



浙江大学  
ZheJiang University

# 大规模软件系统构建技术导论

## 课程说明

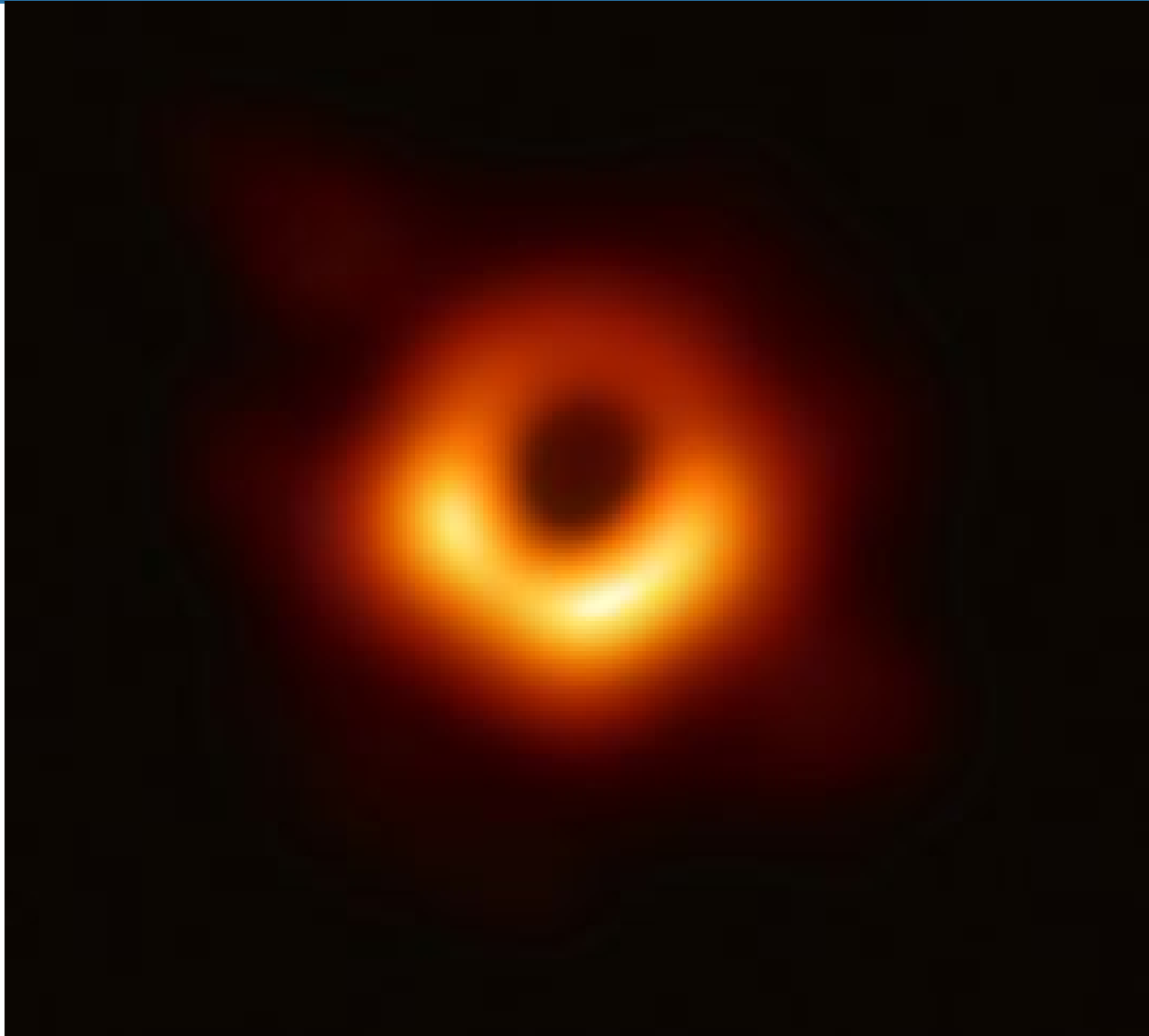
浙江大学计算机学院



# 课程说明

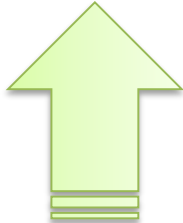
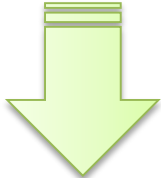
- 课程名称：大规模软件系统构建技术导论
- 主讲老师：鲁伟明
- 联系方式
  - 办公室：玉泉校区数字图书馆教育部工程研究中心（玉泉校区图书馆220）
  - 邮箱：[luwm@zju.edu.cn](mailto:luwm@zju.edu.cn)
  - 电话：13575464647
  - 办公室电话：87953779

# 人类首张黑洞照片



2019年4月10日21点整，天文学家召开全球新闻发布会，宣布首次直接拍摄到黑洞的照片。

## How Small Can We See?

$$\text{Smallest Size} \approx \frac{\text{Wavelength}}{\text{Telescope Size}}$$


How Big Must Our Telescope Be?

