Learning Apache Cassandra

Design documents and implement real world e-commerce applications with Couchbase About This Book Get acquainted with Couchbase architecture and design your document-based data schema Implement full text search using industry standard elastic search plugins Develop critical and high performance applications using this hands-on tutorial guide Who This Book Is For If you are new to the NoSQL document system or have little or no experience in NoSQL development and administration and are planning to deploy Couchbase for your next project, then this book is for you. It would be helpful to have a bit of familiarity with Java. What You Will Learn Get acquainted with the concept of NoSQL databases and configure your Couchbase database cluster Maintain Couchbase effectively using the web-based administrative console with ease Enable partition capabilities by making use of Buckets Analyze important design considerations for maintaining relationship between various documents Use Couchbase SDK Java API to store and retrieve document Write views using map/reduce to retrieve documents efficiently Get familiar with N1QL and how to use it in Java applications Integrate Couchbase with Elasticsearch to implement full text search Configure XDCR for disaster recovery and develop ecommerce application using Couchbase In Detail NoSQL database systems have changed application development in terms of adaptability to dynamics schema and scalability. Compared with the currently available NoSQL database systems, Couchbase is the fastest. Its ease of configuration and powerful features for storing different schema structures, retrieval using map reduce and inbuilt disaster recovery by replicating document across the geographical region, make it one of the most powerful, scalable and comprehensive NoSQL in the market. Couchbase also introduces smart client API for various programming language to integrate the database with the application easily, yet providing very complex features like cluster health awareness. This book achieves its goal by taking up an end-to-end development structure, right from understanding NOSQL document design to implementing full fledged eCommerce application design using Couchbase as a backend. Starting with the architecture of Couchbase to get you up and running, this book quickly takes you through designing a NoSQL document and implementing highly scalable applications using Java API. You will then be introduced to document design and get to know the various ways to administer Couchbase. Followed by this, learn to store documents using bucket. Moving on, you will then learn to store, retrieve and delete documents using smart client base on Java API. You will then retrieve documents using SQL like syntax call N1QL. Next, you will learn how to write map reduce base views. Finally, you will configure XDCR for disaster recovery and implement an eCommerce application using Couchbase. Style and approach The book starts from absolute basics and slowly moves to more advanced topics ensuring at every step that all concepts and terms are understood by the reader to have

complete understanding at every stage. Technical and complex terms are explained in clear and simple language, thus making this book a perfect companion for those who have started their journey to NoSQL using Couchbase When it comes to choosing, using, and maintaining a database, understanding its internals is essential. But with so many distributed databases and tools available today, it's often difficult to understand what each one offers and how they differ. With this practical guide, Alex Petrov guides developers through the concepts behind modern database and storage engine internals. Throughout the book, you'll explore relevant material gleaned from numerous books, papers, blog posts, and the source code of several open source databases. These resources are listed at the end of parts one and two. You'll discover that the most significant distinctions among many modern databases reside in subsystems that determine how storage is organized and how data is distributed. This book examines: Storage engines: Explore storage classification and taxonomy, and dive into B-Tree-based and immutable Log Structured storage engines, with differences and use-cases for each Storage building blocks: Learn how database files are organized to build efficient storage, using auxiliary data structures such as Page Cache, Buffer Pool and Write-Ahead Log Distributed systems: Learn step-by-step how nodes and processes connect and build complex communication patterns Database clusters: Which consistency models are commonly used by modern databases and how distributed storage systems achieve consistency

About Professional Certification of Apache Cassandra: Apache Cassandra is one of the most popular NoSQL Database currently being used by many of the organization, globally in every industry like Aviation, Finance, Retail, Social Networking etc. It proves that there is quite a huge demand for certified Cassandra professionals. Having certification make your selection in the company make much easier. This certification is conducted by the DataStax®, which has the Enterprise Version of the Apache Cassandra and Leader in providing support for the open source Apache Cassandra NoSQL database. Cassandra is one of the Unique NoSQL Database. So go for its certification, it will certainly help in - Getting the Job - Increase in your salary - Growth in your career. - Managing Tera Bytes of Data. - Learning Distributed Database - Using CQL (Cassandra Query Language) Cassandra Certification Information: - Number of questions: 60 Multiple Choice - Time allowed in minutes: 90 - Required passing score: 75% - Languages: English Exam Objectives: There are in total 5 sections and you will be asked total 60 questions in real exam. Please check each section below with regards to the exam objective 1. Apache Cassandra™ data modeling 2. Fundamentals of replication and consistency 3. The distributed and internal architecture of Apache Cassandra™ 4. Installation and configuration 5. Basic tooling

"Cassandra is a NoSQL database with decentralized, fault-tolerant, scalable, and low-cost features, making it a core component of cloud computing systems. The

more recent versions have greatly improved the security features, making it suitable for use in enterprise systems. In this tutorial, you'll see how Cassandra overcomes the challenges that relational databases face during high scalability demand. You will become familiar with the Cassandra terminologies, components, and their roles. Then you will learn how to create a multi-node Cassandra structure, understand the roles and responsibilities of Cassandra components, and see the data flow during database operations that demand speed, accuracy, and durability. You will then see how Cassandra stores data onto files on the disk, how to optimize those files to improve performance, and how to monitor the Cassandra database performance using logs and metrics. We'll demonstrate the factors that could affect the performance SLAs of the Cassandra database. Next, you will learn how to optimize the data model to provide performance guarantees and consistent performance SLA over time. You'll also learn how to build the data model on Cassandra and integrate the database with your application. In the later sections, you'll connect with Cassandra from Spark to read and write data. You'll integrate Cassandra with Spark and learn how to process live streaming data with Spark and persist the data in Cassandra for consumption through the downstream system. By the end of the course, you'll be able to build powerful, scalable Cassandra database layers for your applications. You'll design rich schemes to capture the relationships between different data types and master the advanced features available in Cassandra."--Resource description page.

