

传智播客

《Spark 大数据分析 with 实战》

教学设计

课程名称： Spark 大数据分析 with 实战

授课年级： 20xx 年级

授课学期： 20xx 学年第一学期

教师姓名： 某某老师

2019 年 9 月 1 日

课题名称	第 2 章 Spark 基础	计划学时	7 课时
内容分析	Spark 于 2009 年诞生于美国加州大学伯克利分校的 AMP 实验室，它是一个可应用于大规模数据处理的统一分析引擎。Spark 不仅计算速度快，而且内置了丰富的 API，使得我们能够更加容易编写程序。本章将从 Spark 的发展说起，针对 Spark 集群部署、Spark 运行架构及其原理进行详细讲解。		
教学目标及基本要求	<ol style="list-style-type: none"> 1、掌握 Spark 集群的搭建和配置 2、掌握 Spark HA 集群的搭建和配置 3、掌握 Spark 集群架构 4、理解 Spark 作业提交的工作原理 		
教学重点	<ol style="list-style-type: none"> 1、掌握 Spark 集群的搭建和配置 2、掌握 Spark 集群架构 		
教学难点	Spark 作业提交的工作原理		
教学方式	教师课堂教学要以讲演法讲授为主，并结合多媒体进行教学		
教学过程	<p style="text-align: center;">第一课时</p> <p style="text-align: center;">(Spark 概述、Spark 的特点、Spark 应用场景、Spark 与 Hadoop 对比)</p> <p>一、回顾第一章内容，讲解初识 Spark</p> <ol style="list-style-type: none"> 1. 回顾第一章学习知识，引出本节主题 带领学生回顾第一章与 Scala 相关的知识，要想深入的学习掌握 Scala 语言，首先得学习 Spark 技术。因此，本章将针对 Spark 的基础进行详细讲解。 2. 明确学习目标 <ol style="list-style-type: none"> (1) 了解 Spark 基本概念 (2) 熟悉 Spark 的特点 (3) 了解 Spark 应用场景 (4) 掌握 Spark 与 Hadoop 的区别 <p>二、进行重点知识讲解</p> <ol style="list-style-type: none"> 1. Spark 概述 教师可以参考课件讲解 Spark 及其生态系统。先介绍 Spark 的由来，Spark 于 2009 年诞生于美国加州大学伯克利分校的 AMP 实验室，它是一个可应用于大规模数据处理的统一分析引擎，并于 2014 年正式成为 Apache 软件基金会的顶级项目。再介绍 Spark 的生态系统中包含的组件有 Spark Core、Spark SQL、Spark Streaming、MLlib、GraphX 以及独立调度器，并依次介绍这些组件，最后接收 Spark 具有的显著优势。 2. Spark 的特点 教师可以参考课件讲解 Spark 的特点。Spark 有五大特点，分别是速度快、易用性、通用性以及兼容性。 		

3. Spark 应用场景

教师可以参考课件讲述 Spark 应用场景。Spark 在数据科学应用中，可以进行数据分析与建模；Spark 在数据处理应用中，可以对应用进行分析、效果分析以及定向优化。

4. Spark 与 Hadoop 对比

教师可以参考课件分别从编程方式、数据存储、数据处理以及数据容错这四个方面讲述 Spark 与 Hadoop 的区别对比。

三、归纳总结，随堂练习，布置作业

对课堂上讲解的知识点进行总结，使用高校教辅平台中的随堂练习题巩固本节课的知识点。

第二课时

(环境准备、Spark 的部署方式、Spark 集群安装部署、Spark HA 集群部署)

一、回顾上节课内容，讲述搭建 Spark 开发环境

1. 对上节课留的作业进行答疑
2. 回顾上节课内容，引出本节课的主题

通过上节课的学习，学生对 Spark 有了初步的认识。本节课将针对搭建 Spark 开发环境进行详细讲解。

3. 明确学习目标

- (1) 了解 Spark 开发环境准备
- (2) 熟悉 Spark 的部署方式
- (3) 掌握 Spark 集群安装部署
- (4) 掌握 Spark HA 集群部署

二、进行重点知识的讲解

1. 环境准备

教师可以参考课件以讲演法的方式讲述 Spark 开发环境的准备工作。由于 Spark 仅仅是一种计算框架，不负责数据的存储和管理，因此，通常都会将 Spark 和 Hadoop 进行统一部署，由 Hadoop 中的 HDFS、HBase 等组件负责数据的存储管理，Spark 负责数据计算。安装 Spark 集群之前，需要安装 JDK 以及 Hadoop 环境。

