

# Installing Hadoop

---

# Overview

**Understanding the different install modes available for Hadoop**

**Installing Hadoop in Standalone mode**

**Installing Hadoop in Pseudo-Distributed mode**

# Hadoop Install Modes



**Standalone**

**Pseudo-Distributed**

**Fully Distributed**

# Hadoop Install Modes



The diagram consists of three rectangular boxes arranged horizontally. The first box on the left is dark red and contains the text 'Standalone'. The middle box is light purple and contains the text 'Pseudo-Distributed'. The third box on the right is light blue and contains the text 'Fully Distributed'. All three boxes are of equal height and width, and they are separated by small gaps.

**Standalone**

**Pseudo-Distributed**

**Fully Distributed**

**Standalone**

**The default mode in which  
Hadoop runs**

**Standalone**

**Runs on a single node**

**A single JVM process**

**Local File System for Storage**

**HDFS and YARN do not run**

**Standalone**

**Used to test MapReduce  
programs before running  
them on a cluster**

# Hadoop Install Modes



The diagram consists of three vertical rectangular boxes arranged horizontally. The first box on the left is dark red and contains the text 'Standalone'. The middle box is light purple and contains the text 'Pseudo-Distributed'. The third box on the right is light blue and contains the text 'Fully Distributed'. All text is in a white, sans-serif font and is centered within each box.

**Standalone**

**Pseudo-Distributed**

**Fully Distributed**



# Hadoop Install Modes



Standalone

Pseudo-Distributed

Fully Distributed

## Pseudo-Distributed

**Runs on a single node**

**2 JVM processes to simulate 2 nodes**

**HDFS for storage**

**YARN for managing tasks**

Pseudo-Distributed

**Used as a fully-fledged test  
environment**

# Hadoop Install Modes



Standalone

Pseudo-Distributed

Fully Distributed

# Hadoop Install Modes

Standalone

Pseudo-Distributed

**Fully Distributed**

**Fully Distributed**

## **Runs on a cluster of machines**

- Linux servers in a data center
- VMs requisitioned on a cloud service

**Fully Distributed**

**Manual configuration of a cluster is complicated**

**Usually use enterprise editions**

- Cloudera, MapR, Hortonworks

# Demo

## **Installing Hadoop in Standalone Mode**

- Downloading and unpacking Hadoop
- Running a test MapReduce



# Demo

## **Installing Hadoop in Pseudo-Distributed Mode**

- Setting up Hadoop configuration files
- Setting up an SSH key
- Launching HDFS and YARN processes
- Running a test MapReduce
- Using the UI to monitor jobs

# Summary

**Understood the different install modes available for Hadoop**

**Installed Hadoop in Standalone mode**

**Installed Hadoop in Pseudo-Distributed mode**