If you're an application developer familiar with SQL databases such as MySQL or Postgres, and you want to explore distributed databases such as Cassandra, this is the perfect guide for you. Even if you've never worked with a distributed database before, Cassandra's intuitive programming interface coupled with the step-by-step examples in this book will have you building highly scalable persistence layers for your applications in no time.

Learn about the fastest-growing open source project in the world, and find out how it revolutionizes big data analytics About This Book Exclusive guide that covers how to get up and running with fast data processing using Apache Spark Explore and exploit various possibilities with Apache Spark using real-world use cases in this book Want to perform efficient data processing at real time? This book will be your one-stop solution. Who This Book Is For This guide appeals to big data engineers, analysts, architects, software engineers, even technical managers who need to perform efficient data processing on Hadoop at real time. Basic familiarity with Java or Scala will be helpful. The assumption is that readers will be from a mixed background, but would be typically people with background in engineering/data science with no prior Spark experience and want to understand how Spark can help them on their analytics journey. What You Will Learn Get an overview of big data analytics and its importance for organizations and data professionals Delve into Spark to see how it is different from existing processing platforms Understand the intricacies of various file formats, and how

to process them with Apache Spark. Realize how to deploy Spark with YARN, MESOS or a Stand-alone cluster manager. Learn the concepts of Spark SQL, SchemaRDD, Caching and working with Hive and Parquet file formats Understand the architecture of Spark MLLib while discussing some of the off-the-shelf algorithms that come with Spark. Introduce yourself to the deployment and usage of SparkR. Walk through the importance of Graph computation and the graph processing systems available in the market Check the real world example of Spark by building a recommendation engine with Spark using ALS. Use a Telco data set, to predict customer churn using Random Forests. In Detail Spark juggernaut keeps on rolling and getting more and more momentum each day. Spark provides key capabilities in the form of Spark SQL, Spark Streaming, Spark ML and Graph X all accessible via Java, Scala, Python and R. Deploying the key capabilities is crucial whether it is on a Standalone framework or as a part of existing Hadoop installation and configuring with Yarn and Mesos. The next part of the journey after installation is using key components, APIs, Clustering, machine learning APIs, data pipelines, parallel programming. It is important to understand why each framework component is key, how widely it is being used, its stability and pertinent use cases. Once we understand the individual components, we will take a couple of real life advanced analytics examples such as 'Building a Recommendation system', 'Predicting customer churn' and so on. The objective of these real life examples is to give the reader confidence of using Spark for real-world problems. Style and approach With the help of practical examples and real-world use cases, this guide will take you from scratch to building efficient data applications using Apache Spark. You will learn all about this excellent data processing engine in a step-by-step manner, taking one aspect of it at a time. This highly practical guide will include how to work with data pipelines, dataframes, clustering, SparkSQL, parallel programming, and such insightful topics with the help of real-world use cases.
If you want to efficiently use Storm and Cassandra together and excel at developing production-grade, distributed real-time applications, then this book is for you. No prior knowledge of using Storm and Cassandra together is necessary. However, a background in Java is expected.
"Eric and Russell were early adopters of Cassandra at SimpleReach. In Practical Cassandra, you benefit from their experience in the trenches administering Cassandra, developing against it, and building one of the first CQL drivers. If you are deploying Cassandra soon, or you inherited a Cassandra cluster to tend, spend some time with the deployment, performance tuning, and maintenance chapters… If you are new to Cassandra, I highly recommend the chapters on data modeling and CQL." –From the Foreword by Jonathon Ellis, Apache Cassandra Chair Build and Deploy Massively Scalable, Super-fast Data Management Applications with Apache Cassandra Practical Cassandra is the first hands-on developer's guide to building Cassandra systems and applications that deliver breakthrough speed, scalability, reliability, and performance. Fully up to date, it

reflects the latest versions of Cassandra–including Cassandra Query Language (CQL), which dramatically lowers the learning curve for Cassandra developers. Pioneering Cassandra developers and Datastax MVPs Russell Bradberry and Eric Lubow walk you through every step of building a real production application that can store enormous amounts of structured, semi-structured, and unstructured data. Drawing on their exceptional expertise, Bradberry and Lubow share practical insights into issues ranging from querying to deployment, management, maintenance, monitoring, and troubleshooting. The authors cover key issues, from architecture to migration, and guide you through crucial decisions about configuration and data modeling. They provide tested sample code, detailed explanations of how Cassandra works "under the covers," and new case studies from three cutting-edge users: Ooyala, Hailo, and eBay. Coverage includes Understanding Cassandra's approach, architecture, key concepts, and primary use cases– and why it's so blazingly fast Getting Cassandra up and running on single nodes and large clusters Applying the new design patterns, philosophies, and features that make Cassandra such a powerful data store Leveraging CQL to simplify your transition from SQL-based RDBMSes Deploying and provisioning through the cloud or on bare-metal hardware Choosing the right configuration options for each type of workload Tweaking Cassandra to get maximum performance from your hardware, OS, and JVM Mastering Cassandra's essential tools for maintenance and monitoring Efficiently solving the most common problems with Cassandra deployment, operation, and application development

