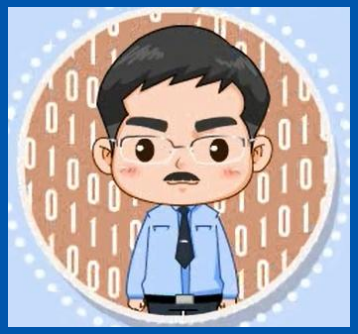


《大数据基础编程、实验和案例教程（第2版）》

教材官网：
http://dblab.xmu.edu.cn/post/bigdatapractice2/

温馨提示：编辑幻灯片母版，可以修改每页PPT的厦大校徽和底部文字

第8章 数据仓库Hive的安装和使用

（PPT版本号：2020年12月版本）

扫一扫访问教材官网

林子雨

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn

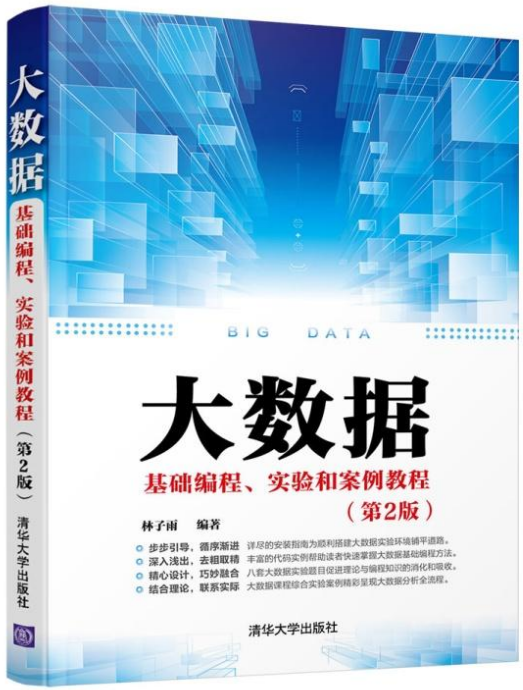
主页：http://dblab.xmu.edu.cn/linziyu

教材简介

本书是与《大数据技术原理与应用（第3版）》教材配套的唯一指定实验指导书

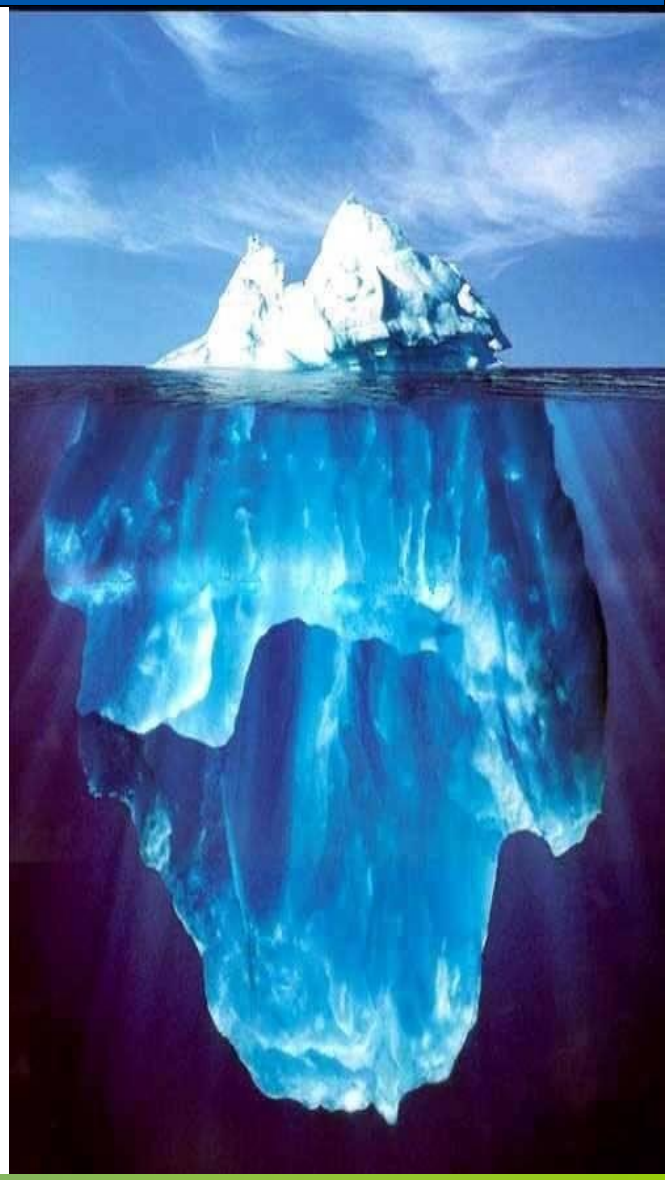
林子雨编著《大数据基础编程、实验和案例教程（第2版）》
清华大学出版社 ISBN:978-7-302-55977-1 定价：69元，2020年10月第2版
教材官网： http://dblab.xmu.edu.cn/post/bigdatapractice2/

