

# Hadoop 2.x 深入浅出基础课程

讲师-梦琪

## 课程大纲

---

- ◆ What is Apache Hadoop、生态系统介绍
- ◆ Hadoop 起源、版本变迁、目录结构
- ◆ Hadoop 2.x 编译、导入源码至Eclipse
- ◆ HDFS 、YARN、MapReduce on YARN 架构原理
- ◆ Hadoop 2.x 伪分布式环境安装、启动脚本和日志分析
- ◆ Hadoop HistoryServer 启动与配置
- ◆ Hadoop 默认配置文件与自定义配置文件讲解
- ◆ 阅读官方环境安装文档、MapReduce Uber 模式讲解

## 课程大纲

---

- ◆ 分布式安装 Hadoop 2.x
- ◆ HDFS Federation 原理与配置、ViewFS
- ◆ HDFS HA 发展、方案、原理、配置（手动）
- ◆ Zookeeper 架构、分布式安装、应用
- ◆ HDFS HA + Federation 配置（自动）
- ◆ HDFS 添加/卸载 节点，从1.x 升级到2.x
- ◆ Hadoop 2.x 其他特性使用、Shell 命令
- ◆ Hadoop 2.x 企业实战运维技巧、JAVA API、网盘

## 课程大纲

---

- ◆ YARN 架构原理、生态系统、MapReduce 运行原理
- ◆ MapReduce 执行过程、Eclipse 插件编译使用
- ◆ MapReduce 编程之自定义数据类型
- ◆ MapReduce 编程之InputFormat、Outputmat
- ◆ MapReduce 编程之Combiner、Partitioner、二次排序
- ◆ MapReduce 编程之分布式缓存、压缩、多Mapper
- ◆ YARN HA 原理、配置和添加/卸载 节点
- ◆ YARN 上部署Spark 1.0.0、运行WordCount