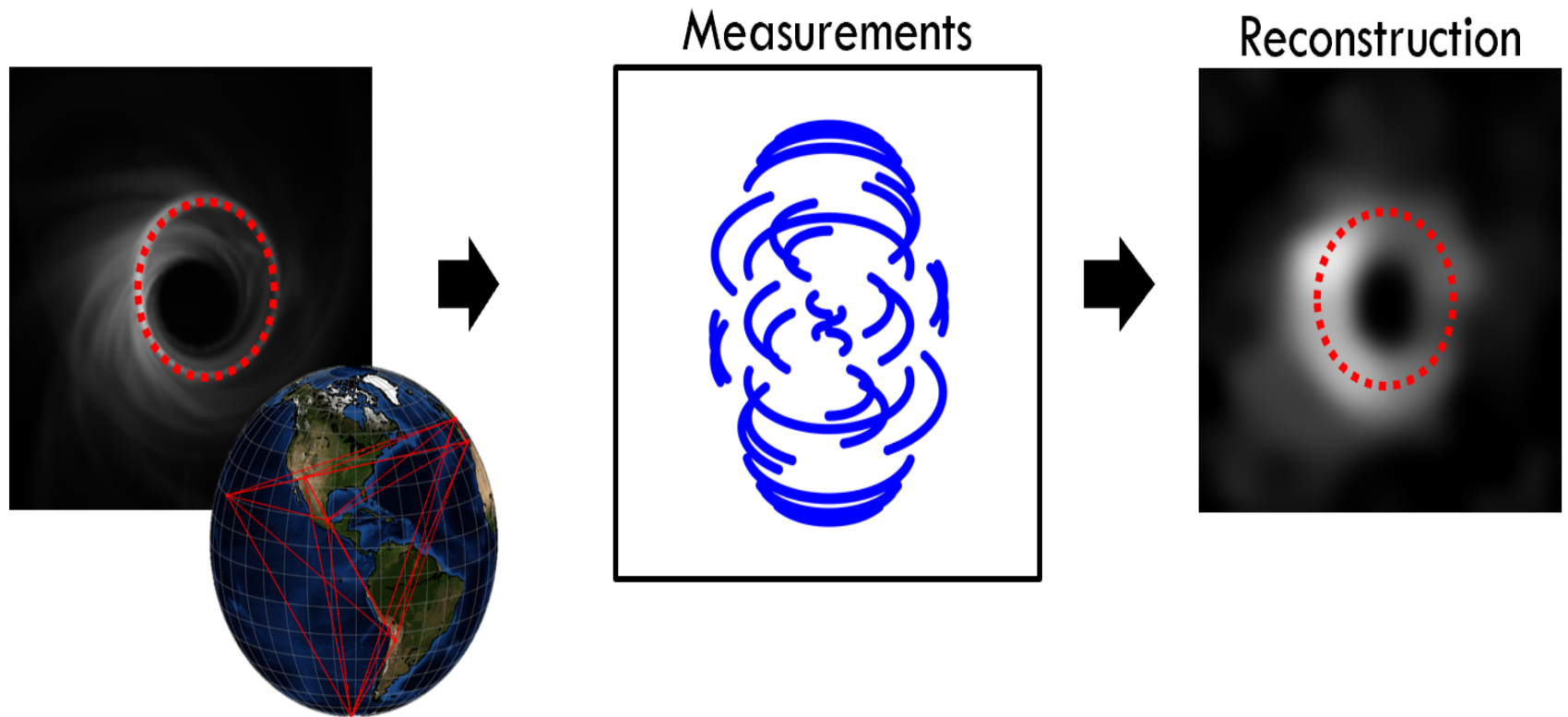

$$\approx \frac{\text{Wavelength}}{\text{Telescope Size}}$$





# 黑洞照片

## Black Hole in the Center of the Milky Way Galaxy



# 黑洞照片

2017年4月份的观测中，每个台站的数据率达到惊人的 $32\text{Gbit} / \text{s}$ ，8个台站在5天观测期间共记录约 $3500\text{TB}$ 数据（相当于350万部电影，至少要几百年才能看完！）。EHT采用专用硬盘来记录数据，再把它们送回数据中心进行处理。





# 课程说明

## □ 教学目标

- 从基础理论、开源软件、应用案例的角度，教授研发大规模软件系统所需的基础构建技术。