扫一扫访问
教材官网



- 步步引导，循序渐进，详尽的安装指南为顺利搭建大数据实验环境铺平道路
- 深入浅出，去粗取精，丰富的代码实例帮助快速掌握大数据基础编程方法
- 精心设计，巧妙融合，八套大数据实验题目促进理论与编程知识的消化和吸收
- 结合理论，联系实际，大数据课程综合实验案例精彩呈现大数据分析全流程

提纲



8.1 Hive的安装

8.1.1 下载安装文件
8.1.2 配置环境变量
8.1.3 修改配置文件
8.1.4 安装并配置MySQL

8.1.1 下载安装文件

访问Hive官网（http://www.apache.org/dyn/closer.cgi/hive/）下载安装文件apache-hive-3.1.2-bin.tar.gz

下载完安装文件以后，需要对文件进行解压。按照Linux系统使用的默认规范，用户安装的软件一般都是存放在“/usr/local/”目录下。请在Linux系统中打开一个终端，执行如下命令：

```
$ sudo tar -zxvf ./apache-hive-3.1.2-bin.tar.gz -C /usr/local   # 解压到/usr/local中
$ cd /usr/local/
$ sudo mv apache-hive-3.1.2-bin hive      # 将文件夹名改为hive
$ sudo chown -R hadoop:hadoop hive        # 修改文件权限
```

8.1.2 配置环境变量

为了方便使用，可以把hive命令加入到环境变量PATH中，从而可以在任意目录下直接使用hive命令启动，请使用vim编辑器打开“~/.bashrc”文件进行编辑，命令如下：

```
$ vim ~/.bashrc
```

在该文件的最前面一行添加如下内容：

```
export HIVE_HOME=/usr/local/hive
export PATH=$PATH:$HIVE_HOME/bin
```

保存该文件并退出vim编辑器，然后，运行如下命令使得配置立即生效：

```
$ source ~/.bashrc
```

8.1.3 修改配置文件

将“/usr/local/hive/conf”目录下的hive-default.xml.template文件重命名为hive-default.xml，命令如下：

```
$ cd /usr/local/hive/conf
$ sudo mv hive-default.xml.template hive-default.xml
```

同时，使用vim编辑器新建一个文件hive-site.xml，命令如下：

```
$ cd /usr/local/hive/conf
$ vim hive-site.xml
```

8.1.3 修改配置文件

在hive-site.xml中输入如下配置信息：

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
  <property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:mysql://localhost:3306/hive?createDatabaseIfNotExist=true</value>
    <description>JDBC connect string for a JDBC metastore</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionDriverName</name>
    <value>com.mysql.jdbc.Driver</value>
    <description>Driver class name for a JDBC metastore</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionUserName</name>
    <value>hive</value>
    <description>username to use against metastore database</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionPassword</name>
    <value>hive</value>
    <description>password to use against metastore database</description>
  </property>
</configuration>
```

8.1.4 安装并配置MySQL

1. 安装MySQL

这里采用MySQL数据库保存Hive的元数据，而不是采用Hive自带的derby来存储元数据，因此，需要安装MySQL数据库。可以参照“附录B：Linux系统中的MySQL安装及常用操作”，完成MySQL数据库的安装，这里不再赘述。

8.1.4 安装并配置MySQL

2. 下载MySQL JDBC驱动程序

为了让Hive能够连接到MySQL数据库，需要下载MySQL JDBC驱动程序。可以到MySQL官网（http://www.mysql.com/downloads/connector/j/）下载mysql-connector-java-5.1.40.tar.gz。

在Linux系统中打开一个终端，在终端中执行如下命令解压缩文件：

