# **AirportRank**

Ranking of airports based on PageRank

## **Upload files to Amazon S3**

* Create a bucket in Amazon S3

1. Upload the “airportrank\_2.11-0.1.jar” to Amazon S3.

The jar file is inside the *target/scala-2.11* folder

1. Upload the “airport\_data.csv” file to S3 bucket

*s3://BUCKET/AirportData/airport\_data.csv*

## **Run the jar file in Amazon EMR**

* Create a cluster in Amazon EMR having Spark 2.4.0, and add steps to be run

1. Add step to run class “AirportRank” with arguments ‘input\_file\_location’, ‘iterations’, and ‘output\_directory\_location’

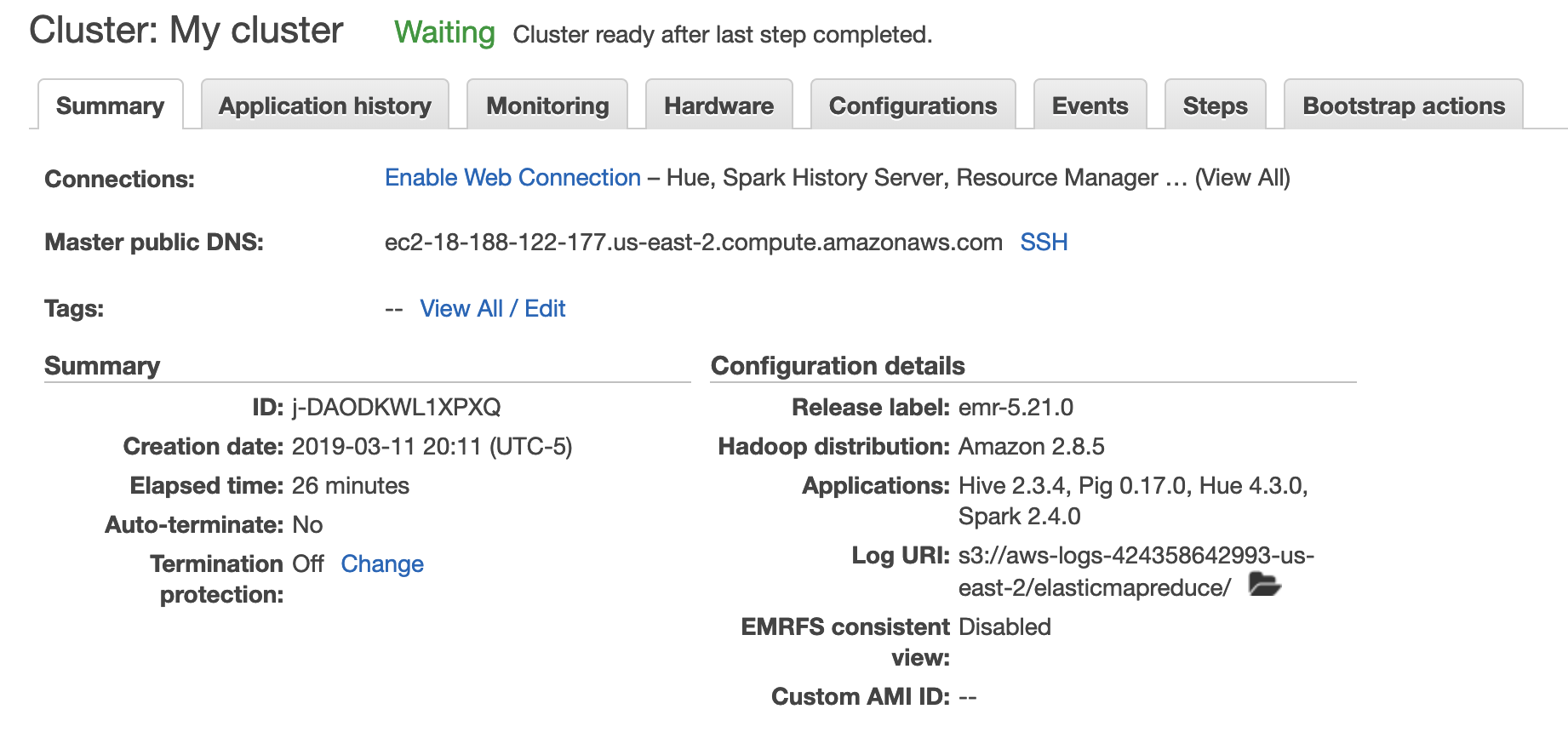
## **Verify the output files**

The output of AirportRank gets stored in specified output folder as well as ‘stdout’

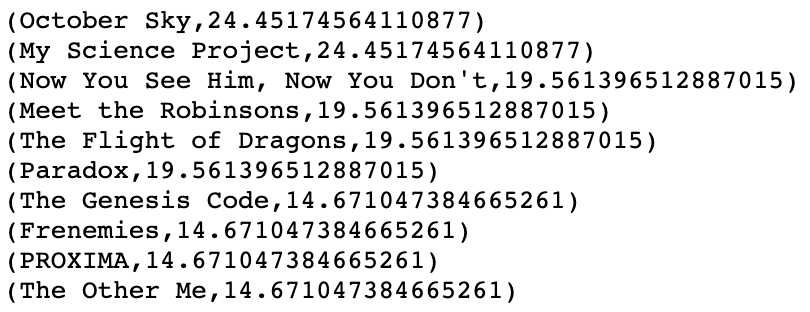
1. The stdout files can be accessed here: *LogURI/containers/application\_id/container\_id/stdout.gz*

*Navigate to Log URI*

* *Navigate to Containers/*
* *Navigate to application\_id/*
* *Navigate to container\_id/*
* *View stdout.gz*



OUTPUT



1. A new folder named output is created under the specified output folder (passed as argument3).

*s3://<bucket>/AirportData/output*  
Inside the output folder are the generated part-0000x files.

