

1. Record Nr.	UNINA9910464753703321
Autore	Menon Rohit
Titolo	Cloudera administration handbook : a complete, hands-on guide to building and maintaining large Apache Hadoop clusters using Cloudera Manager and CDH5 / / Rohit Menon ; cover image by John Michael Harkness
Pubbl/distr/stampa	Birmingham, England : , : Packt Publishing Ltd, , 2014 ©2014
ISBN	1-78355-897-0
Edizione	[1st edition]
Descrizione fisica	1 online resource (254 p.)
Collana	Community Experience Distilled
Disciplina	004.36
Soggetti	Electronic data processing - Distributed processing File organization (Computer science) - Computer programs Electronic books.
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Cover; Copyright; Credits; About the Author; About the Reviewers; www.PacktPub.com; Table of Contents; Preface; Chapter 1: Getting Started with Apache Hadoop; History of Apache Hadoop and its trends; Components of Apache Hadoop; Understanding the Apache Hadoop daemons; Namenode; Secondary namenode; Jobtracker; Tasktracker; Resource Manager; NodeManager; Job submission in YARN; Introducing Cloudera; Introducing CDH; Responsibilities of a Hadoop administrator; Summary; Chapter 2: HDFS and Map Reduce; Essentials of HDFS; Configuring HDFS; The read/write operational flow in HDFS Writing files in HDFS Reading files in HDFS; Understanding the name node UI; Understanding the secondary namenode UI; Exploring HDFS commands; Commonly used HDFS commands; Commands to administer HDFS; Getting acquainted with MapReduce; Understanding the map phase; Understanding the reduce phase; Learning all about the MapReduce job flow; Configuring MapReduce; Understanding the jobtracker UI; Getting MapReduce job information; Summary; Chapter 3: Cloudera's Distribution Including Apache Hadoop - CDH; Getting started with CDH; Understanding the CDH components; Apache Hadoop; Apache Flume NG

Apache Sqoop Apache Pig; Apache Hive; Apache ZooKeeper; Apache HBase; Apache Whirr; Snappy - previously known as Zippy; Apache Mahout; Apache Avro; Apache Oozie; Cloudera Search; Cloudera Impala; Cloudera Hue; Beeswax - Hive UI; Cloudera Impala UI; Pig UI; File Browser; Metastore Manager; Sqoop Jobs; Job Browser; Job Designs; Dashboard; Collection Manager; Hue Shell; HBase Browser; Installing CDH; Stopping Hadoop services; Understanding a YARN cluster; Installing the CDH components; Installing Apache Flume; Installing Apache Sqoop; Installing Apache Sqoop 2; Installing Apache Pig Installing Apache Hive Installing Apache Oozie; Installing Apache ZooKeeper; Summary; Chapter 4: Exploring HDFS Federation and Its High Availability; Implementing HDFS Federation; Configuring HDFS Federation; Configuring ViewFS for federated HDFS; Implementing HDFS High Availability; Quorum-based storage; Configuring HDFS high availability by Quorum-based storage; Shared storage using NFS; Configuring HDFS high availability by shared storage sing NFS; Configuring automatic fail over for HDFS high availability; Jobtracker high availability; Configuring Jobtracker High Availability Configuring automatic fail over for Job tracker high availability Summary; Chapter 5: Using Cloudera Manager; Introducing Cloudera Manager; Understanding the Cloudera Manager architecture; Installing Cloudera Manager; Navigating the Cloudera Manager Web console; Navigating the Home screen; Navigating the Clusters menu; Exploring the Hosts menu; Understanding the Diagnostics menu; Understanding the Audits screen; Understanding the Charts menu; Understanding the Backup menu; Understanding the Administration menu; Configuring High Availability using Cloudera Manager; Summary Chapter 6: Implementing Security Using Kerberos

Sommario/riassunto

An easy-to-follow Apache Hadoop administrator's guide filled with practical screenshots and explanations for each step and configuration. This book is great for administrators interested in setting up and managing a large Hadoop cluster. If you are an administrator, or want to be an administrator, and you are ready to build and maintain a production-level cluster running CDH5, then this book is for you.