If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats with ease from within Spark Learn about alternative modeling frameworks for graph processing, geospatial analysis, and genomics at scale Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark extensions

Mastering Apache Cassandra is a practical, hands-on guide with step-by-step instructions. The smooth and easy tutorial approach focuses on showing people how to utilize Cassandra to its full potential. This book is aimed at intermediate Cassandra

users. It is best suited for startups where developers have to wear multiple hats: programmer, DevOps, release manager, convincing clients, and handling failures. No prior knowledge of Cassandra is required.

If you are a developer or DevOps engineer who understands the basics of Cassandra and are ready to take your knowledge to the next level, then this book is for you. An understanding of the essentials of Cassandra is needed.

"In this Apache Cassandra training course, expert author Ruth Stryker teaches you the basics of working with Cassandra. This course is designed for the absolute beginner, meaning no experience with Cassandra is required. Experience with databases, SQL, and programming is recommended. You will start by learning how to install Cassandra, and then move onto understanding ways to communicate with Cassandra. From there, Ruth will teach you how to create a database, create a table, and insert and model data. This video tutorial will also cover creating an application, updating and deleting data, selecting hardware, and adding nodes to a cluster. You will also learn how to monitor a cluster, repair and remove nodes, and redefine a cluster. Once you have completed this video based training course, you will have a solid understanding of Cassandra, and be able to use Cassandra for your own development projects. Working files are included, allowing you to follow along with the author throughout the lessons."--Resource description page.

This book is a step by step beginners guide to learning Cassandra. The book uses tons of charts, graphs, images and code to aid your Cassandra learning. The book gives a detailed introduction to Cassandra. It proceeds to give step-by-step instructions to installing Cassandra. Cassandra Architecture and Replication Factor Strategy is lucidly explained. Data Modelling, Keyspace CQL are also described in detail. The book will teach you enough to get started with Cassandra. Here is what is included Table Of Content Chapter 1: Introduction 1. Cassandra History 2. Nosql Cassandra Database 3. Nosql Cassandra Database Vs Relational databases 4. Apache Cassandra Features 5. Cassandra Use Cases Chapter 2: Download and Install 1. Prerequisite for Apache Cassandra Installation 2. How to Download and Install Cassandra Chapter 3: Architecture 1. Components of Cassandra 2. Data Replication 3. Write Operation 4. Read Operation Chapter 4: Data Model and Rules 1. Cassandra Data Model Rules 2. Model Your Data in Cassandra 3. Handling One to One Relationship 4. Handling one to many Relationship 5. Handling Many to Many Relationship Chapter 5: Cassandra CQL 1. Create, Alter & Drop Keyspace 2. Cassandra Table: Create, Alter, Drop & Truncate 3. Cassandra Query Language(CQL): Insert, Update, Delete, Read Data 4. Create & Drop INDEX 5. Data Types & Expiration 6. SET, LIST & MAP Chapter 6: Cassandra Cluster 1. Prerequisites for Cassandra Cluster 2. Enterprise Edition Installation 3. Starting Cassandra Node Chapter 7: DevCenter & OpsCenter Installation 1. DevCenter Installation 2. OpsCenter Installation Chapter 8: Security 1. What is Internal Authentication and Authorization 2. Configure Authentication and Authorization 3. Logging in 4. Create New User 5. Authorization 6. Configuring Firewall 7. Enabling JMX Authentication

Deep Learning is a subset of Machine Learning where data sets with several layers of complexity can be processed. This book teaches you the different techniques using which deep learning solutions can be implemented at scale, on Apache Spark. This will help you gain experience of implementing your deep learning models in many real-

world use cases.

Build, manage, and configure high-performing, reliable NoSQL database for your applications with Cassandra Key Features Write programs more efficiently using Cassandra's features with the help of examples Configure Cassandra and fine-tune its parameters depending on your needs Integrate Cassandra database with Apache Spark and build strong data analytics pipeline Book Description With ever-increasing rates of data creation, the demand for storing data fast and reliably becomes a need. Apache Cassandra is the perfect choice for building fault-tolerant and scalable databases. Mastering Apache Cassandra 3.x teaches you how to build and architect your clusters, configure and work with your nodes, and program in a high-throughput environment, helping you understand the power of Cassandra as per the new features. Once you've covered a brief recap of the basics, you'll move on to deploying and monitoring a production setup and optimizing and integrating it with other software. You'll work with the advanced features of CQL and the new storage engine in order to understand how they function on the server-side. You'll explore the integration and interaction of Cassandra components, followed by discovering features such as token allocation algorithm, CQL3, vnodes, lightweight transactions, and data modelling in detail. Last but not least you will get to grips with Apache Spark. By the end of this book, you'll be able to analyse big data, and build and manage high-performance databases for your application. What you will learn Write programs more efficiently using Cassandra's features more efficiently Exploit the given infrastructure, improve performance, and tweak the Java Virtual Machine (JVM) Use CQL3 in your application in order to simplify working with Cassandra Configure Cassandra and fine-tune its parameters depending on your needs Set up a cluster and learn how to scale it Monitor a Cassandra cluster in different ways Use Apache Spark and other big data processing tools Who this book is for Mastering Apache Cassandra 3.x is for you if you are a big data administrator, database administrator, architect, or developer who wants to build a high-performing, scalable, and fault-tolerant database. Prior knowledge of core concepts of databases is required.

