第一图书网, tushu007.com

<<Hadoop权威指南>>

图书基本信息

书名：<<Hadoop权威指南>>

13位ISBN编号：9787564138936

10位ISBN编号：7564138939

出版时间：2013-1

出版时间：东南大学出版社

作者：怀特

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：http://www.tushu007.com

内容概要

《Hadoop权威指南(影印版)(第3版)(修订版)》的内容包括：使用Hadoop分布式文件系统（HDFS）保存大数据集；使用MapReduce运行分布式计算；使用Hadoop的数据和I／O构件实现压缩、数据完整性、序列化（包括Avro）和持久化；了解常见的陷阱和高级特性，以编写实用的MapReduce程序；设计、构建和管理专用的Hadoop集群——或者在云中运行Hadoop；使用Sqoop从关系型数据库载入数据到HDFS；使用Pig查询语言进行大规模数据处理；使用Hadoop的数据仓库系统Hive分析数据集；利用HBase处理结构化和半结构化数据，以及利用ZooKeeper构建分布式系统。

作者简介

作者：（美国）怀特（White T.） 怀特（White T.），Cloudera工程师兼Apache软件基金会成员，自2007年2月起成为ApacheHadoop代码提交者。
他为oreilly网、java网和IBMdeveloperWorks撰写了大量文章，还经常在业界会议上作Hadoop主题演讲。

书籍目录

Foreword Preface 1. Meet Hadoop Data! Data Storage and Analysis Comparison with Other Systems Rational Database Management System Grid Computing Volunteer Computing A Brief History of Hadoop Apache Hadoop and the Hadoop Ecosystem Hadoop Releases What's Covered in This Book Compatibility 2. MapReduce A Weather Dataset Data Format Analyzing the Data with Unix Tools Analyzing the Data with Hadoop Map and Reduce Java MapReduce Scaling Out Data Flow Combiner Functions Running a Distributed MapReduce Job Hadoop Streaming Ruby Python Hadoop Pipes Compiling and Running 3. The Hadoop Distributed Filesystem The Design of HDFS HDFS Concepts Blocks Namenodes and Datanodes HDFS Federation HDFS High-Availability The Command-Line Interface Basic Filesystem Operations Hadoop Filesystems Interfaces The Java Interface Reading Data from a Hadoop URL Reading Data Using the FileSystem API Writing Data Directories Querying the Filesystem Deleting Data Data Flow Anatomy of a File Read Anatomy of a File Write Coherency Model Data Ingest with Flume and Sqoop Parallel Copying with distcp Keeping an HDFS Cluster Balanced Hadoop Archives Using Hadoop Archives Limitations 4. Hadoop I/O Data Integrity Data Integrity in HDFS LocalFileSystem ChecksumFileSystem Compression Codecs Compression and Input Splits Using Compression in MapReduce Serialization The Writable Interface Writable Classes Implementing a Custom Writable Serialization Frameworks Avro Avro Data Types and Schemas In-Memory Serialization and Deserialization Avro Datafiles Interoperability Schema Resolution Sort Order Avro MapReduce Sorting Using Avro MapReduce Avro MapReduce in Other Languages File-Based Data Structures SequenceFile MapFile 5. Developing a MapReduce Application The Configuration API Combining Resources Variable Expansion Setting Up the Development Environment Managing Configuration GenericOptionsParser, Tool, and ToolRunner Writing a Unit Test with MRUnit Mapper Reducer Running Locally on Test Data Running a Job in a Local Job Runner Testing the Driver Running on a Cluster Packaging a Job Launching a Job The MapReduce Web UI Retrieving the Results Debugging a Job Hadoop Logs Remote Debugging Tuning a Job Profiling Tasks MapReduce Workflows Decomposing a Problem into MapReduce Jobs JobControl Apache Oozie 6. How MapReduce Works Anatomy of a MapReduce Job Run Classic MapReduce (MapReduce 1) YARN (MapReduce 2) Failures Failures in Classic MapReduce Failures in YARN Job Scheduling The Fair Scheduler The Capacity Scheduler Shuffle and Sort The Map Side The Reduce Side Configuration Tuning Task Execution The Task Execution Environment Speculative Execution Output Committers Task JVM Reuse Skipping Bad Records 7. MapReduceTypes and Formats MapReduce Types The Default MapReduce Job Input Formats Input Splits and Records Text Input Binary Input Multiple Inputs Database Input (and Output) Output Formats Text Output Binary Output Multiple Outputs Lazy Output Database Output 8. MapReduce Features Counters Built-in Counters User-Defined Java Counters …… 9. Settinq Up a Hadoop Cluster 10. Administering Hadoop 11. Pig 12. Hive 13. HBase 14. ZooKeeper 15. Sqoop 16. Case Studies A. Installing Apache Hadoop B. Cloudera's Distribution Including Apache Hadoop C. Preparing the NCDC Weather Data Index

章节摘录

版权页： Furthermore, blocks fit well with replication for providing fault tolerance and availa-bility. To insure against corrupted blocks and disk and machine failure, each block is replicated to a small number of physically separate machines (typically three). If a block becomes unavailable, a copy can be read from another location in a way that is trans-parent to the client. A block that is no longer available due to corruption or machine failure can be replicated from its alternative locations to other live machines to bring the replication factor back to the normal level. (See "Data Integrity" on page 81 for more on guarding against corrupt data.) Similarly, some applications may choose to set a high replication factor for the blocks in a popular file to spread the read load on the cluster. Like its disk filesystem cousin, HDFS's fsck command understands blocks. For exam-ple, running: hadoop fsck / -files -blocks will list the blocks that make up each file in the filesystem. (See also "Filesystem check (fsck)" on page 347.) Namenodes and Datanodes An HDFS cluster has two types of nodes operating in a master-worker pattern: a name-node (the master) and a number of datanodes (workers). The namenode manages the filesystem namespace. It maintains the filesystem tree and the metadata for all the files and directories in the tree. This information is stored persistently on the local disk in the form of two files: the namespace image and the edit log. The namenode also knows the datanodes on which all the blocks for a given file are located; however, it does not store block locations persistently, because this information is reconstructed from datanodes when the system starts. A client accesses the filesystem on behalf of the user by communicating with the name-node and datanodes. The client presents a filesystem interface similar to a Portable Operating System Interface (POSIX), so the user code does not need to know about the namenode and datanode to function.

编辑推荐

《Hadoop权威指南(影印版)(第3版)(修订版)》中你可以找到富有启发意义的实际案例分析，它们展示了Hadoop用于解决特定问题的各种方式。
第三版涵盖了Hadoop近期更新内容，包括新的MapReduce API相关资料，还有MapReduce2及其更加灵活的执行模型（YARN）。

第一图书网, tushu007.com

<<Hadoop权威指南>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:http://www.tushu007.com