2. Spark 的部署方式

教师可以参考课件以讲演法的方式讲述 Spark 的部署方式。Spark 部署模式分为 Local 模式（本地单机模式）和集群模式，在 Local 模式下，常用于本地开发程序与测试，而集群模式又分为 Standalone 模式（集群单机模式）、Yarn 模式和 Mesos 模式

3. Spark 集群安装部署

教师可以参考课件以讲演法的方式讲述 Spark 集群安装部署。

- (1) 先介绍 Spark 安装包的下载路径；
- (2) 再介绍配置 Spark 安装包的解压安装；

- (3) 介绍修改相关配置文件，部署 Spark 集群；
- (4) 分发文件至其他两台服务器上，完成 Spark 集群的部署；
- (5) 通过执行“sbin/start-all.sh”命令，启动 Spark 集群。

4. Spark HA 集群部署

教师可以参考课件以讲演法的方式讲述 Spark HA 集群部署。

- (1) 先介绍部署 Spark HA 集群的意义；
- (2) 再介绍修改相关配置文件，将 Spark 集群部署成高可用的集群；
- (3) 启动 Spark HA 集群。

三、归纳总结，随堂练习，布置作业

1. 对课堂上讲解的知识点进行总结，使用高校教辅平台中的随堂练习题巩固本节课的知识点。
2. 让学生自己动手搭建 Spark 开发环境，以此来巩固本节的学习内容。

第三课时

(基本概念、Spark 集群运行架构、Spark 运行基本流程)

一、回顾上节课内容，

1. 对上节课留的作业进行答疑
2. 回顾上节课内容，引出本节课的主题

上一节中，主要讲解了 Spark 开发环境的搭建。本节将针对 Spark 运行架构与原理进行详细讲解。

3. 明确学习目标

- (1) 熟悉 Spark 运行架构中的基本概念
- (2) 掌握 Spark 集群运行架构
- (3) 理解 Spark 运行基本流程

二、进行重点知识的讲解

1. 基本概念

教师可以参考课件来讲述 Spark 运行架构中的基本概念，即 Application、Driver Program、Cluster Manager、Worker Node、Executor、Task 以及 Job 和 Stage。

2. Spark 集群运行架构

教师可以参考课件来讲述 Spark 集群运行架构，通过一张图演示 Spark 集群的运行架构，然后详细介绍 Spark 应用程序中的执行器进程和集群管理器。

3. Spark 运行基本流程

教师可以参考课件来讲述 Spark 运行基本流程，通过一张图演示 Spark 运行基本流程，然后分步骤详细讲解 Spark 运行基本流程。

三、归纳总结，随堂练习，布置作业

对课堂上讲解的知识点进行总结，使用高校教辅平台中的随堂练习题巩固本节课的知识点。

第四课时

(体验第一个 Spark 程序)

一、回顾上节课内容，讲解体验第一个 Spark 程序

1. 对上节课留的作业进行答疑。
2. 回顾上节课的内容，引出本节课的主题。

上一节介绍了 Spark 的运行架构与原理，Spark 集群已经部署完毕。下面通过使用 Spark 官方示例 SparkPi，体验 Spark 集群提交任务的流程。

3. 明确学习目标

通过体验第一个 Spark 程序，熟悉 Spark 集群提交任务的流程

二、进行重点知识的讲解

提交 Spark 作业，体验第一个 Spark 程序

教师可以参考课件以讲演法的方式来讲述提交 Spark 作业，体验第一个 Spark 程序。首先介绍执行相关命令提交 Spark 官方提供的示例 Spark Pi，再通过访问 Spark 集群管理界面，观察 Completed Applications 表单 Stats 的值，值为 Finished 时，再去查看控制台的输出信息，发现 Pi 值已经被计算完毕，即出现 Pi is roughly 3.140691140691141。

三、归纳总结，随堂练习，布置作业

1. 对课堂上讲解的知识点进行总结，使用高校教辅平台中的随堂练习题巩固本节课的知识点。
2. 让学生自己动手操作来体验第一个 Spark 程序，以此来巩固本节的学习内容。

