

Instant Apache Hive Essentials How To

Wyatt Earp, the Gunfight near the O.K. Corral, and related events happened in the larger context of an area in turmoil. The Tombstone Mining District was a fluid area, filled with hard working settlers, some of the nations best mining engineers, and some of the nations worst criminals. The Tombstone saga and its greatest conflicts were linked to incidents outside of Tombstone. A shooting near Drew's Station, drew Wyatt Earp more closely into conflict with the local outlaw element. This in turn would

Download File PDF Instant Apache Hive Essentials How To

set in motion a chain of events that would in part lead to the West's most famous gunfight, the Gunfight near the O.K. Corral. The Earps were brought into court at Contention City over the shootings in Tombstone. The all important history of Drew's Station, Contention City, and Fairbank, Arizona Territory are now revealed for the first time ever. Heavily documented and thoroughly researched, many misconceptions about this portion of the wild west are brought to light, and a great deal of previously unpublished research is now available for the first time, bringing to light the key stories of these key locations along the road to Tombstone.

Download File PDF Instant Apache Hive Essentials How To

Describes the features and functions of Apache Hive, the data infrastructure for Hadoop.

Easy, hands-on recipes to help you understand Hive and its integration with frameworks that are used widely in today's big data world

About This Book Grasp a complete reference of different Hive topics.

Get to know the latest recipes in development in Hive including

CRUD operations Understand Hive internals and integration of Hive

with different frameworks used in today's world. Who This Book Is For

The book is intended for those who want to start in Hive or who have basic understanding of Hive

framework. Prior knowledge of basic SQL command is also required

Download File PDF Instant Apache Hive Essentials How To

What You Will Learn Learn different features and offering on the latest Hive Understand the working and structure of the Hive internals Get an insight on the latest development in Hive framework Grasp the concepts of Hive Data Model Master the key concepts like Partition, Buckets and Statistics Know how to integrate Hive with other frameworks such as Spark, Accumulo, etc In Detail Hive was developed by Facebook and later open sourced in Apache community. Hive provides SQL like interface to run queries on Big Data frameworks. Hive provides SQL like syntax also called as HiveQL that includes all SQL capabilities like analytical functions which are the need of the

Download File PDF Instant Apache Hive Essentials How To

hour in today's Big Data world. This book provides you easy installation steps with different types of metastores supported by Hive. This book has simple and easy to learn recipes for configuring Hive clients and services. You would also learn different Hive optimizations including Partitions and Bucketing. The book also covers the source code explanation of latest Hive version. Hive Query Language is being used by other frameworks including spark. Towards the end you will cover integration of Hive with these frameworks. Style and approach Starting with the basics and covering the core concepts with the practical usage, this book is a complete guide to learn and explore Hive offerings.

Download File PDF Instant Apache Hive Essentials How To

How do you start? How should you build a plan for cloud migration for your entire portfolio? How will your organization be affected by these changes? This book, based on real-world cloud experiences by enterprise IT teams, seeks to provide the answers to these questions. Here, you'll see what makes the cloud so compelling to enterprises; with which applications you should start your cloud journey; how your organization will change, and how skill sets will evolve; how to measure progress; how to think about security, compliance, and business buy-in; and how to exploit the ever-growing feature set that the cloud offers to gain strategic and competitive advantage.

Download File PDF Instant Apache Hive Essentials How To

Or, Life in Dakota with General
Custer

Big Data in Practice

A Brief Guide to the Emerging
World of Polyglot Persistence
Programming Hive

Cloudera Administration Handbook

How 45 Successful Companies Used
Big Data Analytics to Deliver
Extraordinary Results

"Apache Hadoop is helping
drive the Big Data revolution.

Now, its data processing has
been completely overhauled:

Apache Hadoop YARN

provides resource

management at data center

scale and easier ways to

create distributed applications

Download File PDF Instant Apache Hive Essentials How To

that process petabytes of data. And now in Apache Hadoop™ YARN, two Hadoop technical leaders show you how to develop new applications and adapt existing code to fully leverage these revolutionary advances." -- From the Amazon

Set in the post-Civil War Arizona "An Apache Princess" tells the tale of the frontier and the life of Native Americans. It follows the story of two young girls, one being the daughter of the Captain in Arizona fort, the other being the Apache girl,

Download File PDF Instant Apache Hive Essentials How To

as they go through numerous adventures in the Wild West. This book takes you on a fantastic journey to discover the attributes of big data using Apache Hive. Key Features Grasp the skills needed to write efficient Hive queries to analyze the Big Data Discover how Hive can coexist and work with other tools within the Hadoop ecosystem Uses practical, example-oriented scenarios to cover all the newly released features of Apache Hive 2.3.3 Book Description In this book, we prepare you for your journey into big data

Download File PDF Instant Apache Hive Essentials How To

by firstly introducing you to backgrounds in the big data domain, alongwith the process of setting up and getting familiar with your Hive working environment. Next, the book guides you through discovering and transforming the values of big data with the help of examples. It also hones your skills in using the Hive language in an efficient manner. Toward the end, the book focuses on advanced topics, such as performance, security, and extensions in Hive, which will guide you on exciting adventures on this worthwhile big data journey.

Download File PDF Instant Apache Hive Essentials How To

By the end of the book, you will be familiar with Hive and able to work effeciently to find solutions to big data problems What you will learn Create and set up the Hive environment Discover how to use Hive's definition language to describe data Discover interesting data by joining and filtering datasets in Hive Transform data by using Hive sorting, ordering, and functions Aggregate and sample data in different ways Boost Hive query performance and enhance data security in Hive Customize Hive to your needs

Download File PDF Instant Apache Hive Essentials How To

by using user-defined functions and integrate it with other tools Who this book is for If you are a data analyst, developer, or simply someone who wants to quickly get started with Hive to explore and analyze Big Data in Hadoop, this is the book for you. Since Hive is an SQL-like language, some previous experience with SQL will be useful to get the most out of this book.

