

大数据技术-第十章：购物网站用户行为分析综合案例 案例简介



CONTENTS

01. 案例目的

03. 时间安排

05. 硬件要求

07. 数据集

02. 适用对象

04. 预备知识

06. 软件工具



01

案例目的



- 熟悉Linux系统、MySQL、Hadoop、HBase、Hive与Sqoop等系统和软件的安装和使用；
- 了解大数据处理的基本流程；
- 熟悉数据预处理方法；
- 熟悉在不同类型数据库之间进行数据相互导入导出；
- 熟悉Sqoop数据迁移的方法。

02

适用对象



适用对象

- 高校（高职）教师
- 大数据与计算机相关专业学生
- 大数据自学者

03

时间安排



本案例可以作为大数据入门级课程结束后的“大作业”，或者可以作为学生暑期或寒假大数据实习实践基础案例，建议在一周左右完成本案例。

04

预备知识



- 需要案例使用者，已经学习过大数据技术的前面9章相关知识点，了解大数据相关技术的基本概念与原理，了解Windows操作系统、Linux操作系统、大数据处理架构Hadoop的关键技术及其基本原理、列族数据库HBase概念及其原理、数据仓库概念与原理、关系型数据库概念与原理等。
- 本案例提供了全部操作细节，包括每个命令和运行结果，所以，即使没有相关背景知识，也可以按照操作说明顺利完成全部实验。

05

硬件要求



>> 硬件要求

本案例可以在单机上完成，也可以在集群环境下完成。本案例采用集群上完成本案例实验时，建议计算机硬件配置为：50GB以上硬盘，8GB以上内存，开启3个虚拟机。

06

软件工具



相关软件版本建议如下：

➤ Linux: CentOS7

➤ MySQL: 5.7

➤ Hadoop: 2.6

➤ HBase: 1.2.0

➤ Hive: 2.0.0

➤ Sqoop: 1.4.7

➤ Zookeeper: 3.4.5

07

数据集



网站用户购物行为数据集，包括2000万条记录。包含字段信息如下：

user_id (用户id)

item_id(商品id)

behaviour_type (包括浏览、收藏、加购物车、购买，对应取值分别是1、2、3、4)

user_geohash(用户地理位置哈希值)

item_category (商品分类)

time (该记录产生时间)

Turing AI 万维图灵 | 大数据系列课程

大数据

BIG
DATA

智 / 能 / 科 / 技 放 / 眼 / 未 / 来