The book is aimed at intermediate developers with an understanding of core database concepts who want to become a master at implementing Cassandra for their application.

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 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 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

Thought-provoking and accessible in approach, this updated and expanded second edition of the Learning Apache Cassandra - Manage Fault Tolerant and Scalable Real-Time Data provides a user-friendly introduction to the subject, Taking a clear structural framework, it guides the reader through the subject's core elements. A flowing writing style combines with the use of illustrations and diagrams throughout the text to ensure the reader understands even the most complex of concepts. This succinct and enlightening overview is a required reading for advanced graduate-level students. We hope you find this book useful in shaping your future career. Feel free to send us your enquiries related to our publications to info@risepress.pw Rise Press

This practical guide explains you to program and understand the power of Apache Cassandra 3.x. You will explore the integration and interaction of Cassandra components, and explore features such as the token allocation algorithm, CQL3, vnodes, lightweight transactions, and data modelling in detail.

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 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 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.

Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be 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 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

Beginning Apache Cassandra Development introduces you to one of the most robust and best-performing NoSQL database platforms on the planet. Apache Cassandra is a

document database following the JSON document model. It is specifically designed to manage large amounts of data across many commodity servers without there being any single point of failure. This design approach makes Apache Cassandra a robust and easy-to-implement platform when high availability is needed. Apache Cassandra can be used by developers in Java, PHP, Python, and JavaScript—the primary and most commonly used languages. In Beginning Apache Cassandra Development, author and Cassandra expert Vivek Mishra takes you through using Apache Cassandra from each of these primary languages. Mishra also covers the Cassandra Query Language (CQL), the Apache Cassandra analog to SQL. You'll learn to develop applications sourcing data from Cassandra, query that data, and deliver it at speed to your application's users. Cassandra is one of the leading NoSQL databases, meaning you get unparalleled throughput and performance without the sort of processing overhead that comes with traditional proprietary databases. Beginning Apache Cassandra Development will therefore help you create applications that generate search results quickly, stand up to high levels of demand, scale as your user base grows, ensure operational simplicity, and—not least—provide delightful user experiences.
Build a scalable, fault-tolerant and highly available data layer for your applications using Apache CassandraAbout This Book* Install Cassandra and use it to set up multi-node clusters* Design rich schemas that capture the relationships between different data types* Master the advanced features available in Cassandra 3.x through a step-by-step tutorial and build a scalable, high performance database layerWho This Book Is ForIf you are a first-time user of Apache Cassandra who wants to learn the basic of it, as well as some not-so-basic features, this book is for you. It does not assume any prior experience in coding or any framework.What you will learn* Install Cassandra and create your first keyspace* Create tables with multiple clustering columns to organize related data* Use secondary indexes and materialized views to avoid denormalization of data* Effortlessly handle concurrent updates with collection columns* Ensure data integrity with lightweight transactions and logged batches* Understand eventual consistency and use the right consistency level for your situation* Understand data distribution with Cassandra and get to know ways to implement application-level optimizationsIn DetailCassandra is a distributed database that stands out thanks to its robust feature set and intuitive interface, while still providing the high availability and scalability of a distributed store. This book will introduce you to the rich features offered by Cassandra, and empower you to create and manage a highly performant, fault-tolerant database layer.The book starts by explaining the new features implemented in Cassandra 3.x, you'll see how to install Cassandra, and you'll understand Lightweight Transactions. Next you'll learn to create tables with composite partition keys, and get to know different methods to avoid denormalization of data. You will then proceed to create user-defined functions and data distribution in Cassandra. Finally, you will set up a multi node cluster and implement application-level optimization using a Java client.By the end of this book, you'll be fully equipped to build powerful, scalable Cassandra database layers for your applications.
What could you do with data if scalability wasn't a problem? With this hands-on guide, you'll learn how Apache Cassandra handles hundreds of terabytes of data while remaining highly available across multiple data centers -- capabilities that have attracted Facebook, Twitter, and other data-intensive companies. Cassandra: The

Definitive Guide provides the technical details and practical examples you need to assess this database management system and put it to work in a production environment. Author Eben Hewitt demonstrates the advantages of Cassandra's nonrelational design, and pays special attention to data modeling. If you're a developer, DBA, application architect, or manager looking to solve a database scaling issue or future-proof your application, this guide shows you how to harness Cassandra's speed and flexibility. Understand the tenets of Cassandra's column-oriented structure Learn how to write, update, and read Cassandra data Discover how to add or remove nodes from the cluster as your application requires Examine a working application that translates from a relational model to Cassandra's data model Use examples for writing clients in Java, Python, and C# Use the JMX interface to monitor a cluster's usage, memory patterns, and more Tune memory settings, data storage, and caching for better performance

