《大数据技术基础》课程实验教学大纲

一、课程基本信息

课程代码：16176903

课程名称：大数据技术基础

英文名称: Fundamentals of Big Data Technology

实验总学时：16

适用专业：软件工程、计算机科学与技术、大数据管理

课程类别：专业课

先修课程：计算机网络，云计算基础，计算机体系结构，数据库原理，JAVA/Python 程序设计

二、实验教学的总体目的和要求

实验教学目的：

**对同学讲授经过实践检验的真理才是正确的，所有理论知识都应该经过实践检验，计算机科学这样，大数据技术基础这门前沿信息技术尤其如此。**本课程实验是为了使学生在课程学习的同时，通过实验教学验证课堂教学的理论，理解和掌握大数据技术基础中最基本、最广泛应用的概念、原理、理论和算法以及基本技术和方法，更好地掌握《大数据技术基础》课程教学大纲要求的内容。

实验要求：

1. 对学生的要求：实验前要充分做好准备工作：
   * 复习和掌握与本实验有关的知识内容；
   * 预习、思考实验内容；
   * 对实验内容进行分析和设计。

实验过程中，实验者必须服从指导教师和实验室工作人员的安排，遵守纪律与实验制度，爱护设备及卫生。在指定的实验时间内，必须到实验室内做实验。

对于上机过程中出现的问题，尽量先独立思考和解决；对于难以解决的问题可以和同学交流或询问老师；对于同一个实验题目，可以考虑多种方法来实现，

然后比较并选择出一种较为有效的方法来实现。

对于设计型和验证型实验，实验时一人一组，独立上机。 2、对实验条件的要求：普通 PC 机房。

三、实验教学内容

实验项目一实验名称：熟悉常用的Linux操作和Hadoop操作

实验内容：在 Windows（Mac）操作系统中安装Ubuntu虚拟机或安装Win/Linux双系统。进行简单Linux系统操作。搭建Hadoop单机版，伪分布式集群和完全分布式集群，并完成简单hadoop命令操作，运行系统例程。

实验性质：验证型实验学时：2

实验目的与要求：必修实验条件：实验机房

实验项目二实验名称：熟悉常用的HDFS操作

实验内容：通过Hadoop SHELL和JAVA 编程同时实现对HDFS上文件的上传，修改，查找等功能操作。

实验性质：设计型

实验学时：2

实验目的与要求：必修实验条件：实验机房

实验项目三实验名称：熟悉常用的HBase操作

实验内容：完成HBase集群搭建。熟练使用Shell命令和JAVA API进行Hadoop上分布式数据库HBase应用开发，实现建表，增改列族，列，增改数据项等分布式数据库操作应用。并完成相应分布式数据库应用开发。

实验性质：设计型

实验学时：4

实验目的与要求：必修实验条件：实验机房

实验项目四实验名称：Nosql与云数据库操作

实验内容：熟练掌握数据库常用shell命令和JAVA API开发。试用华为分布式数据库Gauss集群，试用阿里云数据库PolarDB。通过JAVA程序进行数据库上云和云数据库应用开发。

实验性质：设计型

实验学时：4

实验目的与要求：必修

实验条件：华为Gauss数据库集群，阿里云数据库PolarDB数据分析平台。实验机房。**通过华为Gauss数据库平台和阿里云数据库PolarDB平台介绍与使用，使同学们掌握多个业内最先进的大数据处理，计算及分析挖掘开发平台，进行高质量的理论实践，提高动手能力，同时也可了解国内顶尖大数据分析技术发展现状，增强民族自豪感，激发学习热情。**

实验项目五实验名称： MapReduce编程实战

实验内容：通过实验掌握基本的MapReduce编程方法；掌握用MapReduce解决一些常见的数据处理问题，包括数据去重、数据排序和数据挖掘等。试用阿里大数据服务MaxCompute进行简单大数据处理与分析任务开发。

实验性质：设计型

实验学时：4

实验目的与要求：必修

实验条件：阿里大数据服务MaxCompute。实验机房。**通过阿里大数据服务MaxCompute平台介绍与使用，使同学们掌握业内最先进的大数据处理，计算及分析开发平台，进行高质量的理论实践，提高动手能力和大数据技术应用能力，分析解决问题能力，同时也可了解国内顶尖大数据分析技术发展现状，增强民族自豪感，激发学习热情。**

四、考核方式

1. 实验课的考核方式：每个实验单独计分。验证型和设计型实验依据源程序和运行结果评分，综合性实验要结合课程设计报告综合评分。
2. 实验课考核成绩的确定：实验课成绩计入课程总成绩，验证型和设计型实验成绩计为课程平时成绩。

**五、推荐实验教材和教学参考书**

推荐教材：

1. 《大数据技术原理与应用——大数据概念、存储、处理、分析与应用》厦门大学 林子雨编著，人民邮电出版社，2017年2月第2版，ISBN:978-7-115-44330-4

参考书目：

1. 大数据软件安装和编程实践指南

官网http://dblab.xmu.edu.cn/post/5663/

1. 实验指南

官网：http://dblab.xmu.edu.cn/post/6131/

1. 电子书籍

官网：http://dblab.xmu.edu.cn/post/4782/

1. Spark入门教程

官网：http://dblab.xmu.edu.cn/blog/spark/

1. 大数据课程实验案例《网站用户购物行为分析》

官网：http://dblab.xmu.edu.cn/post/7499/

1. **华为GaussDB分布式数据库**

[**https://e.huawei.com/cn/products/cloud-computing-dc/gaussdb**](https://e.huawei.com/cn/products/cloud-computing-dc/gaussdb)

1. **阿里云数据库PolarDB**

**https://www.aliyun.com/tg/apsaradbsol/6339432.html**

1. **华为云FusionInsight大数据**

[**https://support.huawei.com/enterprise/zh/cloud-computing/fusioninsight-pid-19919550**](https://support.huawei.com/enterprise/zh/cloud-computing/fusioninsight-pid-19919550)

1. **阿里云大数据服务Maxcompute**

[**https://www.aliyun.com/product/odps**](https://www.aliyun.com/product/odps)

1. **其他需说明的**

无。

大纲修订人：胡苏 修订日期: 2022年1月

大纲审定者：郭玉柱 审定日期: 2022年1月

**注：在第二、三、五部分引入课程思政建设内容，其中第三部分的教学内容部分，引入思政教学知识点三个以上，均已标红加粗显示。**