

# Introducció a Hadoop i Big Data amb Cloudera

16 de juny de 2016

## Resum

L'entorn Big Data està evolucionant contínuament a mesura que sorgeixen noves tecnologies i les tecnologies existents maduren. Aquest és un curs que cobreix la arquitectura i l'ecosistema de dades Hadoop. Aquestes tecnologies són la base del moviment Big Data, facilitant una gestió i administració escalable pel processament de gran quantitats de dades.

Els alumnes que completin el curs entendran l'arquitectura dels clústers Hadoop, tan a nivell de hardware com software. Els alumnes aprendran a aplicar Hadoop i les tecnologies Big Data relacionades, com MapReduce, Hive, Impala i Pig en el desenvolupament de l'anàlisi i la resolució dels problemes als que s'enfronten les empreses avui en dia.

Adreçat a docents de CFGS que imparteixin assignatures relacionades amb Big Data i vulguin traslladar la formació en tecnologia Cloudera als seus estudiants a partir del projecte Cloudera Academic.

Requisits previs: Formació en Sistemes Operatius, experiència programant en Java, Python o C/C++ per assignacions: i un coneixement general de les xarxes i els sistemes distribuïts. Es recomana estar familiaritzat amb Linux i Bases de Dades.

Eines: Màquina virtual Cloudera Hadoop configurada completament, incloent conjunts de dades per pràctiques de laboratori.

## Índex

<b>1</b>	<b>Conceptes bàsics de Hadoop</b>	<b>2</b>
<b>2</b>	<b>Ecosistema Hadoop. Introducció a MapReduce</b>	<b>2</b>
<b>3</b>	<b>Programant MapReduce en Java</b>	<b>2</b>
<b>4</b>	<b>Api Hadoop</b>	<b>2</b>
<b>5</b>	<b>Divisors i reductors. Inputs i outputs de Java</b>	<b>2</b>
<b>6</b>	<b>Algoritmes comuns de MapReduce</b>	<b>2</b>
<b>7</b>	<b>Eines Hadoop per l'adquisició de dades</b>	<b>2</b>
<b>8</b>	<b>Introducció a Pig</b>	<b>2</b>
<b>9</b>	<b>Introducció a Hive</b>	<b>2</b>

<b>10</b>	<b>Processament de textos amb Hive</b>	<b>3</b>
<b>11</b>	<b>Introducció a Impala</b>	<b>3</b>
<b>12</b>	<b>Sobre aquest document</b>	<b>3</b>

## **1 Conceptes bàsics de Hadoop**

i++i

## **2 Ecosistema Hadoop. Introducció a MarReduce**

i++i

## **3 Programant MapReduce en Java**

i++i

## **4 Api Hadoop**

i++i

## **5 Divisors i reductors. Inputs i outputs de Java**

i++i

## **6 Algoritmes comuns de MapReduce**

i++i

## **7 Eines Hadoop per l'adquisició de dades**

i++i

## **8 Introducció a Pig**

i++i

## **9 Introducció a Hive**

i++i

## 10 Processament de textos amb Hive

i++i

## 11 Introducció a Impala

i++i

## 12 Conceptes relacionats

### 12.1 Computació distribuïda

i++i

### 12.2 Middleware: SOAP i RPC

i++i

### 12.3 Model de programació MapReduce

i++i

## 13 Sobre aquest document

**Copyright© 2016 Juan Aguilera.**

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation;

with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled “GNU Free Documentation License”.

## Referències

[1]