If you are a system or application developer interested in learning how to solve practical problems using the Hadoop framework,

Download File PDF Instant Apache Hive Essentials How To

then this book is ideal for you. This book is also meant for Hadoop professionals who want to find solutions to the different challenges they come across in their Hadoop projects.

Microsoft Azure Essentials -
Fundamentals of Azure
Up and Running

Apache Hive Third Edition
On the Road to Tombstone

A Guide to Hadoop's Data
Warehouse System

Learn the Essentials of Big
Data Computing in the Apache
Hadoop 2 Ecosystem

An easy-to-follow Apache Hadoop
administrator's guide filled with

Download File PDF Instant Apache Hive Essentials How To

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.

The Complete Guide to Data Science with Hadoop—For Technical Professionals, Businesspeople, and Students
Demand is soaring for professionals who can solve real data science problems with Hadoop and Spark. Practical Data Science with Hadoop® and Spark

Download File PDF Instant Apache Hive Essentials How To

is your complete guide to doing just that. Drawing on immense experience with Hadoop and big data, three leading experts bring together everything you need: high-level concepts, deep-dive techniques, real-world use cases, practical applications, and hands-on tutorials. The authors introduce the essentials of data science and the modern Hadoop ecosystem, explaining how Hadoop and Spark have evolved into an effective platform for solving data science problems at scale. In addition to comprehensive application coverage, the authors also provide useful guidance on the important steps of data ingestion, data

Download File PDF Instant Apache Hive Essentials How To

munging, and visualization. Once the groundwork is in place, the authors focus on specific applications, including machine learning, predictive modeling for sentiment analysis, clustering for document analysis, anomaly detection, and natural language processing (NLP). This guide provides a strong technical foundation for those who want to do practical data science, and also presents business-driven guidance on how to apply Hadoop and Spark to optimize ROI of data science initiatives. Learn What data science is, how it has evolved, and how to plan a data science career How data volume, variety, and velocity shape data

Download File PDF Instant Apache Hive Essentials How To

science use cases Hadoop and its ecosystem, including HDFS, MapReduce, YARN, and Spark Data importation with Hive and Spark Data quality, preprocessing, preparation, and modeling Visualization: surfacing insights from huge data sets Machine learning: classification, regression, clustering, and anomaly detection Algorithms and Hadoop tools for predictive modeling Cluster analysis and similarity functions Large-scale anomaly detection NLP: applying data science to human language The need to handle increasingly larger data volumes is one factor driving the adoption of a new class of nonrelational “NoSQL”

Download File PDF Instant Apache Hive Essentials How To

databases. Advocates of NoSQL databases claim they can be used to build systems that are more performant, scale better, and are easier to program. NoSQL Distilled is a concise but thorough introduction to this rapidly emerging technology. Pramod J. Sadalage and Martin Fowler explain how NoSQL databases work and the ways that they may be a superior alternative to a traditional RDBMS. The authors provide a fast-paced guide to the concepts you need to know in order to evaluate whether NoSQL databases are right for your needs and, if so, which technologies you should explore further. The first part of the book concentrates on

Download File PDF Instant Apache Hive Essentials How To

core concepts, including schemaless data models, aggregates, new distribution models, the CAP theorem, and map-reduce. In the second part, the authors explore architectural and design issues associated with implementing NoSQL. They also present realistic use cases that demonstrate NoSQL databases at work and feature representative examples using Riak, MongoDB, Cassandra, and Neo4j. In addition, by drawing on Pramod Sadalage's pioneering work, NoSQL Distilled shows how to implement evolutionary design with schema migration: an essential technique for applying NoSQL databases. The book concludes by describing

Download File PDF Instant Apache Hive Essentials How To

how NoSQL is ushering in a new age of Polyglot Persistence, where multiple data-storage worlds coexist, and architects can choose the technology best optimized for each type of data access.

Summary Hadoop in Practice, Second Edition provides over 100 tested, instantly useful techniques that will help you conquer big data, using Hadoop. This revised new edition covers changes and new features in the Hadoop core architecture, including MapReduce 2. Brand new chapters cover YARN and integrating Kafka, Impala, and Spark SQL with Hadoop. You'll also get new and updated techniques for Flume, Sqoop, and

Download File PDF Instant Apache Hive Essentials How To

Mahout, all of which have seen major new versions recently. In short, this is the most practical, up-to-date coverage of Hadoop available anywhere. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Book It's always a good time to upgrade your Hadoop skills! Hadoop in Practice, Second Edition provides a collection of 104 tested, instantly useful techniques for analyzing real-time streams, moving data securely, machine learning, managing large-scale clusters, and taming big data using Hadoop. This completely revised edition covers changes

Download File PDF Instant Apache Hive Essentials How To

and new features in Hadoop core, including MapReduce 2 and YARN. You'll pick up hands-on best practices for integrating Spark, Kafka, and Impala with Hadoop, and get new and updated techniques for the latest versions of Flume, Sqoop, and Mahout. In short, this is the most practical, up-to-date coverage of Hadoop available. Readers need to know a programming language like Java and have basic familiarity with Hadoop. What's Inside Thoroughly updated for Hadoop 2 How to write YARN applications Integrate real-time technologies like Storm, Impala, and Spark Predictive analytics using Mahout and RR Readers need to know a

Download File PDF Instant Apache Hive Essentials How To

programming language like Java and have basic familiarity with Hadoop. About the Author Alex Holmes works on tough big-data problems. He is a software engineer, author, speaker, and blogger specializing in large-scale Hadoop projects. Table of Contents

PART 1 BACKGROUND AND FUNDAMENTALS

Hadoop in a heartbeat Introduction to YARN

PART 2 DATA LOGISTICS

Data serialization—working with text and beyond Organizing and optimizing data in HDFS Moving data into and out of Hadoop

PART 3 BIG DATA PATTERNS

Applying MapReduce patterns to big data Utilizing data structures and algorithms at scale Tuning,