```
$ cd ~
$ tar -zxvf mysql-connector-java-5.1.40.tar.gz   #解压
$ #下面将mysql-connector-java-5.1.40-bin.jar拷贝到/usr/local/hive/lib目录下
$ cp mysql-connector-java-5.1.40/mysql-connector-java-5.1.40-bin.jar  /usr/local/hive/lib
```

8.1.4 安装并配置MySQL

3. 启动MySQL

执行如下命令启动MySQL，并进入“mysql>”命令提示符状态：

```
$ service mysql start  #启动MySQL服务
$ mysql -u root -p   #登录MySQL数据库
```

4. 在MySQL中为Hive新建数据库

现在，需要在MySQL数据库中新建一个名称为hive的数据库，用来保存Hive的元数据。MySQL中的这个hive数据库，是与Hive的配置文件hive-site.xml中的“mysql://localhost:3306/hive”对应起来的，用来保存Hive元数据。在MySQL数据库中新建hive数据库的命令，需要在“mysql>”命令提示符下执行，具体如下：

```
mysql> create database hive;
```

8.1.4 安装并配置MySQL

5. 配置MySQL允许Hive接入

需要对MySQL进行权限配置，允许Hive连接到MySQL。

```
mysql> grant all on *.* to hive@localhost identified by 'hive';
mysql> flush privileges;
```

6. 启动Hive

Hive是基于Hadoop的数据仓库，会把用户输入的查询语句自动转换成为MapReduce任务来执行，并把结果返回给用户。因此，启动Hive之前，需要先启动Hadoop集群，命令如下：

```
$ cd /usr/local/hadoop
$ ./sbin/start-dfs.sh
```

8.1.4 安装并配置MySQL

然后，再执行如下命令启动Hive：

```
$ cd /usr/local/hive
$ ./bin/hive
```

8.2 Hive的数据类型

表 Hive的基本数据类型

类型	描述	示例
TINYINT	1个字节（8位）有符号整数	1
SMALLINT	2个字节（16位）有符号整数	1
INT	4个字节（32位）有符号整数	1
BIGINT	8个字节（64位）有符号整数	1
FLOAT	4个字节（32位）单精度浮点数	1.0
DOUBLE	8个字节（64位）双精度浮点数	1.0
BOOLEAN	布尔类型，true/false	true
STRING	字符串，可以指定字符集	“xmu”
TIMESTAMP	整数、浮点数或者字符串	1327882394（Unix新纪元秒）
BINARY	字节数组	[0,1,0,1,0,1,0,1]

8.2 Hive的数据类型

表 Hive的集合数据类型

类型	描述	示例
ARRAY	一组有序字段，字段的类型必须相同	Array(1,2)
MAP	一组无序的键/值对，键的类型必须是原子的，值可以是任何数据类型，同一个映射的键和值的类型必须相同	Map('a',1,'b',2)
STRUCT	一组命名的字段，字段类型可以不同	Struct('a',1,1,0)

8.3 Hive基本操作

8.3.1 创建数据库、表、视图
8.3.2 删除数据库、表、视图
8.3.3 修改数据库、表、视图
8.3.4 查看数据库、表、视图
8.3.5 描述数据库、表、视图
8.3.6 向表中装载数据
8.3.7 查询表中数据
8.3.8 向表中插入数据或从表中导出数据

8.3.1 创建数据库、表、视图

- 创建数据库

① 创建数据库hive

```
hive> create database hive;
```

② 创建数据库hive，因为hive已经存在，所以会抛出异常，加上if not exists关键字，则不会抛出异常

```
hive> create database if not exists hive;
```

- 创建表

① 在hive数据库中，创建表usr，含三个属性id，name，age

```
hive> use hive;
hive>create table if not exists usr(id bigint,name string,age int);
```

② 在hive数据库中，创建表usr，含三个属性id，name，age，存储路径为“/usr/local/hive/warehouse/hive/usr”

```
hive>create table if not exists hive.usr(id bigint,name string,age int)
    >location '/usr/local/hive/warehouse/hive/usr';
```

8.3.1 创建数据库、表、视图

- 创建表

③ 在hive数据库中，创建外部表usr，含三个属性id，name，age，可以读取路径“/usr/local/data”下以“，”分隔的数据。

```
hive>create external table if not exists hive.usr(id bigint,name string,age int)
    >row format delimited fields terminated by ','
     location '/usr/local/data';
```

④ 在hive数据库中，创建分区表usr，含三个属性id，name，age，还存在分区字段sex。

