

coreservlets.com – Hadoop Course

Workflows

In this exercise, you will have a chance to develop MapReduce JobControl workflow.

Approx. Time: 45 minutes

Perform

1. Implement and execute JobControl workflow

`mapRed.workflows.JobControlWorkflow.java`; this workflow is composed of two steps where step 2 uses the output of step 1. `JobControlWorkflow.java` has methods that create Job objects for each MapReduce step in the workflow. Your job is implement JobControl workflow code within run method of `JobControlWorkflow.java`

This workflow can be executed in this manner:

```
$ yarn jar $PLAY_AREA/Exercises.jar mapRed.workflows.JobControlWorkflow \  
  /training/data/hamlet.txt /training/playArea/JobControlWorkflow
```

First step of this workflow calculates number of occurrences for each distinct token. The second step computes the average number of occurrences for each start letter. Your output should look something like this:

```
$ hdfs dfs -cat /training/playArea/JobControlWorkflow/part-r-00000  
"      1  
#      1  
...  
...  
?      2  
A      8  
B      5  
C      3  
...  
...  
Y     17  
z      1
```

Solution

1. The code can be found in the Solutions project:

```
mapRed.workflows.JobControlWorkflow.java
```