第五课时

(运行 Spark-Shell 命令、运行 Spark Shell 读取 HDFS)

一、回顾上节课内容，运行 Spark-Shell 进行编程

1. 对上节课留的作业进行答疑
2. 回顾上节课内容，引出本节课的主题

上一节中，我们体验了第一个是 Spark 程序，深入理解了 Spark 集群提交任务的流程。Spark-Shell 是一个强大的交互式数据分析工具，初学者可以很好的使用它来学习相关 API，用户可以在命令行下使用 Scala 编写 Spark 程序。本节将针对 Spark-Shell 的相关知识进行详细讲解。

3. 明确学习目标

- (1) 熟悉运行 Spark-Shell 命令
- (2) 掌握 Spark Shell 读取 HDFS 文件

二、进行重点知识的讲解

1. 运行 Spark-Shell 命令

教师可以参考课件来讲述 Spark-Shell 的命令，通过执行“bin/spark-shell --master <master-url>”命令进入 Spark-Shell 交互式环境，通过执行“--help”命令获取 Spark-Shell

帮助选项列表。

2. 运行 Spark Shell 读取 HDFS 文件

教师可以参考课件以讲演法的方式讲述运行 Spark Shell 读取 HDFS 文件。

- (1) 准备数据文件 words.txt;
- (2) 整合 Spark 与 HDFS;
- (3) 启动 Hadoop 和 Spark 服务，运行 Spark-Shell 交互式界面;
- (4) 编写程序读取 HDFS 上的数据文件，并实现的单词计数的功能;
- (5) 退出 Spark-Shell 客户端。

三、归纳总结，随堂练习，布置作业

1. 对课堂上讲解的知识点进行总结，使用高校教辅平台中的随堂练习题巩固本节课的知识点。

2. 让学生自己动手进入 Spark-Shell 交互式界面，并进行编写程序，实现单词计数的功能，以此来巩固本节的学习内容。

第六课时

(本地模式执行 Spark 程序、集群模式执行 Spark 程序)

一、回顾上节课内容，通过 IDEA 开发 WordCount 程序

1. 对上节课留的作业进行答疑
2. 回顾上节课内容，引出本节课的主题

上一节中，我们通过 Spark-Shell 交互式界面进行编程，实现单词计数功能。Spark-Shell 通常在测试和验证我们的程序时使用的较多，然而在生产环境中，通常会在 IDEA 开发工具中编写程序，然后打成 Jar 包，最后提交到集群中执行。本节将针对 IDEA 开发 WordCount 程序进行详细讲解。

3. 明确学习目标

- (1) 熟悉本地模式执行 Spark 程序
- (2) 熟悉集群模式执行 Spark 程序

二、进行重点知识的讲解

1. 本地模式执行 Spark 程序

教师可以参考课件以讲演法的方式讲述本地模式执行 Spark 程序。

- (1) 创建 Maven 项目，新建资源文件夹;
- (2) 添加 Spark 相关依赖、打包插件;
- (3) 编写代码，执行代码，查看控制台输出的结果。

2. 集群模式执行 Spark 程序

教师可以参考课件以讲演法的方式讲述集群模式执行 Spark 程序。

- (1) 往 Maven 项目的 pom.xml 配置文件中，添加打包插件;
- (2) 修改词频统计的代码，打包程序;
- (3) 启动 Hadoop 和 Spark 服务，运行 Spark-Shell 交互式界面;
- (4) 通过执行“spark-submit”提交命令，提交 Spark 任务;
- (5) 访问 HDFS WEB 页面，查看/spark/test/out 文件夹，并将结果文件 part-*下

	<p>载到本地进行查看结果。</p> <p>三、归纳总结，随堂练习，布置作业</p> <ol style="list-style-type: none">1. 对课堂上讲解的知识点进行总结，使用高校教辅平台中的随堂练习题巩固本节课的知识点。2. 让学生自己动手使用 IDEA 开发 WordCount 程序，以此来巩固本节的学习内容。 <p style="text-align: center;">第七课时 (上机练习)</p> <p>上机一：搭建 Spark 开发环境 请按照教材中 2.2 小节的内容，独立完成。</p> <p>上机二：使用 IDEA 开发 WordCount 程序 请按照教材中 2.6 小节的内容，独立完成。</p>
思考题和习题	
教 学 后 记	