Download File PDF Instant Apache Hive Essentials How To

debugging, and testing PART 4
BEYOND MAPREDUCE SQL on
Hadoop Writing a YARN
application

Hadoop Essentials

Apache Hive Cookbook

Hadoop MapReduce Cookbook

Professional NoSQL

NoSQL Distilled

Boots and Saddles

Dive into the world of SQL on
Hadoop and get the most out of your
Hive data warehouses. This book is
your go-to resource for using Hive:
authors Scott Shaw, Ankur Gupta,
David Kjerrumgaard, and Andreas
Francois Vermeulen take you
through learning HiveQL, the SQL-
like language specific to Hive, to

Download File PDF Instant Apache Hive Essentials How To

analyze, export, and massage the data stored across your Hadoop environment. From deploying Hive on your hardware or virtual machine and setting up its initial configuration to learning how Hive interacts with Hadoop, MapReduce, Tez and other big data technologies, *Practical Hive* gives you a detailed treatment of the software. In addition, this book discusses the value of open source software, Hive performance tuning, and how to leverage semi-structured and unstructured data. What You Will Learn

- Install and configure Hive for new and existing datasets
- Perform DDL operations
- Execute efficient DML operations
- Use tables,

Download File PDF Instant Apache Hive Essentials How To

partitions, buckets, and user-defined functions Discover performance tuning tips and Hive best practices Who This Book Is For Developers, companies, and professionals who deal with large amounts of data and could use software that can efficiently manage large volumes of input. It is assumed that readers have the ability to work with SQL.

Learn to use Apache Pig to develop lightweight big data applications easily and quickly. This book shows you many optimization techniques and covers every context where Pig is used in big data analytics.

Beginning Apache Pig shows you how Pig is easy to learn and requires relatively little time to develop big

Download File PDF Instant Apache Hive Essentials How To

data applications. The book is divided into four parts: the complete features of Apache Pig; integration with other tools; how to solve complex business problems; and optimization of tools. You'll discover topics such as MapReduce and why it cannot meet every business need; the features of Pig Latin such as data types for each load, store, joins, groups, and ordering; how Pig workflows can be created; submitting Pig jobs using Hue; and working with Oozie. You'll also see how to extend the framework by writing UDFs and custom load, store, and filter functions. Finally you'll cover different optimization techniques such as gathering

Download File PDF Instant Apache Hive Essentials How To

statistics about a Pig script, joining strategies, parallelism, and the role of data formats in good performance. What You Will Learn• Use all the features of Apache Pig• Integrate Apache Pig with other tools• Extend Apache Pig• Optimize Pig Latin code• Solve different use cases for Pig LatinWho This Book Is ForAll levels of IT professionals: architects, big data enthusiasts, engineers, developers, and big data administrators

Boots and Saddles is in reality a bright and sunny sketch of the life of Mrs. Custer's late husband, General George A. Custer, who fell at the battle of Little Big Horn. After the war, General Custer was sent to the

Download File PDF Instant Apache Hive Essentials How To

Indian frontier. His wife was of the party and she is able to give in minute detail the story of her husband's varied career since she was almost always near the scene of his adventures. She touches on themes little canvassed by the civilian, and makes a volume equally redolent of a loving devotion to an honored husband and attractive as a picture of necessary duty by the soldier. Book jacket.

The best-selling author of *Big Data* is back, this time with a unique and in-depth insight into how specific companies use big data. Big data is on the tip of everyone's tongue. Everyone understands its power and importance, but many fail to grasp

Download File PDF Instant Apache Hive Essentials How To

the actionable steps and resources required to utilise it effectively. This book fills the knowledge gap by showing how major companies are using big data every day, from an up-close, on-the-ground perspective. From technology, media and retail, to sport teams, government agencies and financial institutions, learn the actual strategies and processes being used to learn about customers, improve manufacturing, spur innovation, improve safety and so much more. Organised for easy dip-in navigation, each chapter follows the same structure to give you the information you need quickly. For each company profiled, learn what data was used, what problem it

Download File PDF Instant Apache Hive Essentials How To

solved and the processes put it place to make it practical, as well as the technical details, challenges and lessons learned from each unique scenario. Learn how predictive analytics helps Amazon, Target, John Deere and Apple understand their customers Discover how big data is behind the success of Walmart, LinkedIn, Microsoft and more Learn how big data is changing medicine, law enforcement, hospitality, fashion, science and banking Develop your own big data strategy by accessing additional reading materials at the end of each chapter

HBase

Hadoop in Practice

Download File PDF Instant Apache Hive Essentials How To

Geronimo's Story of His Life

Hands-On Big Data Analytics with
PySpark

Learning PySpark

Unbreakable Willpower and Self
Control Techniques! - Erase Bad
Habits and Replace Them with Self
Discipline, Self Esteem, Motivation
and Better Decision Making!

This comprehensive edited volume is the first of its kind, designed to serve as a textbook for long-duration business analytics programs. It can also be used as a guide to the field by practitioners. The book has contributions from experts in top universities and industry. The editors have taken extreme care to ensure continuity across the chapters. The material is organized into three parts: A) Tools, B) Models and C) Applications. In

Download File PDF Instant Apache Hive Essentials How To

Part A, the tools used by business analysts are described in detail. In Part B, these tools are applied to construct models used to solve business problems. Part C contains detailed applications in various functional areas of business and several case studies. Supporting material can be found in the appendices that develop the pre-requisites for the main text. Every chapter has a business orientation. Typically, each chapter begins with the description of business problems that are transformed into data questions; and methodology is developed to solve these questions. Data analysis is conducted using widely used software, the output and results are clearly explained at each stage of development. These are finally transformed into a business solution. The companion website provides examples, data sets and sample code for each chapter.

