

## 海量数据处理中的云计算

# C0. 课程介绍

北京邮电大学信息与通信工程学院

2014年春季学期

# 自我介绍

- 刘军，男，属相：龙，星座：双子

- 经历

– 1994 - 2003	北京邮电大学	本科、博士
– 2003 - 2007	IBM中国研究院	高级研究员、部门经理
– 2007 - 2010	欢城科技	创始人、总经理
– 2010 - 2012	神雕网络	研发副总裁
– 2012至今	北京邮电大学	光荣的人民教师

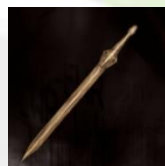
- 特点

- 杂家：研究、开发、教学；项目管理、部门管理、公司管理.....

- 兴趣爱好

- 篮球、滑雪、羽毛球、足球、排球、游戏、旅游、摄影.....

# 课程目标 - 计算机集群解决大规模数据计算问题



**木剑** - 四十岁后，不滞于物，草木竹石均可为剑。

**思想** - 理解实现海量数据计算的关键性问题及解决方法



**玄铁重剑** - 重剑无锋，大巧不工。四十岁前恃之横行天下。

**规律** - 理解Hadoop体系中关键技术的原理、架构、实现



**紫微软剑** - 三十岁前所用，误伤义士不祥，悔恨无已，乃弃之深谷。

**技巧** - 掌握海量数据处理集群优化及高级技术



**利剑** - 凌厉刚猛，无坚不摧，弱冠前以之与河朔群雄争锋。

**技能** - 掌握使用Hadoop技术解决大规模数据计算问题的方法

100小时 → 1小时

2TB数据 → 200TB数据

4G内存 → 400G内存

# 完成课程后的收获



面子



位子



票子

# 2013年美国科技业职业薪酬排名前十

1.	R	\$115,531
2.	NoSQL	\$114,796
3.	MapReduce	\$114,396
4.	PMBok	\$112,382
5.	Cassandra	\$112,382
6.	Omnigraffle	\$111,039
7.	Pig	\$109,561
8.	面向服务的架构	\$108,997
9.	Hadoop	\$108,669
10.	MongoDB	\$107,825

# 需要提前具备的能力

- 理论
  - 计算机原理
  - TCP/IP网络原理
  - 数据结构
- 实践
  - Java编程
  - Linux系统操作



# 课程内容与课时安排

周次	授课	实践
1	C1. 海量数据处理概论	
2	C2. Google云计算核心技术	
3	C3. Hadoop体系架构	E0. 课程设计头脑风暴
4	C4. HDFS	
5	C5. MapReduce ( 1 )	
6	C6. MapReduce ( 2 )	
7		E1. MapReduce作业程序讨论
8		E2. 课程设计方案讨论 ( 1 )
9		E3. 课程设计方案讨论 ( 2 )
10	C7. HBase ( 1 )	
11	C8. HBase ( 2 )	
12	C9. Hive与Pig	
13		E4. HBase作业程序讨论
14	C10. 使用Hadoop的研究实例	
15	C11. 云计算环境下的数据整合与集群管理	
16	C12. 面向未来的海量数据处理	
17		E5. 课程设计结果演讲 ( 1 )
18		E6. 课程设计结果演讲 ( 2 )

# 考核方法

- 平时成绩 ( 50 )
  - 课堂抽查和讨论 ( 10 )
  - Hadoop安装实战 ( 5 )
  - MapReduce实战1 ( 5 )
  - MapReduce实战2 ( 10 )
  - HBase实战1 ( 5 )
  - HBase实战2 ( 10 )
  - Hive/Pig实战 ( 5 )
- 课程设计 ( 30 )
- 笔试成绩 ( 20 )



# 课程设计的内容

- 开题报告（演讲及PPT）
  - 目的：评估课题的内容与难度，集体头脑风暴
  - 内容：问题与数据描述，解决方法和思路，预期结果
- 结果提交
  - 课程报告（Word）：课题描述，解决方法，功能架构，代码说明，结果分析
  - 代码与脚本
  - 源数据与结果数据文件
- 课题总结（演讲及PPT）
  - 讲解课程设计结果

# 课程资源

- 刘军：[liujun@bupt.edu.cn](mailto:liujun@bupt.edu.cn)
- 交流群号：296416804
- 参考资料：
  - **《Hadoop大数据处理》，刘军，人民邮电出版社**
  - 《Hadoop权威指南》，Tom White
  - 《Hadoop实战》，Chuck Lam
  - 《HBase权威指南》，Lars George
  - 《HBase实战》，Nick Dimiduk
  - 《Pig编程指南》，Alan Gates
  - 《Hive编程指南》，Edward Capriolo
  - 《鸟哥的Linux私房菜：基础学习篇》，鸟哥

## 问题？

## 自我介绍

- 学院、专业、研究方向
- 为什么选这个课？
- 想用Hadoop做什么？