- 提高对开源软件框架的分析能力和基于开源软件框架构建大规模软件系统的能力。





# 课程说明

## □ 教学内容

### ■ 高性能计算技术

□ 并行计算、分布式计算、密集计算（GPU计算）

### ■ 大数据管理技术

□ 大规模文件系统、大规模数据库系统

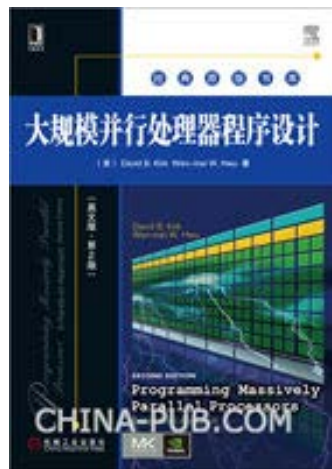
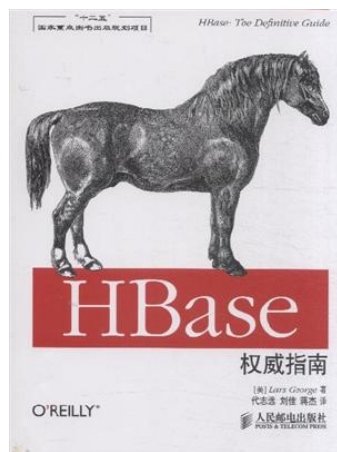
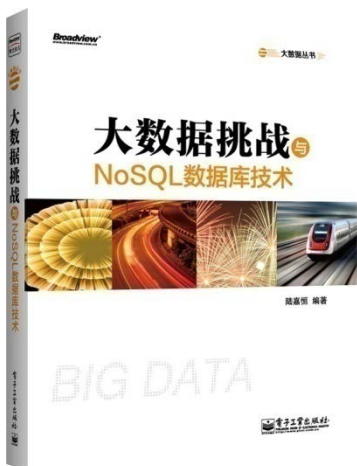
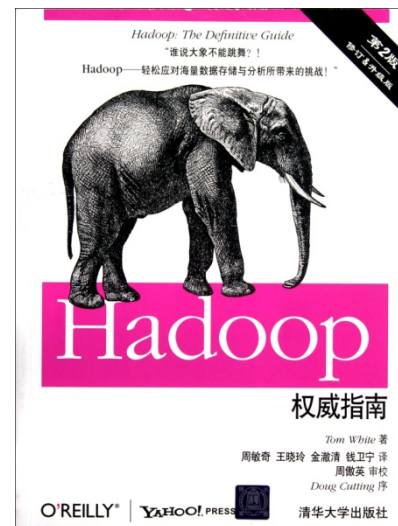
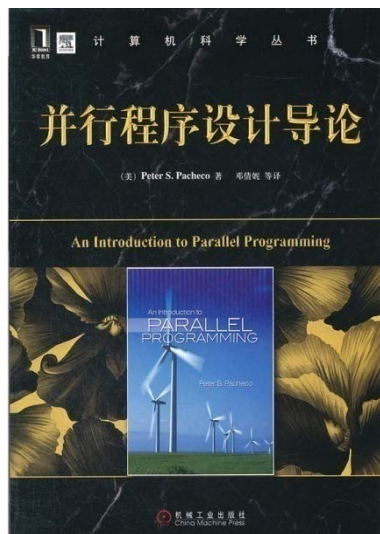
### ■ 大型Web应用构建技术

