1. Record Nr. UNINA9910823169303321

Autore	Jorgensen Adam
Titolo	Microsoft big data solutions / / Adam Jorgensen [and five others] ; executive editor, Robert Elliot ; project editor, Jennifer Lynn ; cover designer, Ryan Sneed
Pubbl/distr/stampa	Indianapolis, Indiana : , : Wiley, , 2014 ©2014
ISBN	1-118-72955-2 1-118-74209-5
Edizione	[1st edition]
Descrizione fisica	1 online resource (410 p.)
Altri autori (Persone)	JorgensenAdam ElliotRobert LynnJennifer SneedRyan
Disciplina	005.74
Soggetti	Cloud computing Computers Web services
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Cover; Title Page; Copyright; Contents; Introduction; Part I What Is Big Data?; Chapter 1 Industry Needs and Solutions; What's So Big About Big Data?; A Brief History of Hadoop; Google; Nutch; What Is Hadoop?; Derivative Works and Distributions; Hadoop Distributions; Core Hadoop Ecosystem; Important Apache Projects for Hadoop; The Future for Hadoop; Summary; Chapter 2 Microsoft's Approach to Big Data; A Story of "Better Together"; Competition in the Ecosystem; SQL on Hadoop Today; Hortonworks and Stinger; Cloudera and Impala; Microsoft's Contribution to SQL in Hadoop; Deploying Hadoop Deployment Factors Deployment Topologies; Deployment Scorecard; Summary; Part II Setting Up for Big Data with Microsoft; Chapter 3 Configuring Your First Big Data Environment; Getting Started; Getting the Install; Running the Installation; On-Premise Installation: Single-Node Installation; HD Insight Service: Installing in the Cloud; Windows Azure Storage Explorer Options; Validating Your New Cluster; Logging

into HD Insight Service; Verify HDP Functionality in the Logs; Common Post-Setup Tasks; Loading Your First Files; Verifying Hive and Pig; Summary; Part III Storing and Managing Big Data
Chapter 4 HDFS, Hive, HBase, and HCatalog Exploring the Hadoop Distributed File System; Explaining the HDFS Architecture; Interacting with HDFS; Exploring Hive: The Hadoop Data Warehouse Platform; Designing, Building, and Loading Tables; Querying Data; Configuring the Hive ODBC Driver; Exploring HCatalog: HDFS Table and Metadata Management; Exploring HBase: An HDFS Column-Oriented Database; Columnar Databases; Defining and Populating an HBase Table; Using Query Operations; Summary; Chapter 5 Storing and Managing Data in HDFS; Understanding the Fundamentals of HDFS; HDFS Architecture Name Nodes and Data Nodes Data Replication; Using Common Commands to Interact with HDFS; Interfaces for Working with HDFS; File Manipulation Commands; Administrative Functions in HDFS; Moving and Organizing Data in HDFS; Moving Data in HDFS; Implementing Data Structures for Easier Management; Rebalancing Data; Summary; Chapter 6 Adding Structure with Hive; Understanding Hive's Purpose and Role; Providing Structure for Unstructured Data; Enabling Data Access and Transformation; Differentiating Hive from Traditional RDBMS Systems; Working with Hive; Creating and Querying Basic Tables Creating Databases Creating Tables; Adding and Deleting Data; Querying a Table; Using Advanced Data Structures with Hive; Setting Up Partitioned Tables; Loading Partitioned Tables; Using Views; Creating Indexes for Tables; Summary; Chapter 7 Expanding Your Capability with HBase and HCatalog; Using HBase; Creating HBase Tables; Loading Data into an HBase Table; Performing a Fast Lookup; Loading and Querying HBase; Managing Data with HCatalog; Working with HCatalog and Hive; Defining Data Structures; Creating Indexes; Creating Partitions; Integrating HCatalog with Pig and Hive
Using HBase or Hive as a Data Warehouse

Sommario/riassunto

Tap the power of Big Data with Microsoft technologies Big Data is here, and Microsoft's new Big Data platform is a valuable tool to help your company get the very most out of it. This timely book shows you how to use HD Insight along with Horton Works Data Platform for Windows to store, manage, analyze, and share Big Data throughout the enterprise. Focusing primarily on Microsoft and Horton Works technologies but also covering open source tools, Microsoft Big Data Solutions explains best practices, covers on-premises and cloud-based solutions, and features valuable case studies