

Spark Cluster on AWS

Software Configuration

Release

<input type="checkbox"/> Hadoop 2.8.5	<input checked="" type="checkbox"/> Zeppelin 0.8.1	<input type="checkbox"/> Livy 0.5.0
<input type="checkbox"/> JupyterHub 0.9.4	<input type="checkbox"/> Tez 0.9.1	<input type="checkbox"/> Flink 1.7.1
<input checked="" type="checkbox"/> Ganglia 3.7.2	<input type="checkbox"/> HBase 1.4.9	<input type="checkbox"/> Pig 0.17.0
<input type="checkbox"/> Hive 2.3.4	<input type="checkbox"/> Presto 0.215	<input type="checkbox"/> ZooKeeper 3.4.13
<input type="checkbox"/> MXNet 1.3.1	<input type="checkbox"/> Sqoop 1.4.7	<input type="checkbox"/> Mahout 0.13.0
<input type="checkbox"/> Hue 4.3.0	<input type="checkbox"/> Phoenix 4.14.1	<input type="checkbox"/> Oozie 5.1.0
<input checked="" type="checkbox"/> Spark 2.4.0	<input type="checkbox"/> HCatalog 2.3.4	<input type="checkbox"/> TensorFlow 1.12.0

Multi-master support
☐ Enable multi-master support

AWS Glue Data Catalog settings (optional)
☐ Use for Spark table metadata

Edit software settings
☒ Enter configuration ☐ Load JSON from S3

`classification=config-file-name,properties=[myKey1=myValue1,myKey2=myValue2]`

Add steps (optional)

Step type **Configure**

☐ Auto-terminate cluster after the last step is completed

1. Launch Hadoop EMR cluster
Go to EMR dashboard and click "Create Cluster".
ClusterName: Adam
Uncheck the "logging" box
Launch mode "Cluster"
Release: 5.8.0
Applications: Spark
Instance Type:
 - Master Node: c4.xlarge instance
 - Core Node: c4.xlarge (8 instances)