

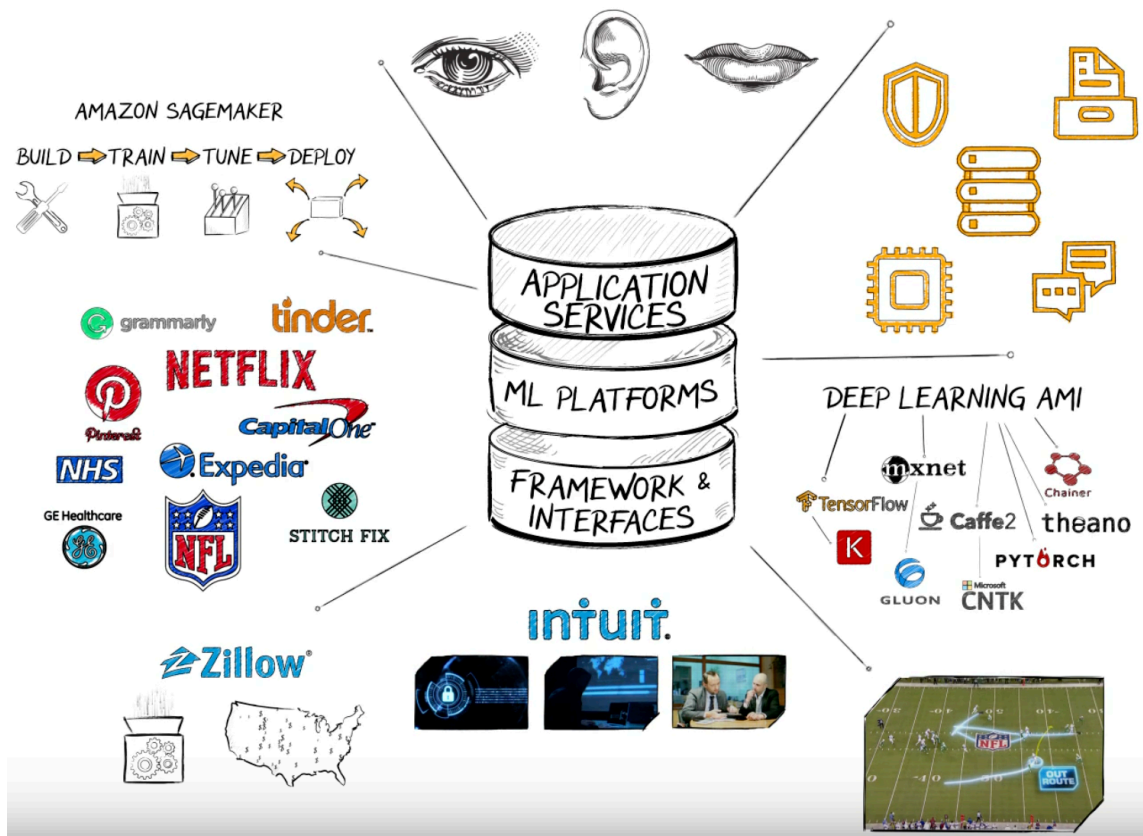
AWS Tutorial: Analyze Big Data with Hadoop

Time spent: 70min

What I learned:

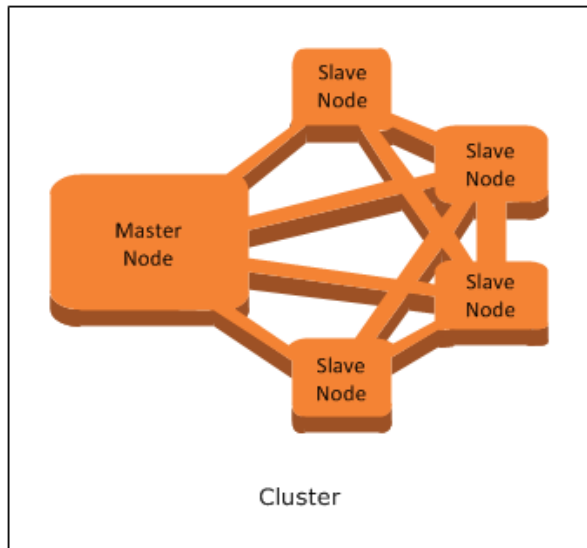
In the overview video, I learned that there was three layers on the AWS ML.

AWS Machine Learning Overview



AWS EMR Overview

The central component of Amazon EMR is the *cluster*. A cluster is a collection of Amazon Elastic Compute Cloud (Amazon EC2) instances. Each instance in the cluster is called a *node*. Each node has a role within the cluster, referred to as the *node type*. Amazon EMR also installs different software components on each node type, giving each node a role in a distributed application like Apache Hadoop.