Download File PDF Instant Apache Hive Essentials How To

This example-driven book offers a thorough introduction to Java's APIs for XML Web Services (JAX-WS) and RESTful Web Services (JAX-RS). Java Web Services: Up and Running takes a clear, pragmatic approach to these technologies by providing a mix of architectural overview, complete working code examples, and short yet precise instructions for compiling, deploying, and executing an application. You'll learn how to write web services from scratch and integrate existing services into your Java applications. With Java Web Services: Up and Running, you will:

- Understand the distinction between SOAP-based and REST-style services*
- Write, deploy, and consume SOAP-based services in core Java*
- Understand the Web Service Definition Language (WSDL) service contract*
- Recognize the structure of a SOAP message*
- Learn how to deliver*

Download File PDF Instant Apache Hive Essentials How To

Java-based RESTful web services and consume commercial RESTful services Know security requirements for SOAP- and REST-based web services Learn how to implement JAX-WS in various application servers Ideal for students as well as experienced programmers, Java Web Services: Up and Running is the concise guide you need to start working with these technologies right away. JSON is becoming the backbone for meaningful data interchange over the internet. This format is now supported by an entire ecosystem of standards, tools, and technologies for building truly elegant, useful, and efficient applications. With this hands-on guide, author and architect Tom Marris shows you how to build enterprise-class applications and services by leveraging JSON tooling and message/document design. JSON at Work provides

Download File PDF Instant Apache Hive Essentials How To

application architects and developers with guidelines, best practices, and use cases, along with lots of real-world examples and code samples. You'll start with a comprehensive JSON overview, explore the JSON ecosystem, and then dive into JSON's use in the enterprise. Get acquainted with JSON basics and learn how to model JSON data Learn how to use JSON with Node.js, Ruby on Rails, and Java Structure JSON documents with JSON Schema to design and test APIs Search the contents of JSON documents with JSON Search tools Convert JSON documents to other data formats with JSON Transform tools Compare JSON-based hypermedia formats, including HAL and jsonapi Leverage MongoDB to store and access JSON documents Use Apache Kafka to exchange JSON-based messages between services

Download File PDF Instant Apache Hive Essentials How To

Every enterprise application creates data, whether it's log messages, metrics, user activity, outgoing messages, or something else. And how to move all of this data becomes nearly as important as the data itself. If you're an application architect, developer, or production engineer new to Apache Kafka, this practical guide shows you how to use this open source streaming platform to handle real-time data feeds. Engineers from Confluent and LinkedIn who are responsible for developing Kafka explain how to deploy production Kafka clusters, write reliable event-driven microservices, and build scalable stream-processing applications with this platform. Through detailed examples, you'll learn Kafka's design principles, reliability guarantees, key APIs, and architecture details, including the replication protocol, the controller, and the storage layer. Understand

Download File PDF Instant Apache Hive Essentials How To

publish-subscribe messaging and how it fits in the big data ecosystem. Explore Kafka producers and consumers for writing and reading messages Understand Kafka patterns and use-case requirements to ensure reliable data delivery Get best practices for building data pipelines and applications with Kafka Manage Kafka in production, and learn to perform monitoring, tuning, and maintenance tasks Learn the most critical metrics among Kafka's operational measurements Explore how Kafka's stream delivery capabilities make it a perfect source for stream processing systems

Hadoop: The Definitive Guide
Principles of Database Management
Willpower
Hadoop in Action
Analyze large datasets and discover techniques for testing, immunizing, and

Download File PDF Instant Apache Hive Essentials How To

parallelizing Spark jobs

Apache Hive Essentials

Perform fast interactive analytics against different data sources using the Trino high-performance distributed SQL query engine. With this practical guide, you'll learn how to conduct analytics on data where it lives, whether it's Hive, Cassandra, a relational database, or a proprietary data store. Analysts, software engineers, and production engineers will learn how to manage, use, and even develop with Trino. Initially developed by

Download File PDF Instant
Apache Hive Essentials How To

Facebook, open source Trino is now used by Netflix, Airbnb, LinkedIn, Twitter, Uber, and many other companies. Matt Fuller, Manfred Moser, and Martin Traverso show you how a single Trino query can combine data from multiple sources to allow for analytics across your entire organization. Get started: Explore Trino's use cases and learn about tools that will help you connect to Trino and query data Go deeper: Learn Trino's internal workings, including how to connect to and query

Download File PDF Instant
Apache Hive Essentials How To

data sources with support for SQL statements, operators, functions, and more Put Trino in production: Secure Trino, monitor workloads, tune queries, and connect more applications; learn how other organizations apply Trino

A hands-on guide to leveraging NoSQL databases NoSQL databases are an efficient and powerful tool for storing and manipulating vast quantities of data. Most NoSQL databases scale well as data grows. In addition, they are often malleable and

flexible enough to accommodate semi-structured and sparse data sets. This comprehensive hands-on guide presents fundamental concepts and practical solutions for getting you ready to use NoSQL databases. Expert author Shashank Tiwari begins with a helpful introduction on the subject of NoSQL, explains its characteristics and typical uses, and looks at where it fits in the application stack. Unique insights help you choose which NoSQL solutions are best for

Download File PDF Instant
Apache Hive Essentials How To

solving your specific data storage needs. Professional NoSQL: Demystifies the concepts that relate to NoSQL databases, including column-family oriented stores, key/value databases, and document databases. Delves into installing and configuring a number of NoSQL products and the Hadoop family of products. Explains ways of storing, accessing, and querying data in NoSQL databases through examples that use MongoDB, HBase, Cassandra, Redis, CouchDB, Google App Engine

Download File PDF Instant
Apache Hive Essentials How To

Datastore and more. Looks at architecture and internals. Provides guidelines for optimal usage, performance tuning, and scalable configurations. Presents a number of tools and utilities relating to NoSQL, distributed platforms, and scalable processing, including Hive, Pig, RRDtool, Nagios, and more.

Get Started Fast with Apache Hadoop® 2, YARN, and Today's Hadoop Ecosystem With Hadoop 2.x and YARN, Hadoop moves beyond MapReduce to

become practical for virtually any type of data processing. Hadoop 2.x and the Data Lake concept represent a radical shift away from conventional approaches to data usage and storage. Hadoop 2.x installations offer unmatched scalability and breakthrough extensibility that supports new and existing Big Data analytics processing methods and models. Hadoop® 2 Quick-Start Guide is the first easy, accessible guide to Apache Hadoop 2.x, YARN, and the modern Hadoop ecosystem.