Learn how to integrate full-stack open source big data architecture and to choose the correct technology—Scala/Spark, Mesos, Akka, Cassandra, and Kafka—in every layer. Big data architecture is becoming a requirement for many different enterprises. So far, however, the focus has largely been on collecting, aggregating, and crunching large data sets in a timely manner. In many cases now, organizations need more than one paradigm to perform efficient analyses. Big Data SMACK explains each of the full-stack technologies and, more importantly, how to best integrate them. It provides detailed coverage of the practical benefits of these technologies and incorporates real-world examples in every situation. This book focuses on the problems and scenarios solved by the architecture, as well as the solutions provided by every technology. It covers the six main concepts of big data architecture and how integrate, replace, and reinforce every layer: The language: Scala The engine: Spark (SQL, MLib, Streaming, GraphX) The container: Mesos, Docker The view: Akka The storage: Cassandra The message broker: Kafka What You Will Learn: Make big data architecture without using complex Greek letter architectures Build a cheap but effective cluster infrastructure Make queries, reports, and graphs that business demands Manage and exploit unstructured and No-SQL data sources Use tools to monitor the performance of your architecture Integrate all technologies and decide which ones replace and which ones reinforce Who This Book Is For: Developers, data architects, and data scientists looking to integrate the most successful big data open stack architecture and to choose the correct technology in every layer

Apache Cassandra is one of the most popular NoSQL Database currently being used by many of the organization, globally in every industry like Aviation, Finance, Retail, Social Networking etc. It proves that there is quite a huge demand for certified Cassandra professionals. Having certification make your selection in the company make much easier. This certification is conducted by the DataStax(R), which has the Enterprise Version of the Apache Cassandra and Leader in providing support for the open source Apache Cassandra NoSQL database. Cassandra is one of the Unique NoSQL Database. So go for its certification, it will certainly help in-Getting the Job-Increase in your salary-Growth in your career.-Managing Tera Bytes of Data.-Learning Distributed Database-Using CQL (Cassandra Query Language)Cassandra Certification Information: -Number of questions: 60 Multiple Choice-Time allowed in minutes: 90-Required passing score: 75%-Languages: English Exam Objectives: There are in

total 5 sections and you will be asked total 60 questions in real exam. Please check each section below with regards to the exam objective1.Apache Cassandra(TM) data modeling2.Fundamentals of replication and consistency3.The distributed and internal architecture of Apache Cassandra(TM)4.Installation and configuration5.Basic toolingAbout Book: In this book, we are covering in total 75 Questions, with detailed explanation. This would help you in clearing your real time Cassandra Certification. You can use Apache Cassandra Documentation to Learn more, while going through the questions in this book. We recommend that, you understand the basic concepts behind each questions and its answers. Dont try to cram or memorize the Questions and Answers for the real exam. Having in depth understanding of the concepts will certainly help you in clearing the exam.

Build a scalable, fault-tolerant and highly available data layer for your applications using Apache Cassandra About This Book Install Cassandra and set up multi-node clusters Design rich schemas that capture the relationships between different data types Master the advanced features available in Cassandra 3.x through a step-by-step tutorial and build a scalable, high performance database layer Who This Book Is For If you are a NoSQL developer and new to Apache Cassandra who wants to learn its common as well as not-so-common features, this book is for you. Alternatively, a developer wanting to enter the world of NoSQL will find this book useful. It does not assume any prior experience in coding or any framework. What You Will Learn Install Cassandra Create keyspaces and tables with multiple clustering columns to organize related data Use secondary indexes and materialized views to avoid denormalization of data Effortlessly handle concurrent updates with collection columns Ensure data integrity with lightweight transactions and logged batches Understand eventual consistency and use the right consistency level for your situation Understand data distribution with Cassandra Develop simple application using Java driver and implement application-level optimizations In Detail Cassandra is a distributed database that stands out thanks to its robust feature set and intuitive interface, while providing high availability and scalability of a distributed data store. This book will introduce you to the rich feature set offered by Cassandra, and empower you to create and manage a highly scalable, performant and fault-tolerant database layer. The book starts by explaining the new features implemented in Cassandra 3.x and get you set up with Cassandra. Then you'll walk through data modeling in Cassandra and the rich feature set available to design a flexible schema. Next you'll learn to create tables with composite partition keys, collections and user-defined types and get to know different methods to avoid denormalization of data. You will then proceed to create user-defined functions and aggregates in Cassandra. Then, you will set up a multi node cluster and see how the dynamics of Cassandra change with it. Finally, you will implement some application-level optimizations using a Java client. By the end of this book, you'll be fully equipped to build powerful, scalable Cassandra database layers for your applications. Style and approach This book takes a step-by- step approach to give you basic to intermediate knowledge of Apache Cassandra. Every concept is explained in depth, and is supplemented with practical examples when required.