```
hive>create table hive.usr(id bigint,name string,age int) partition by(sex boolean);
```

⑤ 在hive数据库中，创建分区表usr1，它通过复制表usr得到。

```
hive> use hive;
hive>create table if not exists usr1 like usr;
```

- 创建视图

①创建视图little_usr，只包含usr表中id，age属性

```
hive>create view little_usr as select id,age from usr;
```

8.3.2 删除数据库、表、视图

- 删除数据库

① 删除数据库hive，如果不存在会出现警告

```
hive> drop database hive;
```

② 删除数据库hive，因为有if exists关键字，即使不存在也不会抛出异常

```
hive>drop database if not exists hive;
```

③ 删除数据库hive，加上cascade关键字，可以删除当前数据库和该数据库中的表

```
hive> drop database if not exists hive cascade;
```

8.3.2 删除数据库、表、视图

- 删除表

① 删除表usr，如果是内部表，元数据和实际数据都会被删除；
如果是外部表，只删除元数据，不删除实际数据

```
hive> drop table if exists usr;
```

- 删除视图

① 删除视图little_usr

```
hive> drop view if exists little_usr;
```

8.3.3 修改数据库、表、视图

- 修改数据库

① 为hive数据库设置dbproperties键值对属性值来描述数据库属性信息

```
hive> alter database hive set dbproperties('edited-by'='lily');
```

- 修改表

① 重命名表usr为user

```
hive> alter table usr rename to user;
```

② 为表usr增加新分区

```
hive> alter table usr add if not exists partition(age=10);
```

③ 删除表usr中分区

```
hive> alter table usr drop if exists partition(age=10);
```

④ 把表usr中列名name修改为username，并把该列置于age列后

```
hive>alter table usr change name username string after age;
```

8.3.3 修改数据库、表、视图

- 修改表

⑤ 在对表usr分区字段之前，增加一个新列sex

hive>alter table usr add columns(sex boolean);

⑥ 删除表usr中所有字段并重新指定新字段newid，newname，newage

hive>alter table usr replace columns(newid bigint,newname string,newage int);

⑥ 为usr表设置tblproperties键值对属性值来描述表的属性信息

hive> alter table usr set tabproperties('notes'='the columns in usr may be null except id');

- 修改视图

① 修改little_usr视图元数据中的tblproperties属性信息

hive> alter view little_usr set tabproperties('create_at'='refer to timestamp');

8.3.4 查看数据库、表、视图

- 查看数据库

① 查看Hive中包含的所有数据库

```
hive> show databases;
```

② 查看Hive中以h开头的所有数据库

```
hive>show databases like 'h.*';
```

- 查看表和视图

① 查看数据库hive中所有表和视图

```
hive> use hive;
hive> show tables;
```

② 查看数据库hive中以u开头的所有表和视图

```
hive> show tables in hive like 'u.*';
```

8.3.5 描述数据库、表、视图

- 描述数据库

① 查看数据库hive的基本信息，包括数据库中文件位置信息等

```
hive> describe database hive;
```

② 查看数据库hive的详细信息，包括数据库的基本信息及属性信息等

```
hive>describe database extended hive;
```

- 描述表和视图

① 查看表usr和视图little_usr的基本信息，包括列信息等

```
hive> describe hive.usr/ hive.little_usr;
```

② 查看表usr和视图little_usr的详细信息，包括列信息、位置信息、属性信息等

```
hive> describe extended hive.usr/ hive.little_usr;
```

③ 查看表usr中列id的信息

```
hive> describe extended hive.usr.id;
```

8.3.6 向表中装载数据

① 把目录’/usr/local/data‘下的数据文件中的数据装载进usr表并覆盖原有数据

hive> load data local inpath ‘/usr/local/data’ overwrite into table usr;

② 把目录’/usr/local/data‘下的数据文件中的数据装载进usr表不覆盖原有数据

hive> load data local inpath ‘/usr/local/data’ into table usr;

③ 把分布式文件系统目录’hdfs://master_srever/usr/local/data‘下的数据文件数据装载进usr表并覆盖原有数据

hive> load data inpath ‘hdfs://master_srever/usr/local/data’

>overwrite into table usr;

8.3.7 查询表中数据

该命令和SQL语句完全相同这里不再赘述。

8.3.8 向表中插入数据或从表中导出数据

① 向表usr1中插入来自usr表的数据并覆盖原有数据

