



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



**UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH**



AI and Predictive Analytics in Data-Center Environments

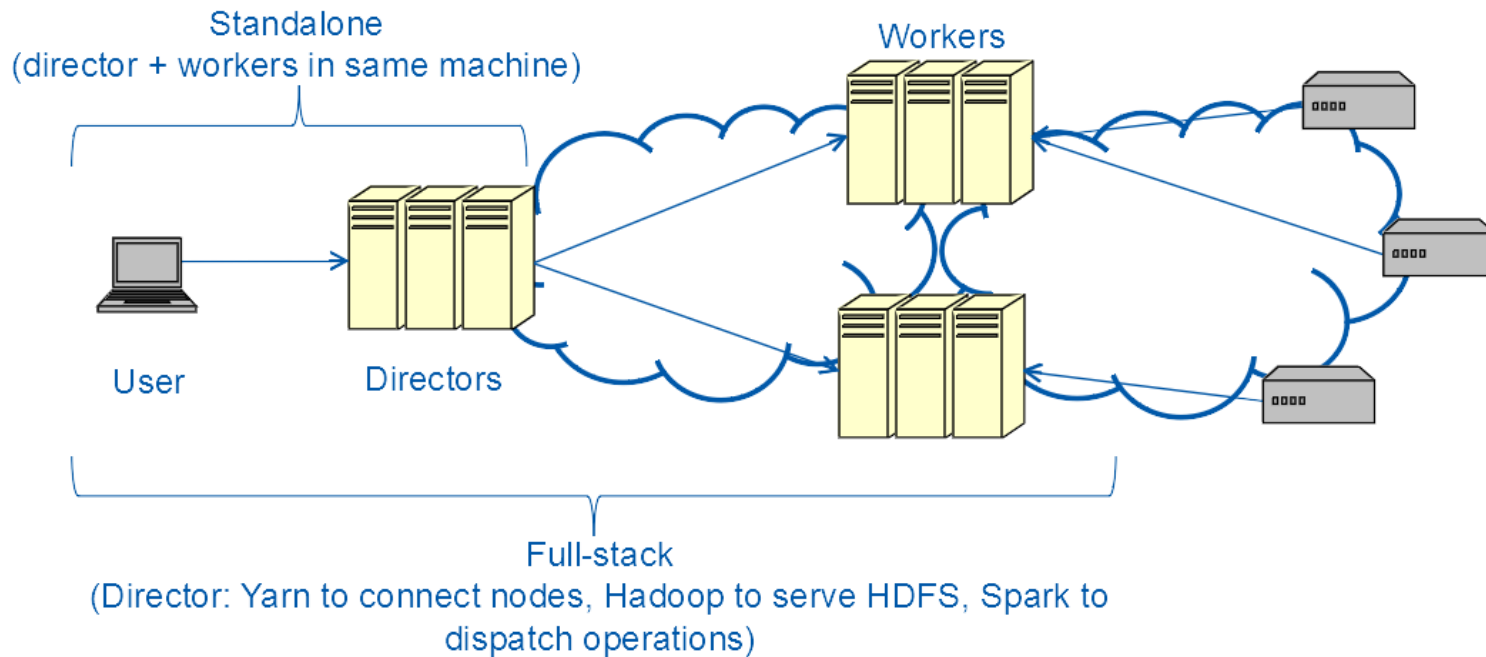
Distributed Computing using Spark

Installing and Running Spark

Hands-On: Installing Spark

- Get a console:
 - We're downloading and setting up Apache Spark

Hands-On: Setting up the Environment



- What we will do now:
 - The “Standalone mode” (single machine)
 - An example of the “Cluster Mode”

Hands-On: Installing Spark

- Our environment:
 - We're working on a Unix Terminal
 - We'll be using Python
- Next steps: Install the Environment and Spark
 - JAVA and OpenSSH → Runtime and Tools
 - Apache Hadoop – HDFS → Distributed File-System
 - Apache Spark → Spark framework

Summary

- What is Spark & Hadoop
 - Distributed Computing Framework
- Installing Spark
 - Dependencies
 - Also Hadoop DFS
- Leverage Virtualization
 - VirtualBox
 - Images at Vagrant repositories