Create your own massively scalable Cassandra database with highly responsive database queries About This Book Create a Cassandra cluster and tweak its configuration to get the best performance based on your environment Analyze the key concepts and architecture of

Cassandra, which are essential to create highly responsive Cassandra databases A fast-paced and step-by-step guide on handling huge amount of data and getting the best out of your database applications Who This Book Is For If you are a developer who is working with Cassandra and you want to deep dive into the core concepts and understand Cassandra's non-relational nature, then this book is for you. A basic understanding of Cassandra is expected. What You Will Learn Install and set up your Cassandra Cluster using various installation types Use Cassandra Query Language (CQL) to design Cassandra database and tables with various configuration options Design your Cassandra database to be evenly loaded with the lowest read/write latencies Employ the available Cassandra tools to monitor and maintain a Cassandra cluster Debug CQL queries to discover why they are performing relatively slowly Choose the best-suited compaction strategy for your database based on your usage pattern Tune Cassandra based on your deployment operation system environment In Detail Apache Cassandra Essentials takes you step-by-step from from the basics of installation to advanced installation options and database design techniques. It gives you all the information you need to effectively design a well distributed and high performance database. You'll get to know about the steps that are performed by a Cassandra node when you execute a read/write query, which is essential to properly maintain of a Cassandra cluster and to debug any issues. Next, you'll discover how to integrate a Cassandra driver in your applications and perform read/write operations. Finally, you'll learn about the various tools provided by Cassandra for serviceability aspects such as logging, metrics, backup, and recovery. Style and approach This step-by-step guide is packed with examples that explain the core concepts as well as advanced concepts, techniques, and usages of Apache Cassandra.

Learn Azure in a Month of Lunches, Second Edition, is a tutorial on writing, deploying, and running applications in Azure. In it, you'll work through 21 short lessons that give you real-world experience. Each lesson includes a hands-on lab so you can try out and lock in your new skills. Summary You can be incredibly productive with Azure without mastering every feature, function, and service. Learn Azure in a Month of Lunches, Second Edition gets you up and running quickly, teaching you the most important concepts and tasks in 21 practical bite-sized lessons. As you explore the examples, exercises, and labs, you'll pick up valuable skills immediately and take your first steps to Azure mastery! This fully revised new edition covers core changes to the Azure UI, new Azure features, Azure containers, and the upgraded Azure Kubernetes Service. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Microsoft Azure is vast and powerful, offering virtual servers, application templates, and prebuilt services for everything from data storage to AI. To navigate it all, you need a trustworthy guide. In this book, Microsoft engineer and Azure trainer Iain Foulds focuses on core skills for creating cloud-based applications. About the book Learn Azure in a Month of Lunches, Second Edition, is a tutorial on writing, deploying, and running applications in Azure. In it, you'll work through 21 short lessons that give you real-world experience. Each lesson includes a hands-on lab so you can try out and lock in your new skills. What's inside Understanding Azure beyond point-and-click Securing applications and data Automating your environment Azure services for machine learning, containers, and more About the reader This book is for readers who can write and deploy simple web or client/server applications. About the author Iain Foulds is an engineer and senior content developer with Microsoft. Table of Contents PART 1 - AZURE CORE SERVICES 1 Before you begin 2 Creating a virtual machine 3 Azure Web Apps 4 Introduction to Azure Storage 5 Azure Networking basics PART 2 - HIGH AVAILABILITY AND SCALE 6 Azure Resource Manager 7 High availability and redundancy 8 Load-balancing applications 9 Applications that scale 10 Global databases with Cosmos DB 11 Managing network traffic and routing 12 Monitoring and troubleshooting PART 3 - SECURE BY DEFAULT 13 Backup, recovery, and replication 14 Data encryption 15 Securing information with Azure Key Vault 16

Azure Security Center and updates PART 4 - THE COOL STUFF 17 Machine learning and artificial intelligence 18 Azure Automation 19 Azure containers 20 Azure and the Internet of Things 21 Serverless computing

What value this book provides? This book absolutely provides tremendous value in terms its usefulness. This book takes away the pain associated with learning and mastering Cassandra. All complexity has been digested by the author and simplified for the reader with very useful and practical information that can be absorbed just by glancing through the pages. Years of author's experience and practical knowledge in Cassandra has been gifted to the reader in this book with great diligence and generosity. If you are planning to undergo expensive Cassandra training, think again, spending few hours with this book will change your mind, this book has been written with great care to reduce the learning curve. The aim of this book is multi fold, just to mention a few: Provide confidence to the reader in Cassandra concepts and architecture Provide a flexible, practical framework and context customizable for various situations Provide practical guidance to manage Cassandra platforms of various hues, sizes, shapes Provide real world examples to reduce guess work Provide executable query statements and command line statements at every step Provide practical outcomes to help the reader to gain instance understanding of what to expect Provide insights into making Cassandra environment robust and scalable Provide tricks and tips to implement and maintain seamlessly Provide security and vulnerability mitigation tips and steps Provide best practices to follow for optimal Cassandra use There is no doubt - this book makes the reader very productive Cassandra professional in very short span of time. This book essentially bridges the training gap as the industry is moving fast to take full advantage of what Cassandra can offer to fulfill emerging business needs. This book can be very helpful to Data administrators, Developers, Data modelers/Application Architects, Project Managers and Data Custodians. This book has range of topics, many are listed here: Cassandra concepts and architecture Cassandra Installation and Configuration Cassandra system architecture depicting gossip protocol, replication, consistency, tombstones, hinted handoff, compactions, repairs, memtables, commit log, read and write path functions Cassandra oriented data modelling Cassandra QL (CQL) tutorial Handling of Primary and Partition keys in Cassandra covering No joins, Static columns and TTL aspects Configuring authentication, authorization to access Cassandra in addition, steps to set up node-node and client-node SSL Configuring nodes addition, removal, decommission in single token and vnode setup modes in Cassandra Instructions to add new data center and delink the existing data center from a multi-dc cluster arrangement Cassandra backup and recovery functions with real examples of restoring tables after truncation events Cassandra utilities such as sstabledump, sstablemetadata, sstablesplit, cqlsh and cassandra-stress Troubleshooting methods such as Node down, Read latency and Recover truncated table Upgrading Cassandra to higher versions Additional Cassandra architecture II methods such as Read and Write path, Compactions and Repairs

