple shell script to submit the Oozie workflow. Inspect the run.sh script and then run:

$ ./run.sh lab1-java-mapreduce

Notice that Oozie returns a job identification number.

1. Inspect the progress of the job:

$ oozie job -oozie http://localhost:11000/oozie

-info *job\_id*

1. When the job has completed, review the job output directory in HDFS to confirm that the output has been produced as expected.
2. Repeat the above procedure for lab2-sort-wordcount. Notice when you inspect workflow.xml that this workflow includes two MapReduce jobs which run one after the other, in which the output of the first is the input for the second. When you inspect the output in HDFS you will see that the second job sorts the output of the first job into descending numerical order.

**Bonus Exercises**

The exercises in this section are provided as a way to explore topics in further depth than they were covered in classes. You may work on these exercises at your convenience: during class if you have extra time, or after the course is over.

**END**