□ 负载均衡、缓存机制、容错机制

### ■ 云计算技术、虚拟化技术

# 课程说明

## 参考图书





# 课程说明

## □ 联系方式

- QQ群：678048864
- 课间加群
- 请改实名





# 课程说明

## ☐ 课程考核

### ■ 平时出勤

### ■ 开源框架项目分析与实践（形成报告、PPT和视频）

☐ 设计理念

☐ 代码分析

☐ 应用实例

### ■ 小程序？

## ☐ 评分形式

☐ 考勤

☐ 报告

☐ PPT

☐ 视频（10分钟）

☐ 小程序？



# 开源框架项目分析与实践

## □设计理念

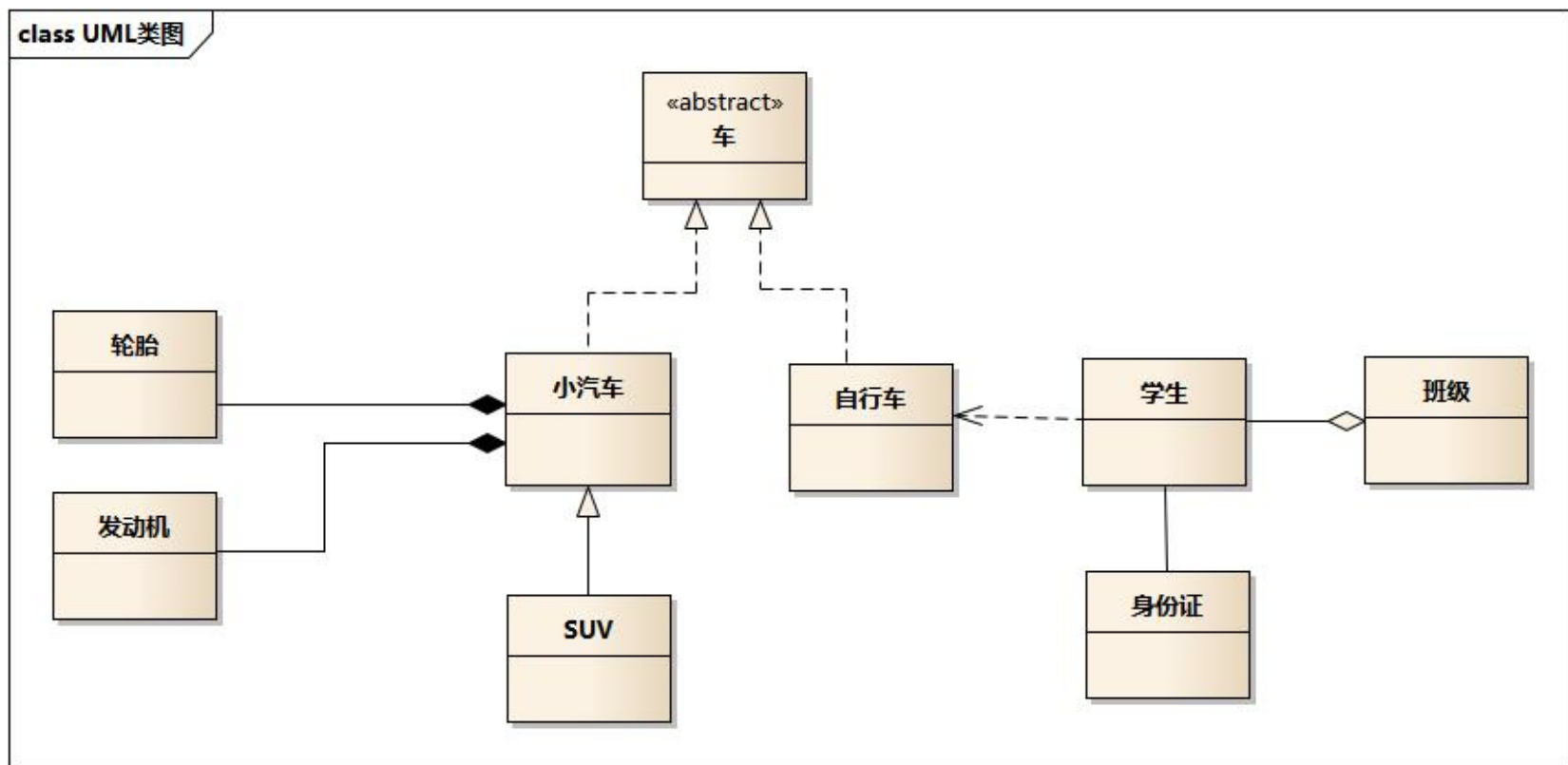
- 分析一个开源大规模软件框架项目（框架、模块、功能）
- 例如
  - YARN或MapReduce等中资源调度策略
  - MapReduce中的任务调度策略（MapReduce的实现）
  - 图计算框架中BSP计算模式的实现
  - Hadoop中HDFS中读写数据的实现
  - .....