Building on his unsurpassed experience teaching Hadoop and Big Data, author Douglas Eadline covers all the basics you need to know to install and use Hadoop 2 on personal computers or servers, and to navigate the powerful technologies that complement it. Eadline concisely introduces and explains every key Hadoop 2 concept, tool, and service, illustrating each with a simple “beginning-to-end” example and identifying trustworthy, up-to-date resources for learning more. This guide is ideal if you

Download File PDF Instant
Apache Hive Essentials How To

want to learn about Hadoop 2 without getting mired in technical details. Douglas Eadline will bring you up to speed quickly, whether you're a user, admin, devops specialist, programmer, architect, analyst, or data scientist.

Coverage Includes

Understanding what Hadoop 2 and YARN do, and how they improve on Hadoop 1 with MapReduce

Understanding Hadoop-

based Data Lakes versus

RDBMS Data Warehouses

Installing Hadoop 2 and core services on Linux machines,

Download File PDF Instant
Apache Hive Essentials How To

**virtualized sandboxes, or
clusters Exploring the
Hadoop Distributed File
System (HDFS)
Understanding the
essentials of MapReduce
and YARN application
programming Simplifying
programming and data
movement with Apache Pig,
Hive, Sqoop, Flume, Oozie,
and HBase Observing
application progress,
controlling jobs, and
managing workflows
Managing Hadoop efficiently
with Apache
Ambari-including recipes for
HDFS to NFSv3 gateway,**

Download File PDF Instant
Apache Hive Essentials How To

**HDFS snapshots, and YARN
configuration Learning basic
Hadoop 2 troubleshooting,
and installing Apache Hue
and Apache Spark**

**Oklahoman S.M Barret wrote
down and edited Apache
Chief Geronimo's story of his
life.**

**Instant Apache Hive
Essentials How-to
Essential techniques to help
you process, and get unique
insights from, big data, 2nd
Edition**

**Apache Oozie Essentials
Essentials of Business
Analytics**

Kafka: The Definitive Guide

Download File PDF Instant Apache Hive Essentials How To

Learning Spark

Apache Hadoop is the technology at the heart of the Big Data revolution, and Hadoop skills are in enormous demand. Now, in just 24 lessons of one hour or less, you can learn all the skills and techniques you'll need to deploy each key component of a Hadoop platform in your local environment or in the cloud, building a fully functional Hadoop cluster and using it with real programs and datasets. Each short, easy lesson builds on all that's come before, helping you master all of Hadoop's essentials, and extend it to meet your unique challenges. Apache Hadoop in 24 Hours, Sams Teach Yourself covers all this, and much more: Understanding Hadoop and the Hadoop Distributed File System (HDFS) Importing data into Hadoop,

Download File PDF Instant Apache Hive Essentials How To

and process it there Mastering basic MapReduce Java programming, and using advanced MapReduce API concepts Making the most of Apache Pig and Apache Hive Implementing and administering YARN Taking advantage of the full Hadoop ecosystem Managing Hadoop clusters with Apache Ambari Working with the Hadoop User Environment (HUE) Scaling, securing, and troubleshooting Hadoop environments Integrating Hadoop into the enterprise Deploying Hadoop in the cloud Getting started with Apache Spark Step-by-step instructions walk you through common questions, issues, and tasks; Q-and-As, Quizzes, and Exercises build and test your knowledge; "Did You Know?" tips offer insider advice and shortcuts; and "Watch Out!" alerts help you avoid pitfalls. By the time you're finished,

Download File PDF Instant Apache Hive Essentials How To

you'll be comfortable using Apache Hadoop to solve a wide spectrum of Big Data problems.

Feature engineering is a crucial step in the machine-learning pipeline, yet this topic is rarely examined on its own.

With this practical book, you'll learn techniques for extracting and transforming features—the numeric representations of raw data—into formats for machine-learning models. Each chapter guides you through a single data problem, such as how to represent text or image data.

Together, these examples illustrate the main principles of feature engineering.

Rather than simply teach these principles, authors Alice Zheng and Amanda Casari focus on practical application with exercises throughout the book. The closing chapter brings everything together by tackling a real-

Download File PDF Instant Apache Hive Essentials How To

world, structured dataset with several feature-engineering techniques.

Python packages including numpy, Pandas, Scikit-learn, and Matplotlib are used in code examples. You'll examine:

- Feature engineering for numeric data: filtering, binning, scaling, log transforms, and power transforms
- Natural text techniques: bag-of-words, n-grams, and phrase detection
- Frequency-based filtering and feature scaling for eliminating uninformative features
- Encoding techniques of categorical variables, including feature hashing and bin-counting
- Model-based feature engineering with principal component analysis
- The concept of model stacking, using k-means as a featurization technique
- Image feature extraction with manual and deep-learning techniques

Download File PDF Instant Apache Hive Essentials How To

Microsoft Azure Essentials from Microsoft Press is a series of free ebooks designed to help you advance your technical skills with Microsoft Azure. The first ebook in the series, Microsoft Azure Essentials: Fundamentals of Azure, introduces developers and IT professionals to the wide range of capabilities in Azure. The authors - both Microsoft MVPs in Azure - present both conceptual and how-to content for key areas, including: Azure Websites and Azure Cloud Services Azure Virtual Machines Azure Storage Azure Virtual Networks Databases Azure Active Directory Management tools Business scenarios Watch Microsoft Press's blog and Twitter (@MicrosoftPress) to learn about other free ebooks in the "Microsoft Azure Essentials" series. Unleash the power of Apache Oozie to

