

Final Takeaway



Xavier Morera

PASSIONATE ABOUT TEACHING

@xmorera www.xaviermorera.com





Great Job! First **Cloudera** Cluster

Big Data, Hadoop & Cloudera



Digital revolution

Information explosion

Data has valuable applications

Store, process and analyze

Limitations with existing technologies

Big Data with Hadoop



Big Data, Hadoop & Cloudera



Hadoop is a complex ecosystem

Cloudera comes to fill this gap

CDH

- **Cloudera Distribution including Hadoop**

Hadoop for the masses

New competitive advantage



Cloudera QuickStart VM



Everything you need to get started

Pseudo-cluster

- Hadoop, Hive, Pig, Spark, Solr, ...

Cloudera Manager

- Express, Enterprise Data Hub, Enterprise

What's next? A real cluster

- On-premises or cloud
- Cloud? → Cloudera Director
- Requirements



Linux Machines Cluster



Create a base cluster

Linux

- CentOS

VirtualBox

Install from scratch

- ISO → Install → Configure → Clone
- NAT Network and final configurations

We have a clean cluster



First Big Data Cluster with CDH



Installation Paths

- A, B and C
- Production and non production

Path A

- Cloudera Manager Server
- Deploy agents to hosts
- Install CDH
- Configure
- Select services and roles

Your First Big Data Hadoop Cluster Using Cloudera CDH





Looking Ahead

Big Data is the future (and the present)

Hadoop is a huge ecosystem

Pick your battles

- Hadoop Admin
- Data Engineer
- Data Analyst & Scientist

Many alternatives

Cloudera is a top choice

