AWS Tutorial: Analyze Big Data with Hadoop

Step 1: Set Up Prerequisites for your sample Cluster

Create a bucket

Bucket name ⊤=	Access ⓘ ⊤=	Region ⊤=	Date created ⊤=
🪣 test950607			Nov 11, 2018 5:45:38 PM GMT-0500

Create an Amazon EC2 Key Pair

☑	Key pair name ▲	Fingerprint ▼	
☑	KeyPair1	21:b5:64:97:c7:20:e3:de:df:4c:f2:e8:e5:3e:a3:28:ee:e4:62:c7	

Step 2: Launch your sample Amazon EMR Cluster

Launch the sample Amazon EMR cluster

		Name	ID	Status	Creation time (UTC-5) ▼
☐ ▶ 🟢		My First EMR Cluster	j-3EAE9T1MRJ4L0	Waiting Cluster ready	2018-11-11 18:47 (UTC-5)

Step 3: Allow SSH Connections to the Cluster From your client

SSH	TCP	22	108.45.68.108/32

Step 4: Process Data By Running the Hive Script as a Step

Submit the Hive Script as a Step

	ID	Name	Status	Start time (UTC-5) ▼	Elapsed time	Log files ↗
○ ▶ 🟢	s-277DWCCCPRR32	Hive program	Pending		0 seconds	controller	syslog*	stderr
○ ▶	s-3I3GGR70RKC5F	Setup hadoop debugging	Completed	2018-11-11 18:55 (UTC-5)	2 seconds	View logs

Get result in bucket

	Name ↑☰	Last modified ↑☰	Size ↑☰	Storage class ↑☰
	📂 Test1	--	--	--
	📄 os_requests_$folder$	Nov 11, 2018 8:27:30 PM GMT-0500	0 B	Standard

Viewing 1 to 2

Viewing 1 to 2