```
hive> insert overwrite table usr1
    > select * from usr where age=10;
```

② 向表usr1中插入来自usr表的数据并追加在原有数据后

```
hive> insert into table usr1
    > select * from usr
    > where age=10;
```

8.4 Hive应用实例：WordCount

现在我们通过一个实例——词频统计，来深入学习一下Hive的具体使用。首先，需要创建一个需要分析的输入数据文件，然后编写HiveQL语句实现WordCount算法，在Unix下实现步骤如下：

（1）创建input目录，其中input为输入目录。命令如下：

```
$ cd /usr/local/hadoop
$ mkdir input
```

（2）在input文件夹中创建两个测试文件file1.txt和file2.txt，命令如下：

```
$ cd  /usr/local/hadoop/input
$ echo "hello world" > file1.txt
$ echo "hello hadoop" > file2.txt
```

8.4 Hive应用实例：WordCount

（3）进入hive命令行界面，编写HiveQL语句实现WordCount算法，命令如下：

```
$ hive
hive> create table docs(line string);
hive> load data inpath 'input' overwrite into table docs;
hive>create table word_count as
    select word, count(1) as count from
    (select explode(split(line,' '))as word from docs) w
    group by word
    order by word;
```

执行完成后，用select语句查看运行结果如下：

```
OK
Time taken: 2.662 seconds
hive> select * from word_count;
OK
hadoop  1
hello   2
world   1
Time taken: 0.043 seconds, Fetched: 3 row(s)
```

8.5 Hive编程的优势

词频统计算法是最能体现MapReduce思想的算法之一，接下来，我们将比较WordCount算法在MapReduce中的编程实现和Hive中编程实现的主要不同点：

1. 采用Hive实现WordCount算法需要编写较少的代码量

- 在MapReduce中，wordcount类由63行Java代码编写而成代码位置：%HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.1.jar；
- 而在Hive中只需要编写7行代码

2. 在MapReduce的实现中，需要进行编译生成jar文件来执行算法，而在Hive中不需要。

- HiveQL语句的最终实现需要转换为MapReduce任务来执行，这都是由Hive框架自动完成的，用户不需要了解具体实现细节。

8.6 本章小结

Hive是一个构建于Hadoop顶层的数据仓库工具，主要用于对存储在Hadoop 文件中的数据集进行数据整理、特殊查询和分析处理。Hive在某种程度上可以看作是用户编程接口，本身不存储和处理数据，依赖HDFS存储数据，依赖MapReduce处理数据。

本章介绍了Hive的安装方法，包括下载安装文件、配置环境变量、修改配置文件、安装并配置MySQL等。Hive支持关系数据库中的大多数基本数据类型，同时Hive还支持关系数据库中不常出现的的3种集合数据类型。Hive提供了类似SQL的语句——HiveQL，可以很方便地对Hive进行操作，包括创建、修改、删除数据库、表、视图等。Hive的一大突出优点是，可以把查询语句自动转化成相应的MapReduce任务去执行得到结果，这样就可以大大节省用户的编程工作量，本章最后通过一个WordCount应用实例，充分展示了Hive的这一优点。

附录A：主讲教师林子雨简介

主讲教师：林子雨

单位：厦门大学计算机科学系
E-mail: ziyulin@xmu.edu.cn