Download File PDF Instant Apache Hive Essentials How To

create and manage your big data and machine learning pipelines in one go
About This Book Teaches you everything you need to know to get started with Apache Oozie from scratch and manage your data pipelines effortlessly Learn to write data ingestion workflows with the help of real-life examples from the author's own personal experience Embed Spark jobs to run your machine learning models on top of Hadoop
Who This Book Is For If you are an expert Hadoop user who wants to use Apache Oozie to handle workflows efficiently, this book is for you. This book will be handy to anyone who is familiar with the basics of Hadoop and wants to automate data and machine learning pipelines. What You Will Learn Install and configure Oozie from source code on your Hadoop cluster

Download File PDF Instant Apache Hive Essentials How To

Dive into the world of Oozie with Java MapReduce jobs Schedule Hive ETL and data ingestion jobs Import data from a database through Sqoop jobs in HDFS Create and process data pipelines with Pig, hive scripts as per business requirements. Run machine learning Spark jobs on Hadoop Create quick Oozie jobs using Hue Make the most of Oozie's security capabilities by configuring Oozie's security In Detail As more and more organizations are discovering the use of big data analytics, interest in platforms that provide storage, computation, and analytic capabilities is booming exponentially. This calls for data management. Hadoop caters to this need. Oozie fulfils this necessity for a scheduler for a Hadoop job by acting as a cron to better analyze data. Apache Oozie Essentials starts off

Download File PDF Instant Apache Hive Essentials How To

with the basics right from installing and configuring Oozie from source code on your Hadoop cluster to managing your complex clusters. You will learn how to create data ingestion and machine learning workflows. This book is sprinkled with the examples and exercises to help you take your big data learning to the next level. You will discover how to write workflows to run your MapReduce, Pig ,Hive, and Sqoop scripts and schedule them to run at a specific time or for a specific business requirement using a coordinator. This book has engaging real-life exercises and examples to get you in the thick of things. Lastly, you'll get a grip of how to embed Spark jobs, which can be used to run your machine learning models on Hadoop. By the end of the book, you will have a good knowledge of Apache Oozie.

Download File PDF Instant Apache Hive Essentials How To

You will be capable of using Oozie to handle large Hadoop workflows and even improve the availability of your Hadoop environment. Style and approach This book is a hands-on guide that explains Oozie using real-world examples. Each chapter is blended beautifully with fundamental concepts sprinkled in-between case study solution algorithms and topped off with self-learning exercises.

An Apache Princess

An Introduction to the Methodology and its Applications

Drew's Station, Contention City and Fairbank

Practical Data Science with Hadoop and Spark

JSON at Work

Hadoop in 24 Hours, Sams Teach Yourself

Ready to unlock the power

Download File PDF Instant Apache Hive Essentials How To

of your data? With this comprehensive guide, you'll learn how to build and maintain reliable, scalable, distributed systems with Apache Hadoop. This book is ideal for programmers looking to analyze datasets of any size, and for administrators who want to set up and run Hadoop clusters. You'll find illuminating case studies that demonstrate how Hadoop is used to solve specific problems. This third edition covers recent changes to Hadoop, including material on the

Download File PDF Instant Apache Hive Essentials How To

new MapReduce API, as well as MapReduce 2 and its more flexible execution model (YARN). Store large datasets with the Hadoop Distributed File System (HDFS) Run distributed computations with MapReduce Use Hadoop's data and I/O building blocks for compression, data integrity, serialization (including Avro), and persistence Discover common pitfalls and advanced features for writing real-world MapReduce programs Design, build, and administer a dedicated Hadoop

Download File PDF Instant Apache Hive Essentials How To

cluster—or run Hadoop in the cloud Load data from relational databases into HDFS, using Sqoop Perform large-scale data processing with the Pig query language Analyze datasets with Hive, Hadoop's data warehousing system Take advantage of HBase for structured and semi-structured data, and ZooKeeper for building distributed systems Introductory, theory-practice balanced text teaching the fundamentals of databases to advanced undergraduates or graduate students in information

Download File PDF Instant Apache Hive Essentials How To

systems or computer science.

Filled with practical, step-by-step instructions and clear explanations for the most important and useful tasks. This book provides quick recipes for using Hive to read data in various formats, efficiently querying this data, and extending Hive with any custom functions you may need to insert your own logic into the data pipeline. This book is written for data analysts and developers who want to use their current knowledge of SQL to be

Download File PDF Instant Apache Hive Essentials How To

more productive with Hadoop. It assumes that readers are comfortable writing SQL queries and are familiar with Hadoop at the level of the classic WordCount example. If you are a data analyst, developer, or simply someone who wants to use Hive to explore and analyze data in Hadoop, this is the book for you. Whether you are new to big data or an expert, with this book, you will be able to master both the basic and the advanced features of Hive. Since Hive is an SQL-like

Download File PDF Instant Apache Hive Essentials How To

language, some previous experience with the SQL language and databases is useful to have a better understanding of this book.

Principles and Techniques
for Data Scientists

Big Data Processing Made
Easy

Apache Hadoop YARN

Enterprise Cloud epUB _1

Beginning Apache Pig

Practical Data Integration
for the Web

Hadoop in Action teaches readers how to use Hadoop and write MapReduce programs. The intended readers are

Download File PDF Instant Apache Hive Essentials How To

programmers, architects, and project managers who have to process large amounts of data offline. Hadoop in Action will lead the reader from obtaining a copy of Hadoop to setting it up in a cluster and writing data analytic programs. The book begins by making the basic idea of Hadoop and MapReduce easier to grasp by applying the default Hadoop installation to a few easy-to-follow tasks, such as analyzing changes in word

Download File PDF Instant Apache Hive Essentials How To

frequency across a body of documents. The book continues through the basic concepts of MapReduce applications developed using Hadoop, including a close look at framework components, use of Hadoop for a variety of data analysis tasks, and numerous examples of Hadoop in action. Hadoop in Action will explain how to use Hadoop and present design patterns and practices of programming MapReduce. MapReduce is a complex idea both

Download File PDF Instant Apache Hive Essentials How To

conceptually and in its implementation, and Hadoop users are challenged to learn all the knobs and levers for running Hadoop. This book takes you beyond the mechanics of running Hadoop, teaching you to write meaningful programs in a MapReduce framework. This book assumes the reader will have a basic familiarity with Java, as most code examples will be written in Java. Familiarity with basic statistical concepts (e.g.