This updated and expanded second edition of the Learning Apache Cassandra - Manage Fault Tolerant and Scalable Real-Time Data provides a user-friendly introduction to the subject Taking a clear structural framework, it guides the reader through the subject's core elements. A flowing writing style combines with the use of illustrations and diagrams throughout the text to ensure the reader understands even the most complex of concepts. This succinct and enlightening overview is a required reading for all those interested in the subject . We hope you find this book useful in shaping your future career & Business.

This book is a step by step beginners guide to learning Cassandra. The book uses tons of charts, graphs, images and code to aid your Cassandra learning. The book gives a detailed introduction to Cassandra. It proceeds to give step-by-step instructions to installing Cassandra. Cassandra Architecture and Replication Factor Strategy is lucidly explained. Data Modelling, Keyspace CQL are also described in detail. The book will teach you enough to get started with

Cassandra. Here is what is included Chapter 1: Introduction Cassandra History Nosql Cassandra Database Nosql Cassandra Database Vs Relational databases Apache Cassandra Features Cassandra Use Cases Chapter 2: Download and Install Prerequisite for Apache Cassandra Installation How to Download and Install Cassandra Chapter 3: Architecture Components of Cassandra Data Replication Write Operation Read Operation Chapter 4: Data Model and Rules Cassandra Data Model Rules Model Your Data in Cassandra Handling One to One Relationship Handling one to many Relationship Handling Many to Many Relationship Chapter 5: Cassandra CQL Create, Alter & Drop Keyspace Cassandra Table: Create, Alter, Drop & Truncate Cassandra Query Language(CQL): Insert, Update, Delete, Read Data Create & Drop INDEX Data Types & Expiration SET, LIST & MAP Chapter 6: Cassandra Cluster Prerequisites for Cassandra Cluster Enterprise Edition Installation Starting Cassandra Node Chapter 7: DevCenter & OpsCenter Installation DevCenter Installation OpsCenter Installation Chapter 8: Security What is Internal Authentication and Authorization Configure Authentication and Authorization Logging in Create New User Authorization Configuring Firewall Enabling JMX Authentication ???Download Free - For Kindle Unlimited Subscribers!???

Learning Apache CassandraPackt Publishing Ltd

Python makes machine learning easy for beginners and experienced developers With computing power increasing exponentially and costs decreasing at the same time, there is no better time to learn machine learning using Python. Machine learning tasks that once required enormous processing power are now possible on desktop machines. However, machine learning is not for the faint of heart—it requires a good foundation in statistics, as well as programming knowledge. Python Machine Learning will help coders of all levels master one of the most in-demand programming skillsets in use today. Readers will get started by following fundamental topics such as an introduction to Machine Learning and Data Science. For each learning algorithm, readers will use a real-life scenario to show how Python is used to solve the problem at hand. • Python data science—manipulating data and data visualization • Data cleansing • Understanding Machine learning algorithms • Supervised learning algorithms • Unsupervised learning algorithms • Deploying machine learning models Python Machine Learning is essential reading for students, developers, or anyone with a keen interest in taking their coding skills to the next level.

Written in a clear, step-by-step manner, this 400-page course provides an excellent starting point for people that want to get into Apache Cassandra and learn best by doing. A high-quality, project-based, hands-on training courseware book, Apache Cassandra Hands-On Training Level One is designed to be used as the student book for a 2-day introductory level Cassandra course delivered by a Cassandra instructor. Having said that, this book can also be done as a self-paced training course. Recommended prerequisites for this training book are experience with databases, SQL, and programming. This hands-on training book takes people through the basics of working with Cassandra as they learn how to install Cassandra, create a database, create tables, insert, update, and delete data, and create an application, as well as create and modify a multiple-node cluster. Unit 1: Understanding What Cassandra is For Unit 2: Getting Started with the Architecture Unit 3: Installing Cassandra Unit 4: Communicating with Cassandra Unit 5: Creating a Database Unit 6: Creating a Table Unit 7: Inserting Data Unit 8: Modeling Data Unit 9: Creating an Application Unit 10: Updating and Deleting Data Unit 11: Selecting Hardware Unit 12: Adding Nodes to a Cluster Unit 13: Repairing Nodes Unit 14: Removing a Node Unit 15: Monitoring a Cluster Unit 16: Adding a Data Center As virtual machine images are used extensively throughout this hands-on course, including for the creation of a multiple-node Cassandra cluster, any computer used for the exercises in this course needs to be relatively high spec. Specifically, a computer with the following is needed: 64-bit operating system (Mac, Windows, or Linux) 8GB (or more) of RAM 30GB (or more) of free hard drive space Latest version of VMware Player installed and working A way to unzip

