Installing Your First Big Data Cluster Using Cloudera CDH

Xavier Morera

PASSIONATE ABOUT TEACHING

@xmorera   www.xaviermorera.com

The Three Cloudera Installation Paths

A

B

C

Path A

Path C

Path C

The Three Cloudera Installation Paths

A

B

B

C

Path A

Path B

Path B

Path C

Installation Path A

Automated installation

Using Cloudera Manager
- Cloudera Manager Server
- JDK, Agents, CDH and services

Automatic embedded PostgreSQL

Mainly for POCs, dev and demos

Not intended for production
- Scaling limitations
- May require database migration

Installation Path B

B

Automated or manual installation

Parcels or Packages

Puppet or Chef to help install across cluster

Production and Non-Production

Non-Production
- Embedded PostgreSQL

Production
- Manual database installation & configuration

Installation Path C

Manual installation
- JDK, Cloudera Manager Server & Agents
- Tarballs used for installation

Use Cloudera Manager to install and deploy
- CDH
- Managed software services (parcels)

Next Step?

```
wget https://archive.cloudera.com/
  cm5/installer/latest/cloudera-manager-installer.bin
```

Getting the Installation File

1

Latest version

Find the one that you need

```
chmod u+x cloudera-manager-installer.bin
```

Execute Permissions

2

chmod gives execute permissions to the user

.bin is a self extracting binary file

Execute this file to install Cloudera Manager

```
sudo ./cloudera-manager-installer.bin
```

Install Cloudera Manager

3

Execute .bin file which will start the process

Notice the ./ and sudo is required
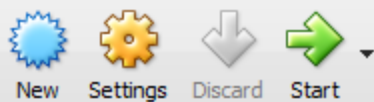
Use --skip_repo_package=1 for local package installation

Demo

**Exercise 7
Installing Cloudera Manager
using Path A**

File   Machine   Help

New   Settings   Discard   Start

Details   Snapshots (1)

cloudera-quickstart-v...
Saved

Base Linux Machine (F...)
Powered Off

mgt01 (mgt01 clean)
Powered Off

dn01 (clean dn01)
Powered Off

dn02 (clean dn02)
Powered Off

dn03 (clean dn03)
Powered Off

dn04 (clean dn04)
Powered Off

General

Name: mgt01
Operating System: Red Hat (64-bit)

System

Base Memory: 2048 MB
Processors: 2
Boot Order: Floppy, Optical, Hard Disk
Acceleration: VT-x/AMD-V, Nested Paging, PAE/NX, KVM Paravirtualization

Preview

mgt01

Display

Video Memory: 16 MB
Remote Desktop Server: Disabled
Video Capture: Disabled

Storage

Controller: IDE
  IDE Primary Master: [Optical Drive] Empty
  IDE Secondary Master: [Optical Drive] Empty
Controller: SATA
  SATA Port 0:   mgt01.vdi (Normal, 80.00 GB)

Audio

Host Driver: Windows DirectSound
Controller: ICH AC97

Network

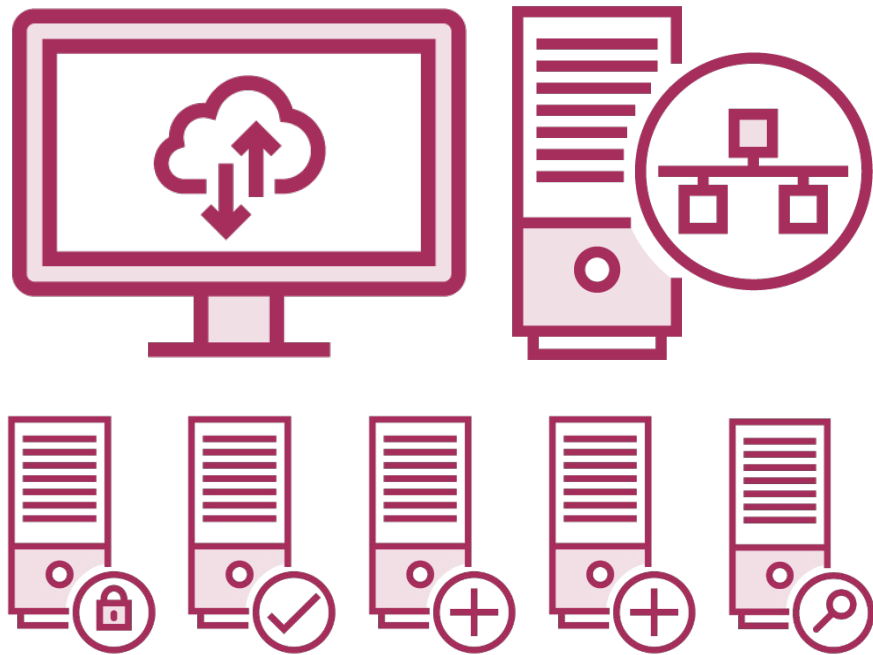Adapter 1:   Intel PRO/1000 MT Desktop (NAT Network, 'PSNATNetwork')

USB

Shared folders

None

Description

Installation Step Two: Deploying Your Cluster

Cloudera Manager is installed

Deploy agents to all nodes in our cluster

Automated installation
- But you need to make a few decision

Cloudera Manager provides
- Easy to understand interface
- Detailed log during installation

What's left?

Steps

Select Cloudera Manager version

Specify hosts

Select repository

Accept licenses for JDK installation

Single user mode

Login credentials

Simultaneous installations

Steps

Parcels: downloaded, distributed, unpacked and activated

Inspect hosts for correctness

Select combination of services to install

Select nodes for services

Database setup

Configuration parameters

Cluster setup complete!

Xavier

localhost:7180/cmf/login

cloudera MANAGER

admin

•••••

Log In

☐ Remember me

Takeaway

Set up a Hadoop cluster
- 5 nodes

CDH
- Cloudera Distribution including Hadoop

Cloudera Manager

Automated install
- Path A
- Multiple steps
- Embedded database
- Non prod but works great for dev

Great first step to Big Data