Download File PDF Instant Apache Hive Essentials How To

histogram, correlation) will help the reader appreciate the more advanced data processing examples. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book.

Use PySpark to easily crush messy data at-scale and discover proven techniques to create testable, immutable, and easily parallelizable Spark jobs

Key Features

Work

Download File PDF Instant Apache Hive Essentials How To

with large amounts of agile data using distributed datasets and in-memory cachingSource data from all popular data hosting platforms, such as HDFS, Hive, JSON, and S3Employ the easy-to-use PySpark API to deploy big data Analytics for productionBook Description Apache Spark is an open source parallel-processing framework that has been around for quite some time now. One of the many uses of Apache

Download File PDF Instant Apache Hive Essentials How To

Spark is for data analytics applications across clustered computers. In this book, you will not only learn how to use Spark and the Python API to create high-performance analytics with big data, but also discover techniques for testing, immunizing, and parallelizing Spark jobs. You will learn how to source data from all popular data hosting platforms, including HDFS, Hive, JSON, and S3, and deal with large

Download File PDF Instant Apache Hive Essentials How To

datasets with PySpark to gain practical big data experience. This book will help you work on prototypes on local machines and subsequently go on to handle messy data in production and at scale. This book covers installing and setting up PySpark, RDD operations, big data cleaning and wrangling, and aggregating and summarizing data into useful reports. You will also learn how to implement some practical

Download File PDF Instant Apache Hive Essentials How To

and proven techniques to improve certain aspects of programming and administration in Apache Spark. By the end of the book, you will be able to build big data analytical solutions using the various PySpark offerings and also optimize them effectively. What you will learn

Get practical big data experience while working on messy datasets

Analyze patterns with Spark SQL to improve your business intelligence

Use

Download File PDF Instant Apache Hive Essentials How To

PySpark's interactive shell to speed up development time
Create highly concurrent Spark programs by leveraging immutability
Discover ways to avoid the most expensive operation in the Spark API: the shuffle operation
Re-design your jobs to use reduceByKey instead of groupBy
Create robust processing pipelines by testing Apache Spark jobs
Who this book is for
This book is for developers, data scientists, business

Download File PDF Instant Apache Hive Essentials How To

analysts, or anyone who needs to reliably analyze large amounts of large-scale, real-world data. Whether you're tasked with creating your company's business intelligence function or creating great data platforms for your machine learning models, or are looking to use code to magnify the impact of your business, this book is for you. Can we add value to the current Apache Hive decision-making process (largely qualitative) by

Download File PDF Instant Apache Hive Essentials How To

incorporating
uncertainty modeling
(more quantitative)?
Apache Hive in
management -Strategic
planning How will the
Apache Hive team and the
organization measure
complete success of
Apache Hive? Will Apache
Hive deliverables need
to be tested and, if so,
by whom? Who will be
responsible for deciding
whether Apache Hive goes
ahead or not after the
initial investigations?
This premium Apache Hive
self-assessment will

Download File PDF Instant Apache Hive Essentials How To

make you the credible Apache Hive domain auditor by revealing just what you need to know to be fluent and ready for any Apache Hive challenge. How do I reduce the effort in the Apache Hive work to be done to get problems solved? How can I ensure that plans of action include every Apache Hive task and that every Apache Hive outcome is in place? How will I save time investigating strategic and tactical options and ensuring

Download File PDF Instant Apache Hive Essentials How To

Apache Hive costs are low? How can I deliver tailored Apache Hive advice instantly with structured going-forward plans? There's no better guide through these mind-expanding questions than acclaimed best-selling author Gerard Blokdyk. Blokdyk ensures all Apache Hive essentials are covered, from every angle: the Apache Hive self-assessment shows succinctly and clearly that what needs to be clarified to organize the required activities

Download File PDF Instant Apache Hive Essentials How To

and processes so that Apache Hive outcomes are achieved. Contains extensive criteria grounded in past and current successful projects and activities by experienced Apache Hive practitioners. Their mastery, combined with the easy elegance of the self-assessment, provides its superior value to you in knowing how to ensure the outcome of any efforts in Apache Hive are maximized with professional results.

Download File PDF Instant Apache Hive Essentials How To

Your purchase includes access details to the Apache Hive self-assessment dashboard download which gives you your dynamically prioritized projects-ready tool and shows you exactly what to do next. Your exclusive instant access details can be found in your book. You will receive the following contents with New and Updated specific criteria: - The latest quick edition of the book in PDF - The latest complete edition of the

Download File PDF Instant Apache Hive Essentials How To

book in PDF, which
criteria correspond to
the criteria in... - The
Self-Assessment Excel
Dashboard, and... -
Example pre-filled Self-
Assessment Excel
Dashboard to get
familiar with results
generation ...plus an
extra, special, resource
that helps you with
project managing.

INCLUDES LIFETIME SELF
ASSESSMENT UPDATES Every
self assessment comes
with Lifetime Updates
and Lifetime Free
Updated Books. Lifetime

Download File PDF Instant Apache Hive Essentials How To

Updates is an industry-first feature which allows you to receive verified self assessment updates, ensuring you always have the most accurate information at your fingertips.

Individual self-contained code recipes. Solve specific problems using individual recipes, or work through the book to develop your capabilities. If you are a big data enthusiast and striving to use Hadoop to solve your problems, this book is

Download File PDF Instant Apache Hive Essentials How To

for you. Aimed at Java programmers with some knowledge of Hadoop MapReduce, this is also a comprehensive reference for developers and system admins who want to get up to speed using Hadoop.