How to create my cluster

Add SSH Access

aws Services Resource Groups expolee0705

EC2 Dashboard Events Tags Reports Limits INSTANCES Instances Launch Templates Spot Requests Reserved Instances Dedicated Hosts Scheduled Instances Capacity Reservations IMAGES AMIs Bundle Tasks ELASTIC BLOCK STORE Volumes Snapshots Lifecycle Manager NETWORK & SECURITY Security Groups Elastic IPs Placement Groups Key Pairs Network Interfaces

Create Security Group Actions

search : sg-0087d6f2ad15070cb Add filter

Name	Group ID	Group Name	VPC ID	Description
	sg-0087d6f2ad15070cb	ElasticMapReduce-master	vpc-7502aa0f	Master group for Elastic MapReduce created on 2018-12-09T13:26:11.103Z

Edit inbound rules

Type	Protocol	Port Range	Source	Description
All TCP	TCP	0 - 65535	Custom sg-0087d6f2ad15070cb	e.g. SSH for Admin Desktop
All TCP	TCP	0 - 65535	Custom sg-0792d7d141fd13f58	e.g. SSH for Admin Desktop
All UDP	UDP	0 - 65535	Custom sg-0087d6f2ad15070cb	e.g. SSH for Admin Desktop
All UDP	UDP	0 - 65535	Custom sg-0792d7d141fd13f58	e.g. SSH for Admin Desktop
All ICMP - IPv	ICMP	0 - 65535	Custom sg-0087d6f2ad15070cb	e.g. SSH for Admin Desktop
All ICMP - IPv	ICMP	0 - 65535	Custom sg-0792d7d141fd13f58	e.g. SSH for Admin Desktop
SMTP	TCP	25	My IP 128.164.212.97/32	e.g. SSH for Admin Desktop

Add Rule

NOTE: Any edits made on existing rules will result in the edited rule being deleted and a new rule created with the new details. This will cause traffic that depends on that rule to be dropped for a very brief period of time until the new rule can be created.

Cancel Save

Start the script by adding steps

Clone

Terminate

AWS CLI export

Cluster: My First EMR Cluster Waiting Cluster ready after last step completed.

Summary

Application history

Monitoring

Hardware

Events

Steps

Configurations

Bootstrap actions

Add step

Clone step

Cancel step

Steps

Filter: All steps Filter steps ... 2 steps (all loaded)

	ID	Name	Status	Start time (UTC-5)	Elapsed time	Log files	Actions
	s-2WPNWJIUZUQ44	ShiboHive	Pending		0 seconds	View logs	View jobs

JAR location : command-runner.jar
Main class : None
hive-script --run-hive-script --args -f s3://us-east-1.elasticmapreduce.samples/cloudfront/code/Hive_CloudFront.q -d INPUT=s3://us-east-1.elasticmapreduce.samples -d
Arguments : OUTPUT=s3://bigdata0705/HiveQueryResults/
Action on failure: Continue

Clone

Terminate

AWS CLI export

Cluster: My First EMR Cluster Starting

Summary

Application history

Monitoring

Hardware

Events

Steps

Configurations

Bootstrap actions

Connections: --
Master public DNS: --
Tags: -- [View All / Edit](#)

Summary

Configuration details

Network and hardware

Key name: shibokeypair
EC2 instance profile: EMR_EC2_DefaultRole
EMR role: EMR_DefaultRole
Visible to all users: All [Change](#)
Security groups for Master:
Security groups for Core & Task:

Release label: emr-5.19.0
Hadoop distribution: Amazon 2.8.5
Applications: Ganglia 3.7.2, Hive 2.3.3, Hue 4.2.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.8.4
Log URI: s3://aws-logs-238665400611-us-east-1/elasticmapreduce/
EMRFS consistent view: Disabled
Custom AMI ID: --

Availability zone: --
Subnet ID: [subnet-b64bd098](#)
Master: Provisioning 1 m4.large
Core: Provisioning 2 m4.large
Task: --

Get the result in S3

[Amazon S3](#) > [bigdata0705](#) / [HiveQueryResults](#) / [os_requests](#)

Overview

 Type a prefix and press Enter to search. Press ESC to clear.



Upload



Create folder

Download

Actions 




Name 



 000000_0



000000_0 (1) 

Android855
Linux813
MacOS852
OSX799
Windows883
iOS794

Summary

In this tutorial, I learned the mechanism and structure of AWS EMR. With the help of it, I finished the data process by the Hadoop.