## □代码分析

- 从源代码来分析上述功能模块是如何实现的
  - 数据结构
  - 类图，时序图等等

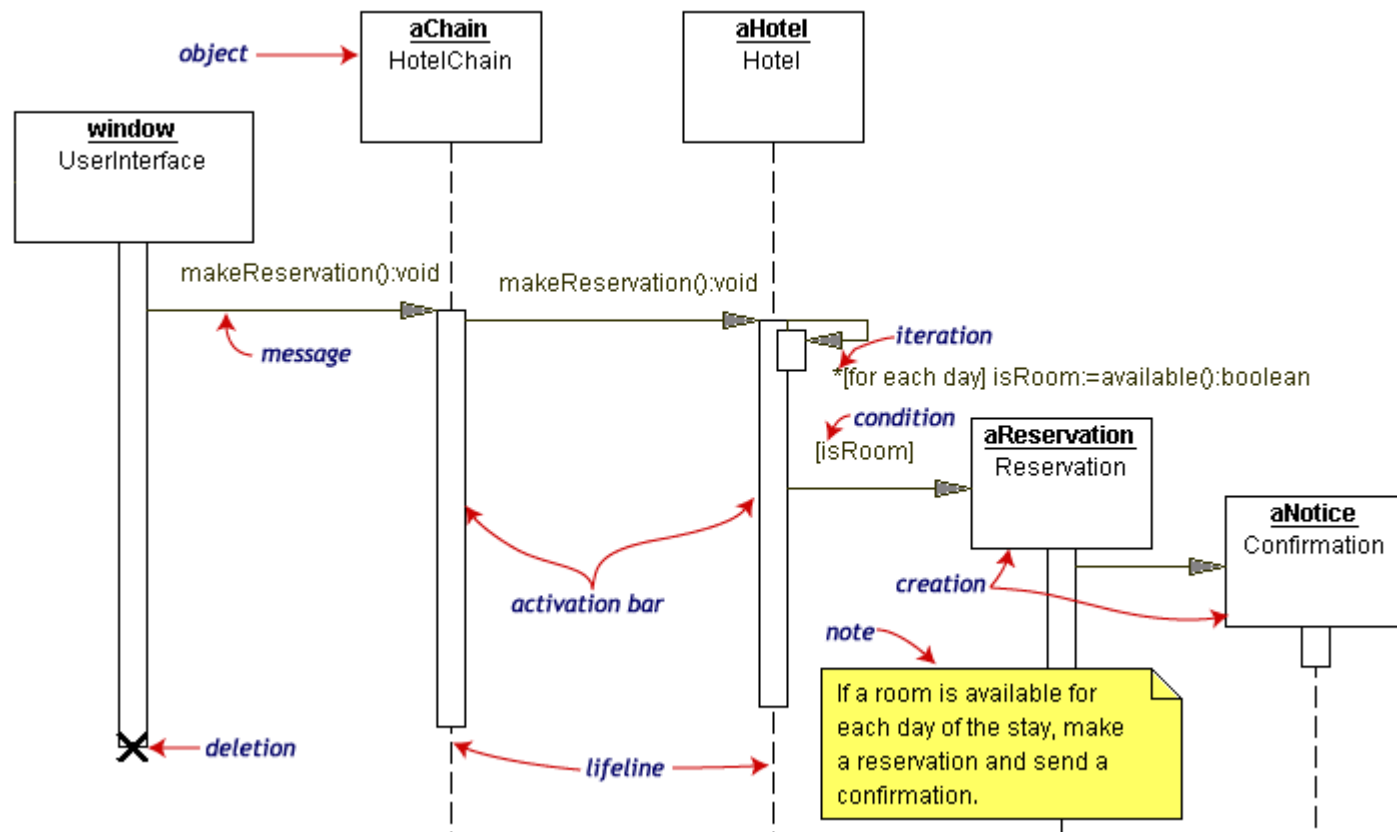
# 开源框架项目分析与实践

□采用UML绘制的类图和时序图等图来分析源码，说明架构和流程



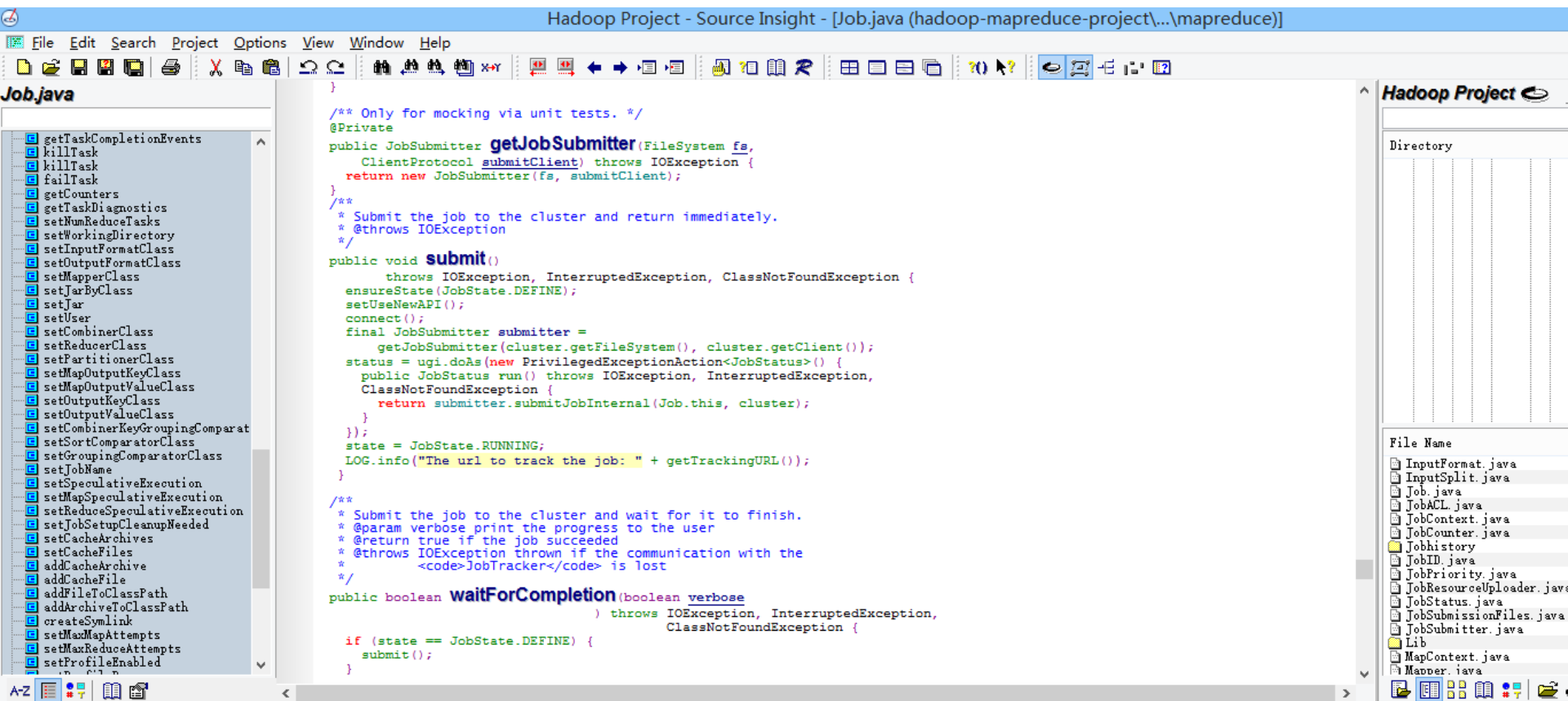
# 开源框架项目分析与实践

□ 采用UML绘制的类图和时序图等图来分析源码，说明架构和流程



# 开源框架项目分析与实践

## □可采用SourceInsight来分析代码



submitJobInternal Method of JobSubmitter in JobSubmitter.java (hadoop-mapreduce-project\...\mapreduce) at line 135 (121 lines)

```
/**
 * @throws IOException
 */
JobStatus submitJobInternal(Job job, Cluster cluster)
throws ClassNotFoundException, InterruptedException, IOException {

    //validate the jobs output specs
    checkSpecs(job);