Hadoop 2 Quick-Start
Guide

Briggs

Java Web Services: Up
and Running

Moving Beyond MapReduce
and Batch Processing
with Apache Hadoop 2
The Definitive Guide

Real-Time Data and

Download File PDF Instant Apache Hive Essentials How To

Stream Processing at Scale

Willpower Ultimate Guide! This "Willpower" book contains proven steps and strategies on how to quickly and easily develop new habits of self discipline, self esteem, and better decision making! Today only, get this Amazing Amazon book for this incredibly discounted price! If you don't have strong willpower, you will often be very discouraged and frustrated with yourself. It is hard to have a good self image and self esteem if you go back on your word every time you decide that you want to do something. This implies that you do not trust yourself and causes you to feel depressed and other negative

Download File PDF Instant Apache Hive Essentials How To

emotions. Don't allow this to happen anymore! Stop the bad cycle you are in and learn these simple techniques to increasing your willpower, self esteem, self discipline and decision making NOW! I am confident you will be happy that you did. This book contains easy to follow steps to help enhance your willpower and develop self-control, two important aspects in achieving your goals. Learn how to take advantage of the power of your mind to help you attain the things you want in life. No need to suffer stress and anxiety anymore and learn to live the life you want with the help of this book. Here Is A Preview Of What You'll Learn...
Understanding Willpower And The Benefits You Can Gain From

Download File PDF Instant Apache Hive Essentials How To

Having It How To Build
Unbreakable Willpower To Erase
Bad Habits And Replace Them
With Good Ones Eliminate
Barriers To Your Willpower
Success Such As Perfectionism
And Self-Doubt How To Develop
Amazing Self-Control And Resist
The Biggest Temptations Using
Willpower Stop Procrastination
NOW With These Unstoppable
Willpower Tips Time Management
Techniques To Keep You On Track
And Reinforce Self-Discipline
Meditation Strategies To Help You
Visualize Yourself Full of
Willpower And Able To Make
Better Decisions Setting Daily
Goals To Build Self-Esteem And
Get Laser Focused On What You
Want To Accomplish And Be In
Control Of Brain Training

Download File PDF Instant Apache Hive Essentials How To

Strategies To Increase Motivation
And Willpower To Succeed Putting
It All Together Into A Quick And
Easy Daily Routine To Increase
Willpower And Self-Discipline
Much, Much More! Get Your Copy
Today!

Build data-intensive applications
locally and deploy at scale using
the combined powers of Python
and Spark 2.0 About This Book
Learn why and how you can
efficiently use Python to process
data and build machine learning
models in Apache Spark 2.0
Develop and deploy efficient,
scalable real-time Spark solutions
Take your understanding of using
Spark with Python to the next
level with this jump start guide
Who This Book Is For If you are a
Python developer who wants to

Download File PDF Instant Apache Hive Essentials How To

learn about the Apache Spark 2.0 ecosystem, this book is for you. A firm understanding of Python is expected to get the best out of the book. Familiarity with Spark would be useful, but is not mandatory. What You Will Learn

- Learn about Apache Spark and the Spark 2.0 architecture
- Build and interact with Spark DataFrames using Spark SQL
- Learn how to solve graph and deep learning problems using GraphFrames and TensorFrames respectively
- Read, transform, and understand data and use it to train machine learning models
- Build machine learning models with MLlib and ML
- Learn how to submit your applications programmatically using spark-submit
- Deploy locally built

Download File PDF Instant Apache Hive Essentials How To

applications to a cluster In Detail Apache Spark is an open source framework for efficient cluster computing with a strong interface for data parallelism and fault tolerance. This book will show you how to leverage the power of Python and put it to use in the Spark ecosystem. You will start by getting a firm understanding of the Spark 2.0 architecture and how to set up a Python environment for Spark. You will get familiar with the modules available in PySpark. You will learn how to abstract data with RDDs and DataFrames and understand the streaming capabilities of PySpark. Also, you will get a thorough overview of machine learning capabilities of PySpark using ML and MLlib,

Download File PDF Instant Apache Hive Essentials How To

graph processing using GraphFrames, and polyglot persistence using Blaze. Finally, you will learn how to deploy your applications to the cloud using the spark-submit command. By the end of this book, you will have established a firm understanding of the Spark Python API and how it can be used to build data-intensive applications. Style and approach This book takes a very comprehensive, step-by-step approach so you understand how the Spark ecosystem can be used with Python to develop efficient, scalable solutions. Every chapter is standalone and written in a very easy-to-understand manner, with a focus on both the hows and the whys of each concept. If your organization is looking for

Download File PDF Instant Apache Hive Essentials How To

a storage solution to accommodate a virtually endless amount of data, this book will show you how Apache HBase can fulfill your needs. As the open source implementation of Google's BigTable architecture, HBase scales to billions of rows and millions of columns, while ensuring that write and read performance remain constant. HBase: The Definitive Guide provides the details you require, whether you simply want to evaluate this high-performance, non-relational database, or put it into practice right away. HBase's adoption rate is beginning to climb, and several IT executives are asking pointed questions about this high-capacity database. This is the

Download File PDF Instant Apache Hive Essentials How To

only book available to give you meaningful answers. Learn how to distribute large datasets across an inexpensive cluster of commodity servers Develop HBase clients in many programming languages, including Java, Python, and Ruby Get details on HBase's primary storage system, HDFS—Hadoop's distributed and replicated filesystem Learn how HBase's native interface to Hadoop's MapReduce framework enables easy development and execution of batch jobs that can scan entire tables Discover the integration between HBase and other facets of the Apache Hadoop project Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be

Download File PDF Instant Apache Hive Essentials How To

processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walk-throughs, code snippets, and notebooks, you'll be able to:

- Learn Python, SQL, Scala, or Java high-level Structured APIs
- Understand Spark operations and SQL Engine
- Inspect, tune, and debug Spark operations with Spark configurations and Spark UI

Download File PDF Instant Apache Hive Essentials How To

Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLflow

Trino: The Definitive Guide
The Practical Guide to Storing, Managing and Analyzing Big and Small Data

Designing and Building Effective Analytics at Scale

Feature Engineering for Machine Learning

Practical Hive