

# 分布式编程模型与系统



徐辰  
cxu@dase.ecnu.edu.cn

华东师范大学



## 课程名称

### 2 研究生

- 2018、2019年：大数据处理系统
- 2020年：大规模数据处理系统
- 2021年：分布式计算系统



### 本科生

- 2018、2019年：分布式模型与编程
- 2021、2022年：分布式编程模型与系统
- 2023年：分布式计算系统

## 课程背景

### 3 大数据处理系统 → 分布式计算系统

- Hadoop、Spark、Flink等
- “大数据”的涵义过于宽泛
- 其它类似课程/教材
  - 英文论文的翻译：不同论文的体系可能不一致
  - 针对某一系统的工具手册：时效性
- 本课程/教材
  - 强调系统设计、原理、编程的结合

## 课程目的

### 4 培养“系统思维”



## 课程内容

- When: 背景
- Why: 设计
- What: 架构
- How:
  - 原理：系统层面
  - 编程：用户层面

## 课程安排

- 理论课程：每周3学时，多记笔记
  - 设计思想：为什么？
  - 系统架构：是什么？不同系统的联系与区别
  - 编程思路：不是教API
- 实践课程：每周2学时
  - 开源系统部署：保持耐心、坑很多，不要奢望照着实验说明就能一步到位
  - 基本编程开发、代码调试：
    - 使用Java开发，动手能力强的自学Scala
    - 熟练使用IntelliJ IDE、maven

## 课程安排(tentative)

7

周	周一	周四	周	周一	周四
1	绪论	准备工作	10	五一放假	作业点评
2	HDFS	Hadoop 1部署	11	Flink	Flink部署
3	MapReduce	Hadoop 2部署	12	Flink	Flink编程
4	MapReduce	Hadoop 2编程	13	批流融合	Flink+Yarn
5	Spark	编程作业说明	14	批流融合	作业点评
6	清明放假	Spark部署	15	新技术介绍	课程项目
7	Spark	Spark编程	16	端午放假	课程项目
8	Spark	作业点评	17	总结	课程项目验收
9	Yarn	Spark+Yarn	18	考试	

## 课后训练

8

### 理论复习

- 提供—些参考资料、论文（但不作要求）
- 多看笔记、多思考，不要死记硬背

### 动手编程是最关键的

- 上机作业：在线提交
- 自己多看开源系统自带的例子
- 增强调试代码的能力
  - 设断点调试
  - 遇到错误，多用Google搜索

## 课程成绩评定

9

### 平时成绩：50%

- 考勤：10%
- 作业：40%
- 实验报告：30%
- 课程项目：20%

### 期末成绩：50%

- 期末考试：100%

## 实验报告

10

### 实验目的：学习XXX

### 实验任务：给定

### 使用环境：Ubuntu 18.04

### 实验过程：主体部分

- 不是抄实验说明，主要记录你踩过什么坑
- 截图（图中需包含个人账户名称）、简要的文字说明、重要的代码片段，

### 思考题

### 总结：个人的体会和感想

## 实验报告提交

11

### 实验报告文件命名格式

- 组号-学号-姓名-实验X-版本Y.pdf
  - X取值范围[1-5]
  - Y取值范围[1, +∞)
- 例如：1-10061-张三-实验1-版本2.pdf

- 需通过大夏学堂网上提交，规定的deadline之前，逾期提交的实验报告影响评分

## 教材及参考书目

12

### 教材

- 徐辰：分布式计算系统—面向数据密集型应用
- 徐辰：分布式计算系统实践教程

### 参考书

- 分布式系统概念与设计，George Coulouris等著，金蓓弘等译
- 设计数据密集型应用，Martin Kleppmann



## 教学信息

13

### □ 教学团队

✚ 主讲教师：徐辰

✚ 助教：

➢ 杨溢：2019级硕士

➢ 毕倪飞：2019级硕士

### □ 课程信息

✚ 主页：大夏学堂<https://elearning.ecnu.edu.cn/>

✚ 邮箱：dasebigdata@163.com

13

DAS

## 谢谢！Q&A



14

14

DAS