

美国大数据工程师面试攻略

董飞

2014-09-07



大纲

- 个人介绍
- 硅谷高科技公司介绍
- 大数据相关技术
- 求职经验
- 面试流程
- 面试准备
- 选择工作



个人背景

- 软件工程本科@



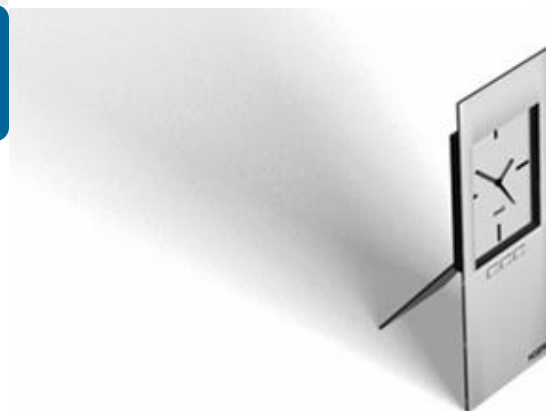
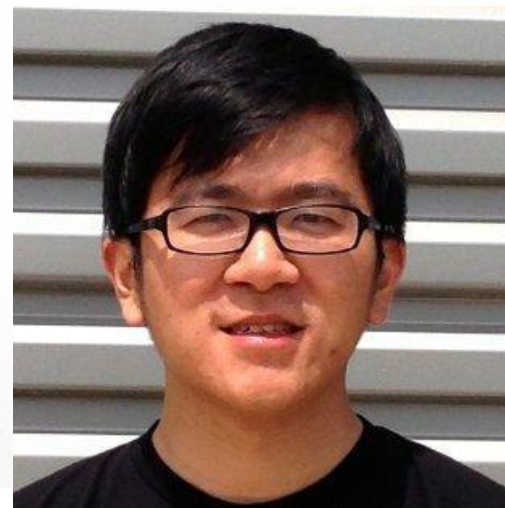
- 计算机硕士@