个人网页：http://dblab.xmu.edu.cn/post/linziyu
数据库实验室网站：http://dblab.xmu.edu.cn

扫一扫访问个人主页

林子雨，男，1978年出生，博士（毕业于北京大学），全国高校知名大数据教师，现为厦门大学计算机科学系副教授，曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革局副局长。中国计算机学会数据库专业委员会委员，中国计算机学会信息系统专业委员会委员。国内高校首个“数字教师”提出者和建设者，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员，2013年度、2017年度和2020年度厦门大学教学类奖教金获得者，荣获2019年福建省精品在线开放课程、2018年厦门大学高等教育成果特等奖、2018年福建省高等教育教学成果二等奖、2018年国家精品在线开放课程。主要研究方向为数据库、数据仓库、数据挖掘、大数据、云计算和物联网，并以第一作者身份在《软件学报》《计算机学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文。作为项目负责人主持的科研项目包括1项国家自然科学青年基金项目(No.61303004)、1项福建省自然科学青年基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049)，主持的教改课题包括1项2016年福建省教改课题和1项2016年教育部产学协作育人项目，同时，作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。中国高校首个“数字教师”提出者和建设者，2009年至今，“数字教师”大平台累计向网络免费发布超过1000万字高价值的研究和教学资料，累计网络访问量超过1000万次。打造了中国高校大数据教学知名品牌，编著出版了中国高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》，并成为京东、当当网等网店畅销书籍；建设了国内高校首个大数据课程公共服务平台，为教师教学和学生学习大数据课程提供全方位、一站式服务，年访问量超过200万次，累计访问量超过1000万次。

附录B：大数据学习路线图

大数据学习路线图访问地址：http://dblab.xmu.edu.cn/post/10164/

附录C：林子雨大数据系列教材

了解全部教材信息：http://dblab.xmu.edu.cn/post/bigdatabook/

附录D：《大数据导论（通识课版）》教材

开设全校公共选修课的优质教材

本课程旨在实现以下几个培养目标：

- 引导学生步入大数据时代，积极投身大数据的变革浪潮之中
- 了解大数据概念，培养大数据思维，养成数据安全意识
- 认识大数据伦理，努力使自己的行为符合大数据伦理规范要求
- 熟悉大数据应用，探寻大数据与自己专业的应用结合点
- 激发学生基于大数据的创新创业热情

高等教育出版社 ISBN:978-7-04-053577-8 定价：32元

教材官网：http://dblab.xmu.edu.cn/post/bigdataintroduction/

附录E：《大数据导论》教材

- 林子雨 编著《大数据导论》
- 人民邮电出版社，2020年9月第1版
- ISBN:978-7-115-54446-9 定价：49.80元

教材官网：http://dblab.xmu.edu.cn/post/bigdata-introduction/

开设大数据专业导论课的优质教材

扫一扫访问教材官网

附录F：《大数据技术原理与应用》教材

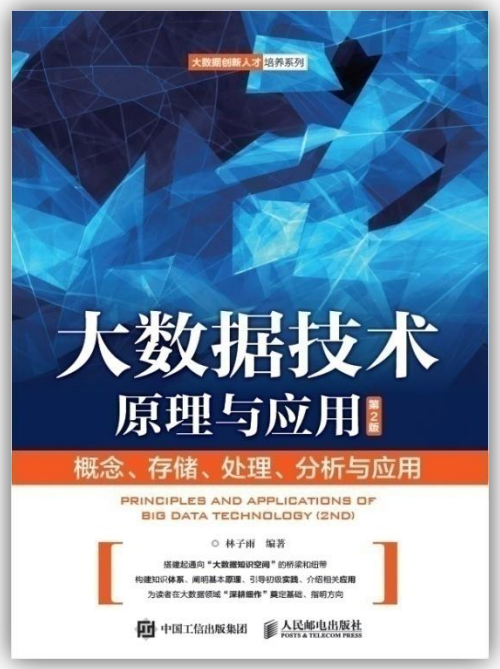
《大数据技术原理与应用——概念、存储、处理、分析与应用（第2版）》，由厦门大学计算机科学系林子雨博士编著，是国内高校第一本系统介绍大数据知识的专业教材。人民邮电出版社 ISBN:978-7-115-44330-4 定价：49.80元

全书共有15章，系统地论述了大数据的基本概念、大数据处理架构Hadoop、分布式文件系统HDFS、分布式数据 库HBase、NoSQL数据库、云数据库、分布式并行编程模型MapReduce、Spark、流计算、图计算、数据可视化以及大数据在互联网、生物医学和物流等各个领域的应用。在Hadoop、HDFS、HBase和MapReduce等重要章节，安排了入门级的实践操作，让读者更好地学习和掌握大数据关键技术。