    Configuration conf = job.getConfiguration();
    addMRFrameworkToDistributedCache(conf);

    Path jobStagingArea = JobSubmissionFiles.getStagingDir(cluster, conf);
    //configure the command line options correctly on the submitting dfs
```



# 开源框架项目分析与实践

## □ 开源软件

### ■ Hadoop生态圈中的软件

- Flume等

### ■ Spark生态圈中的软件

### ■ 并行计算框架

- MR, 流式计算(Storm), 图计算(Giraph, GraphX等),.....

### ■ 分布式文件系统

- HDFS等

### ■ NoSQL系统

- Redis是一个开源的使用ANSI C语言编写、支持网络、可基于内存亦可持久化的日志型、Key-Value数据库，并提供多种语言的API。

- *MongoDB* 是一个基于分布式文件存储的数据库。由 C++ 语言编写。旨在为 WEB 应用提供可扩展的高性能数据存储解决方案。

- Impala

- Hive

- Hbase

- .....

### ■ 人工智能学习框架

- Caffe, MXNet等

# 开源框架项目分析与实践

## ■ 大型Web应用构建涉及的系统

- 如负载均衡的nginx，Nginx代码完全用C语言从头写成，已经移植到许多体系结构和操作系统。Nginx有自己的函数库，并且除了zlib、PCRE和OpenSSL之外，标准模块只使用系统C库函数。
- Memcached 是一个高性能的分布式内存对象缓存系统，用于动态 Web 应用以减轻 数据库负载。它通过在内存中缓存数据和对象来减少读取数据库的次数，从而提供动态、 数据库驱动网站的速度。  
Memcached 基于一个存储键/值对的 hashmap。  
memcached 采用 libevent 的事件处理机制



# 开源框架项目分析与实践

## □课程FTP

■ ftp://software:123456@10.15.82.27

□ Slides – 课件

□ Software – 相关软件

□ Homework – 学生上传作业

■ 最终上传文件为“学号-姓名.zip”(finalreport文件夹),内容包括:

□ 分析报告(word)

□ 汇报slide(ppt)

□ 汇报视频（约10分钟，针对PPT内容的汇报）

■ 期间，在5月31日之前上传开题报告(proposal文件夹)

□ 说明要分析什么软件（的什么模块），架构和流程大概是怎么样的

□ 开题报告格式为word，文件名：学号-姓名-开题报告.docx

谢谢！