- 国内工作三年@

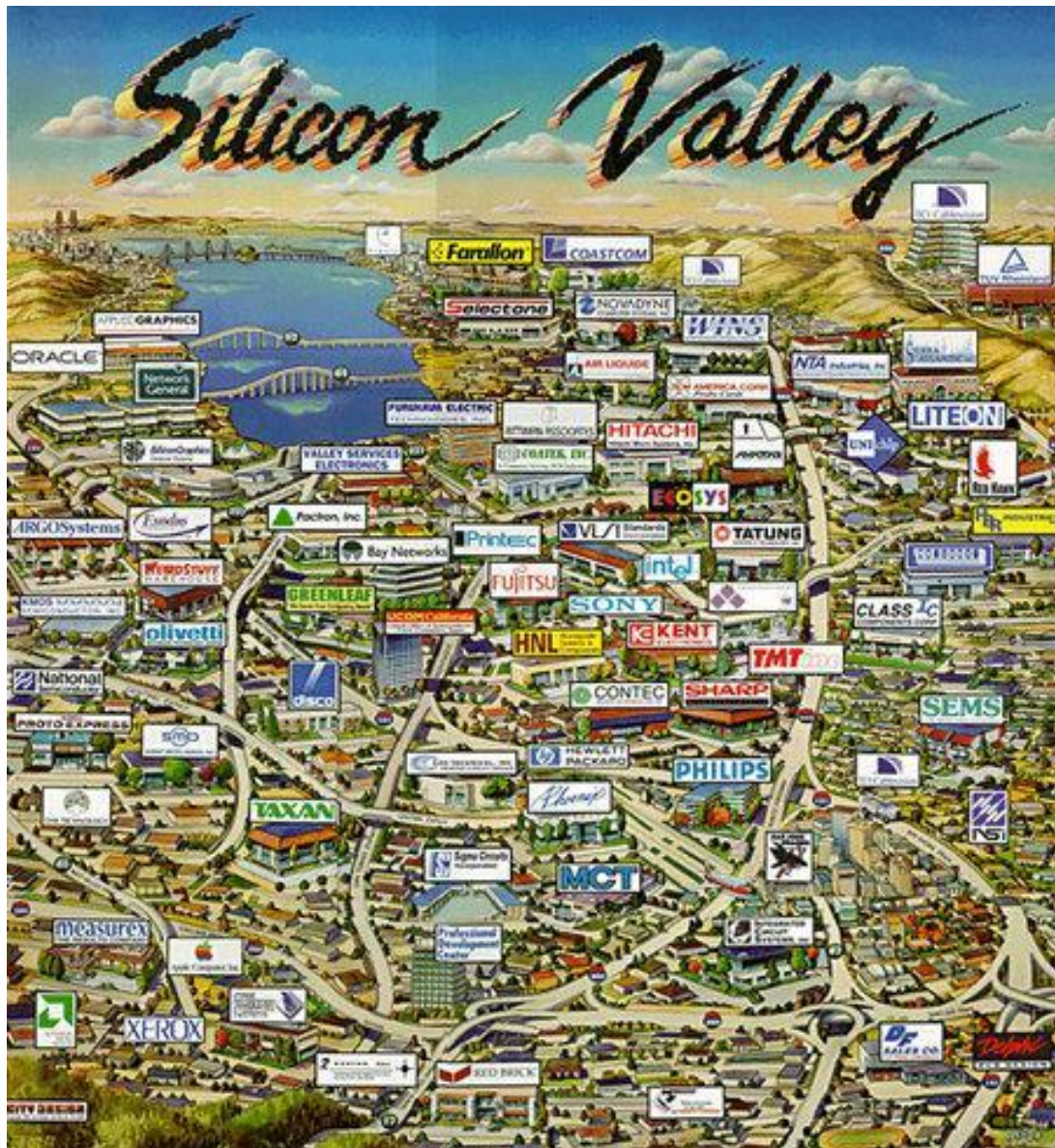


- 美国工作两年+@



硅谷高科技公司介绍1

- 20年前
Intel, Oracle, AMD
Apple, Cisco
- 10年前
Google, Yahoo, eBay
- 现在
Tesla, Facebook
Twitter, LinkedIn

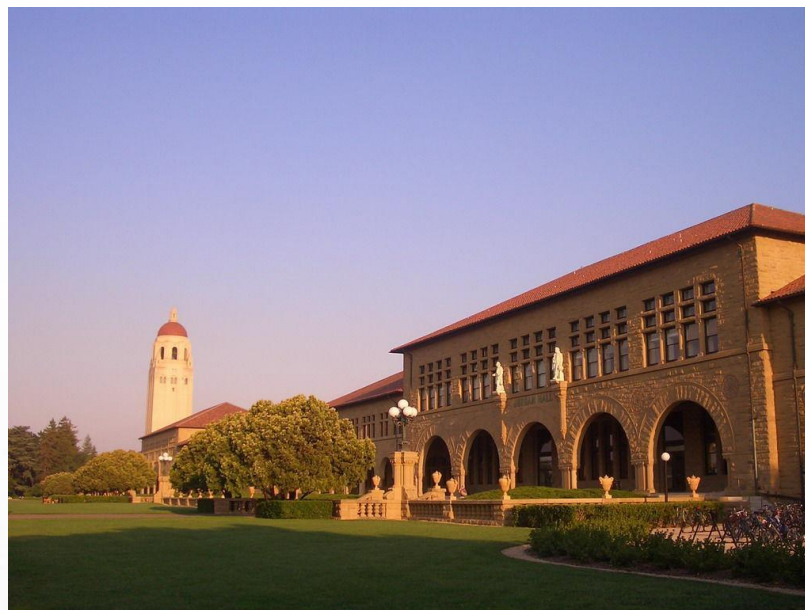


硅谷高科技公司介绍2

- 创新源泉
高校，人才！

- 创业公司
机会，VC

- 未来
NEXT BIG THING



硅谷高科技公司介绍 - LinkedIn

- 技术公司(FLG)
 - Vision: connect the world's professionals to make them more productive and successful.
- Employee Benefits
- Data Driven Products
- Business model



硅谷高科技公司介绍 — Startup

现在什么公司最火？云计算，大数据，consumer internet，健康，通讯，支付，生活类app

<http://graphics.wsj.com/billion-dollar-club/>

TOTAL EQUITY FUNDING		LATEST VALUATION	LATEST VALUATION DATE	YEAR FOUNDED	ROUNDS OF FUNDING
Uber	\$1.6 billion	\$18.2 billion	June 2014	2009	5
Airbnb	\$0.8 billion	\$10.0 billion	Apr. 2014	2008	7
Dropbox	\$507 million	\$10.0 billion	Jan. 2014	2007	5
Snapchat	\$150 million	\$10.0 billion	Nov. 2013	2012	3
Xiaomi	\$347 million	\$10.0 billion	Aug. 2013	2010	4
Palantir	\$0.9 billion	\$9.3 billion	Sept. 2013	2004	7
Jingdong	\$2.2 billion	\$7.3 billion	Nov. 2012	2004	4
Pinterest	\$0.8 billion	\$5.0 billion	May 2014	2008	8
Zalando	\$49 million	\$4.9 billion	Oct. 2013	2008	8
SpaceX	\$115 million	\$4.8 billion	Dec. 2012	2002	4
Cloudera	\$1.2 billion	\$4.1 billion	Mar. 2014	2008	7
Spotify	\$521 million	\$4.0 billion	Nov. 2013	2006	6