本书可以作为高等院校计算机专业、信息管理等相关专业的大数据课程教材，也可供相关技术人员参考、学习、培训之用。

欢迎访问《大数据技术原理与应用——概念、存储、处理、分析与应用》教材官方网站：
http://dblab.xmu.edu.cn/post/bigdata

扫一扫访问教材官网



附录G：《大数据基础编程、实验和案例教程（第2版）》

本书是与《大数据技术原理与应用（第3版）》教材配套的唯一指定实验指导书

大数据教材

1+1黄金组合

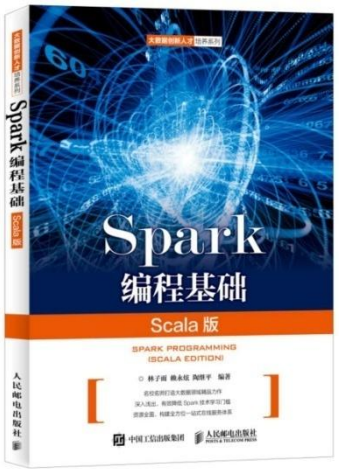
厦门大学林子雨编著

配套实验指导书

- 步步引导，循序渐进，详尽的安装指南为顺利搭建大数据实验环境铺平道路
- 深入浅出，去粗取精，丰富的代码实例帮助快速掌握大数据基础编程方法
- 精心设计，巧妙融合，八套大数据实验题目促进理论与编程知识的消化和吸收
- 结合理论，联系实际，大数据课程综合实验案例精彩呈现大数据分析全流程

林子雨编著《大数据基础编程、实验和案例教程（第2版）》
清华大学出版社 ISBN:978-7-302-55977-1 定价：69元 2020年10月第2版

附录H：《Spark编程基础（Scala版）》



《Spark编程基础（Scala版）》

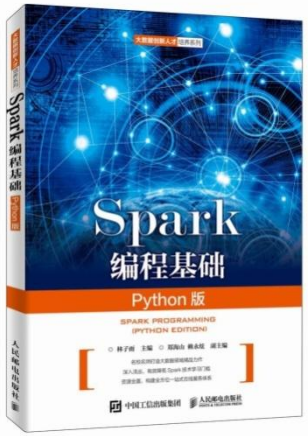
厦门大学 林子雨，赖永炫，陶继平 编著

披荆斩棘，在大数据丛林中开辟学习捷径
填沟削坎，为快速学习Spark技术铺平道路
深入浅出，有效降低Spark技术学习门槛
资源全面，构建全方位一站式在线服务体系

人民邮电出版社出版发行，ISBN:978-7-115-48816-9
教材官网：http://dblab.xmu.edu.cn/post/spark/

本书以Scala作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Scala语言基础、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。

附录I：《Spark编程基础（Python版）》



《Spark编程基础（Python版）》

厦门大学 林子雨，郑海山，赖永炫 编著

披荆斩棘，在大数据丛林中开辟学习捷径
填沟削坎，为快速学习Spark技术铺平道路
深入浅出，有效降低Spark技术学习门槛
资源全面，构建全方位一站式在线服务体系

人民邮电出版社出版发行，ISBN:978-7-115-52439-3
教材官网： http://dblab.xmu.edu.cn/post/spark-python/

本书以Python作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Structured Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、上机实验指南等。

附录J：高校大数据课程公共服务平台

高校大数据课程

公 共 服 务 平 台

http://dblab.xmu.edu.cn/post/bigdata-teaching-platform/

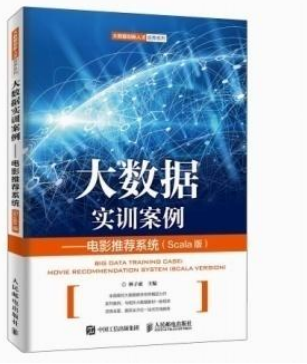
扫一扫访问平台主页

扫一扫观看3分钟FLASH动画宣传片

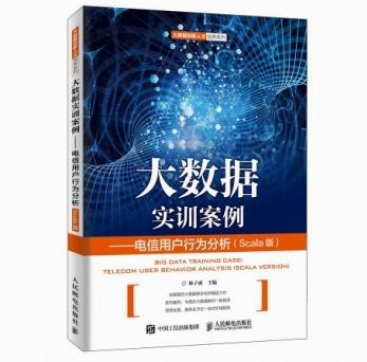
附录K：高校大数据实训课程系列案例教材

为了更好满足高校开设大数据实训课程的教材需求，厦门大学数据库实验室林子雨老师团队联合企业共同开发了《高校大数据实训课程系列案例》，目前已经完成开发的系列案例包括：

《电影推荐系统》（已经于2019年5月出版）
《电信用户行为分析》（已经于2019年5月出版）
《实时日志流处理分析》
《微博用户情感分析》
《互联网广告预测分析》
《网站日志处理分析》



系列案例教材将于2019年陆续出版发行，教材相关信息，敬请关注网页后续更新！
http://dblab.xmu.edu.cn/post/shixunkecheng/



扫一扫访问大数据实训课程系列案例教材主页


Thank You!