

[Read and download] Learning Big Data with Amazon Elastic MapReduce

Learning Big Data with Amazon Elastic MapReduce

Von Amarkant Singh, Vijay Rayapati
audiobook / *ebooks / Download PDF / ePub / DOC





Learning Big Data with Amazon Elastic MapReduce

Easily learn, build, and execute real-world Big Data solutions using Hadoop and AWS EMR

Amarkant Singh Vijay Rayapati

[PACKT] enterprise
PUBLISHING professional expertise distilled

 Download

 Read Online

Produktinformation -Verkaufsrang: #1193265 in eBooksVerffentlicht am: 2014-10-30Erscheinungsdatum: 2014-10-30File Name: B00P2ER354 | File size: 40.Mb

Von Amarkant Singh, Vijay Rayapati : Learning Big Data with Amazon Elastic MapReduce before purchasing it in order to gage whether or not it would be worth my time, and all praised Learning Big Data with Amazon Elastic MapReduce:

KurzbeschreibungEasily learn, build, and execute real-world Big Data solutions using Hadoop and AWS EMRAbout This BookLearn how to solve big data problems using Apache HadoopUse Elastic MapReduce to create and maintain cluster infrastructure for big data analyticsA step-by-step guide exploring the vast set of services provided by on the cloudWho This Book Is ForThis book is aimed at developers and system administrators who want to learn about Big Data analysis using Elastic MapReduce. Basic Java programming knowledge is required. You should be comfortable

with using command-line tools. Prior knowledge of AWS, API, and CLI tools is not assumed. Also, no exposure to Hadoop and MapReduce is expected.

What You Will Learn

- Create and access your account on AWS and learn about its various services
- Launch a machine on the cloud infrastructure of AWS, get login credentials, and communicate with that machine
- Learn about the logical dataflow of MapReduce and how it uses distributed computing effectively
- Understand the benefits of EMR over a local Hadoop cluster
- Discover the best practices that should be kept in mind while planning and executing a cluster/job on EMR
- Launch a cluster on EMR, submit the Hello World wordcount job for processing, and download and view the results
- Execute jobs on EMR using the two primary methods provided by EMR

In Detail Elastic MapReduce is a web service used to process and store vast amount of data, and it is one of the largest Hadoop operators in the world. With the increase in the amount of data generated and collected by many businesses and the arrival of cost-effective cloud-based solutions for distributed computing, the feasibility to crunch large amounts of data to get deep insights within a short span of time has increased greatly.

This book will get you started with AWS so that you can quickly create your own account and explore the services provided, many of which you might be delighted to use. This book covers the architectural details of the MapReduce framework, Apache Hadoop, various job models on EMR, how to manage clusters on EMR, and the command-line tools available with EMR. Each chapter builds on the knowledge of the previous one, leading to the final chapter where you will learn about solving a real-world use case using Apache Hadoop and EMR. This book will, therefore, get you up and running with major Big Data technologies quickly and efficiently.

Kurzbeschreibung Easily learn, build, and execute real-world Big Data solutions using Hadoop and AWS EMR

About This Book Learn how to solve big data problems using Apache Hadoop

Use Elastic MapReduce to create and maintain cluster infrastructure for big data analytics

A step-by-step guide exploring the vast set of services provided by on the cloud

Who This Book Is For This book is aimed at developers and system administrators who want to learn about Big Data analysis using Elastic MapReduce. Basic Java programming knowledge is required. You should be comfortable with using command-line tools. Prior knowledge of AWS, API, and CLI tools is not assumed. Also, no exposure to Hadoop and MapReduce is expected.

What You Will Learn

- Create and access your account on AWS and learn about its various services
- Launch a machine on the cloud infrastructure of AWS, get login credentials, and communicate with that machine
- Learn about the logical dataflow of MapReduce and how it uses distributed computing effectively
- Understand the benefits of EMR over a local Hadoop cluster
- Discover the best practices that should be kept in mind while planning and executing a cluster/job on EMR
- Launch a cluster on EMR, submit the Hello World wordcount job for processing, and download and view the results
- Execute jobs on EMR using the two primary methods provided by EMR

In Detail Elastic MapReduce is a web service used to process and store vast amount of data, and it is one of the largest Hadoop operators in the world. With the increase in the amount of data generated and collected by many businesses and the arrival of cost-effective cloud-based solutions for distributed computing, the feasibility to crunch large amounts of data to get deep insights within a short span of time has increased greatly.

This book will get you started with AWS so that you can quickly create your own account and explore the services provided, many of which you might be delighted to use. This book covers the architectural details of the MapReduce framework, Apache Hadoop, various job models on EMR, how to manage clusters on EMR, and the command-line tools available with EMR. Each chapter builds on the knowledge of the previous one, leading to the final chapter where you will learn about solving a real-world use case using Apache Hadoop and EMR. This book will, therefore, get you up and running with major Big Data technologies quickly and efficiently.

ber den Autor und weitere Mitwirkende

Amarkant Singh Amarkant Singh is a Big Data specialist. Being one of the initial users of Elastic MapReduce, he has used it extensively to build and deploy many Big Data solutions. He has been working with Apache Hadoop and EMR for almost 4 years now. He is also a certified AWS Solutions Architect. As an engineer, he has designed and developed enterprise applications of various scales. He is currently leading the product development team at one of the most happening cloud-based enterprises in the Asia-Pacific region. He is also an all-time top user on Stack Overflow for EMR at the time of writing this book. He blogs at <http://www.bigdataspeak.com/> and is active on Twitter as @singh_amarkant.

Vijay Rayapati Vijay Rayapati is the CEO of Minjar Cloud Solutions Pvt. Ltd., one of the leading providers of cloud and Big Data solutions on public cloud platforms. He has over 10 years of experience in building business rule engines, data analytics platforms, and real-time analysis systems used by many leading enterprises across the world, including Fortune 500 businesses. He has worked on various technologies such as LISP, .NET, Java, Python, and many NoSQL databases. He has rearchitected and led the initial development of a large-scale location intelligence and analytics platform using Hadoop and AWS EMR. He has worked with many ad networks, e-commerce, financial, and retail companies to help them design, implement, and scale their data analysis and BI platforms on the AWS Cloud. He is passionate about open source software, large-scale systems, and performance engineering. He is active on Twitter as @amnigos, he blogs at amnigos.com, and his GitHub profile is <https://github.com/amnigos>.