files Acrobat Reader (or equivalent, for viewing a PDF file) For the full outline, and class files download, see ruthstryker.com/books/achotl1. For a sample unit, see ruthstryker.com/books/achotl1/achotl1_ch06_20140717.pdf (Unit 6) or ruthstryker.com/books/achotl1/achotl1_ch15_20140717.pdf (Unit 15). For the setup steps, see ruthstryker.com/books/achotl1/achotl1_apC_20140722.pdf (Appendix C). Student comments about the book: "Excellent starter course that has taken me from knowing nothing of Cassandra to feeling confident in setting up and using it." "Level covered in book is just right." "Course material was good. It had a wide range of labs and was very helpful in understanding the agenda." "Course material was well-written and easy to follow." "Excellent introduction into Cassandra filled with hands-on exercises for all topics." "Material covers the basics quite well." Follow this handbook to build, configure, tune, and secure Apache Cassandra databases. Start with the installation of Cassandra and move on to the creation of a single instance, and then a cluster of Cassandra databases. Cassandra is increasingly a key player in many big data environments, and this book shows you how to use Cassandra with Apache Spark, a popular big data processing framework. Also covered are day-to-day topics of importance such as the backup and recovery of Cassandra databases, using the right compression and compaction strategies, and loading and unloading data. Expert Apache Cassandra Administration provides numerous step-by-step examples starting with the basics of a Cassandra database, and going all the way through backup and recovery, performance optimization, and monitoring and securing the data. The book serves as an authoritative and comprehensive guide to the building and management of simple to complex Cassandra databases. The book: Takes you through building a Cassandra database from installation of the software and creation of a single database, through to complex clusters and data centers Provides numerous examples of actual commands in a real-life Cassandra environment that show how to confidently configure, manage, troubleshoot, and tune Cassandra databases Shows how to use the Cassandra configuration properties to build a highly stable, available, and secure Cassandra database that always operates at peak efficiency What You'll Learn Install the Cassandra software and create your first database Understand the Cassandra data model, and the internal architecture of a Cassandra database Create your own Cassandra cluster, step-by-step Run a Cassandra cluster on Docker Work with Apache Spark by connecting to a Cassandra database Deploy Cassandra clusters in your data center, or on Amazon EC2 instances Back up and restore mission-critical Cassandra databases Monitor, troubleshoot, and tune production Cassandra databases, and cut your spending on resources such as memory, servers, and storage Who This Book Is For Database administrators, developers, and architects who are looking for an authoritative and comprehensive single volume for all their Cassandra administration needs. Also for administrators who are tasked with setting up and maintaining highly reliable and high-performing Cassandra databases. An excellent choice for big data administrators, database administrators, architects, and developers who use Cassandra as their key data store, to support high volume online transactions, or as a decentralized, elastic data store.

Over 150 recipes to design and optimize large scale Apache Cassandra deployments.

This book will introduce Redis and help you understand its various facets. Starting with an introduction to NoSQL, you will learn how to install Redis and how to classify and work with data structures. By working with real world scenarios pertaining to using Redis, you will discover sharding and indexing techniques, along with how to improve scalability and performance through persistent strategies and data migration techniques. With the help of multiple

examples, you will learn to design web and business applications. You will also learn how to configure Redis for setting up clusters and tuning it for performance. At the end of this book, you will find essential tips on backup and recovery strategies for the Redis environment.

Learn Big Data from the ground up with this complete and up-to-date resource from leaders in the field Big Data: Concepts, Technology, and Architecture delivers a comprehensive treatment of Big Data tools, terminology, and technology perfectly suited to a wide range of business professionals, academic researchers, and students. Beginning with a fulsome overview of what we mean when we say, "Big Data," the book moves on to discuss every stage of the lifecycle of Big Data. You'll learn about the creation of structured, unstructured, and semi-structured data, data storage solutions, traditional database solutions like SQL, data processing, data analytics, machine learning, and data mining. You'll also discover how specific technologies like Apache Hadoop, SQOOP, and Flume work. Big Data also covers the central topic of big data visualization with Tableau, and you'll learn how to create scatter plots, histograms, bar, line, and pie charts with that software. Accessibly organized, Big Data includes illuminating case studies throughout the material, showing you how the included concepts have been applied in real-world settings. Some of those concepts include: The common challenges facing big data technology and technologists, like data heterogeneity and incompleteness, data volume and velocity, storage limitations, and privacy concerns Relational and non-relational databases, like RDBMS, NoSQL, and NewSQL databases Virtualizing Big Data through encapsulation, partitioning, and isolating, as well as big data server virtualization Apache software, including Hadoop, Cassandra, Avro, Pig, Mahout, Oozie, and Hive The Big Data analytics lifecycle, including business case evaluation, data preparation, extraction, transformation, analysis, and visualization Perfect for data scientists, data engineers, and database managers, Big Data also belongs on the bookshelves of business intelligence analysts who are required to make decisions based on large volumes of information. Executives and managers who lead teams responsible for keeping or understanding large datasets will also benefit from this book.

This book is a practical, hands-on guide, taking the reader from the basics of using Cassandra though to the installation and the running.Learning Cassandra for Administrators is for administrators who manage a large deployment of Cassandra clusters, and support engineers who would like to install the monitoring tools and who are also in charge of making sure the cluster stays the same, ensuring that the service is always up and running.

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 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

Copyright: cc319e0ae18e2a2b254cf98a08a505f3