软工实验课堂记录

# 2017/3/17

|  |  |  |  |
| --- | --- | --- | --- |
| **软工实验第3周课堂记录** | | | |
| 时间 | 2017/3/17 | 地点 | 北航新主楼A209 |
| 成员 | 全体成员 | 时间 | 17:00-17:30 |
| 记录 | 王春柳 | 内容 | 本组发言内容 |
| 发言人：组长  我们的项目选择的是Spark，今天介绍的主要内容是spark的研究与应用;  1.简介：Apache Spark是用于大规模数据处理的快速的通用引擎，是一种基于内存的开源计算框架。它能快速处理多种场景下的大数据问题，高效挖掘大数据中的价值，为业务发展提供决策支持。  2.spark的发展史：2009年诞生于加州大学伯克利分校AMPLab，2010年项目开源，Spark系统思想的论文发表 ；2013年成为Apache孵化项目；2014年成为Apache的顶级项目；  2015年Spark1.0.0版正式上线。  3.为什么选择我们的理由是流行，Spark被使用案例很多，许多大公司用spark。  4.功能模块分为四块，包括Scheduler模块、Deploy模块、Executor模块、Storage模块  5.项目目标我们也分为四块，包括需求、测试、源码以及应用。  6.工作计划大体上包括项目计划书、线上沟通、线下讨论、项目进步、文档管理更新、分工协作六大方面。此处不一一介绍了。  **提问环节：**  同学提问：  1.代码量大，四个模块，怎么着手？回答：对其不了解，四个人擅长的都不一样  2.建议字体有点小，白色字体看不清。回答：我们下次改正  3.spark的优势在哪？回答：spark不需要编写，内存处理数据，效率高  老师提问：  1.任老师建议：你觉得所选的项目以后会怎么用到？  Spark会浪费大量的时间用在配置环境上，需要花费大量的时间来学习新东西，建议规定到某一模块或者换一个项目。如果是tensorflow就专门挑一个问题。  2.刘老师建议：课堂重点在于八个实验，，太难得不到高分，太简单学不到东西，如果选spark需要与前人不同，包括node.js。 | | | |