

<<Hive编程指南>>

图书基本信息

书名：<<Hive编程指南>>

13位ISBN编号：9787115333834

10位ISBN编号：7115333831

出版时间：2013-12-1

出版时间：人民邮电出版社

作者：卡普廖洛 (Edward Capriolo), 万普勒 (Dean Wampler), 卢森格林 (Jason Rutherglen)

译者：曹坤

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>

<<Hive编程指南>>

内容概要

市场中第一本Hive图书。

Hive在Hadoop系统中的应用趋势比较可观。

<<Hive编程指南>>

作者简介

Edward Capriolo : Media6degrees公司系统管理员，他是Apache软件基金会成员，还是Hadoop-Hive项目成员。

Dean Wampler : Think Big Analytics公司总顾问，对大数据问题以及Hadoop和机器学习有专门的研究。

Jason Rutherglen : Think Big Analytics公司软件架构师，对大数据、Hadoop、搜索和安全有专门的研究。

。

<<Hive编程指南>>

书籍目录

第1章基础知识

1.1Hadoop和MapReduce综述

1.2Hadoop生态系统中的Hive

1.2.1Pig

1.2.2HBase

1.2.3Cascading、Crunch及其他

1.3Java和Hive：词频统计算法

1.4后续事情

第2章基础操作

2.1安装预先配置好的虚拟机

2.2安装详细步骤

2.2.1装Java

2.2.2安装Hadoop

2.2.3本地模式、伪分布式模式和分布式模式

2.2.4测试Hadoop

2.2.5安装Hive

2.3Hive内部是什么

2.4启动Hive

2.5配置Hadoop环境

2.5.1本地模式配置

2.5.2分布式模式和伪分布式模式配置

2.5.3使用JDBC连接元数据

2.6Hive命令

2.7命令行界面

2.7.1CLI选项

2.7.2变量和属性

2.7.3Hive中“一次使用”命令

2.7.4从文件中执行Hive查询

2.7.5hiverc文件

2.7.6使用HiveCLI的更多介绍

2.7.7查看操作命令历史

2.7.8执行shell命令

2.7.9在Hive内使用Hadoop的dfs命令

2.7.10Hive脚本中如何进行注释

2.7.11显示字段名称

第3章数据类型和文件格式

3.1基本数据类型

3.2集合数据类型

3.3文本文件数据编码

3.4读时模式

第4章HiveQL：数据定义

4.1Hive中的数据库

4.2修改数据库

4.3创建表

4.3.1管理表

<<Hive编程指南>>

4.3.2外部表

4.4分区表、管理表

4.4.1外部分区表

4.4.2自定义表的存储格式

4.5删除表

4.6修改表

4.6.1表重命名

4.6.2增加、修改和删除表分区

4.6.3修改列信息

4.6.4增加列

4.6.5删除或者替换列

4.6.6修改表属性

4.6.7修改存储属性

4.6.8众多的修改表语句

第5章HiveQL：数据操作

5.1向管理表中装载数据

5.2通过查询语句向表中插入数据

5.3单个查询语句中创建表并加载数据

5.4导出数据

第6章HiveQL：查询

6.1SELECT...FROM语句

6.1.1使用正则表达式来指定列

6.1.2使用列值进行计算

6.1.3算术运算符

6.1.4使用函数

6.1.5LIMIT语句

6.1.6列别名

6.1.7嵌套SELECT语句

6.1.8CASE...WHEN...THEN句式

6.1.9什么情况下Hive可以避免进行MapReduce

6.2WHERE语句

6.2.1谓词操作符

6.2.2关于浮点数比较

6.2.3LIKE和RLIKE

6.3GROUPBY语句

6.4JOIN语句

6.4.1INNERJOIN

6.4.2JOIN优化

6.4.3LEFTOUTERJOIN

6.4.4OUTERJOIN

6.4.5RIGHTOUTERJOIN

6.4.6FULLOUTERJOIN

6.4.7LEFTSEMI—JOIN

6.4.8笛卡尔积JOIN

6.4.9map—sideJOIN

6.5ORDERBY和SORTBY

6.6含有SORTBY的DISTRIBUTEY

<<Hive编程指南>>

6.7 CLUSTER BY

6.8 类型转换

6.9 抽样查询

6.9.1 数据块抽样

6.9.2 分桶表的输入裁剪

6.10 UNION ALL

第7章 HiveQL：视图

7.1 使用视图来降低查询复杂度

7.2 使用视图来限制基于条件过滤的数据

7.3 动态分区中的视图和map类型

7.4 视图零零碎碎相关的事情

第8章 HiveQL：索引

8.1 创建索引

8.2 重建索引

8.3 显示索引

8.4 删除索引

8.5 实现一个定制化的索引处理器

第9章 模式设计

9.1 按天划分的表

9.2 关于分区

9.3 唯一键和标准化

9.4 同一份数据多种处理

9.5 对于每个表的分区

9.6 分桶表数据存储

9.7 为表增加列

9.8 使用列存储表

9.8.1 重复数据

9.8.2 多列

9.9 (几乎) 总是使用压缩

第10章 调优

10.1 使用EXPLAIN

10.2 EXPLAIN EXTENDED

10.3 限制调整

10.4 JOIN 优化

10.5 本地模式

10.6 并行执行

10.7 严格模式

10.8 调整mapper和reducer个数

10.9 JVM 重用

10.10 索引

10.11 动态分区调整

10.12 推测执行

10.13 单个MapReduce中多个GROUP BY

10.14 虚拟列

第11章 其他文件格式和压缩方法

11.1 确定安装编解码器

11.2 选择一种压缩编 / 解码器

<<Hive编程指南>>

11.3开启中间压缩

11.4最终输出结果压缩

11.5sequencefile存储格式

11.6使用压缩实践

11.7存档分区

11.8压缩：包扎

.....

第12章开发

第13章函数

第14章Streaming

第15章自定义Hive文件和记录格式

第16章Hive的Thrift服务

第17章存储处理程序和NoSQL

第18章安全

第19章锁

第20章Hive和Oozie整合

第21章Hive和亚马逊网络服务系统（AWS）

第22章HCatalog

第23章案例研究

术语词汇表

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>