硅谷高科技公司介绍 — 大数据

- Hadoop 公司: 三巨头

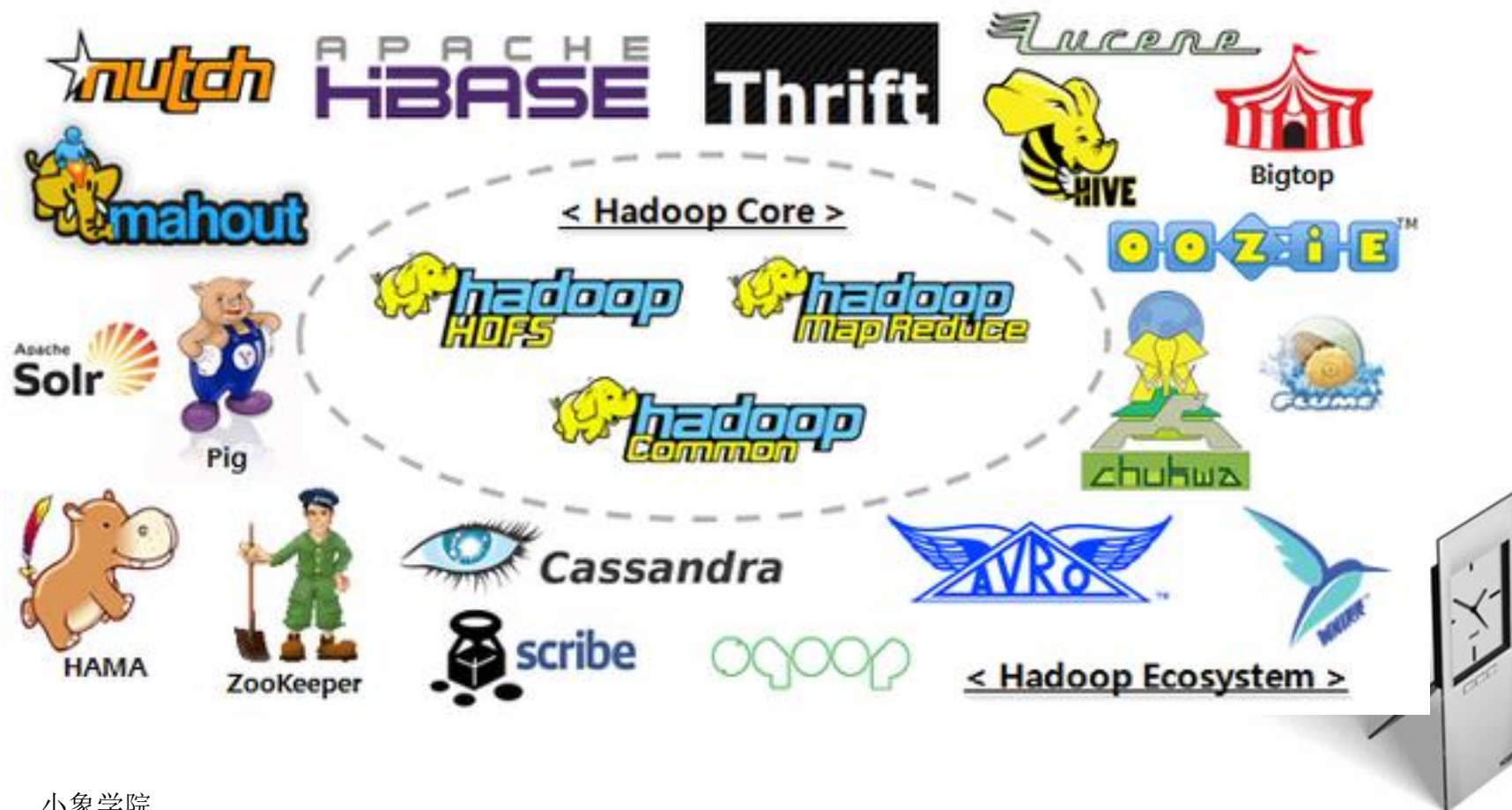
cloudera



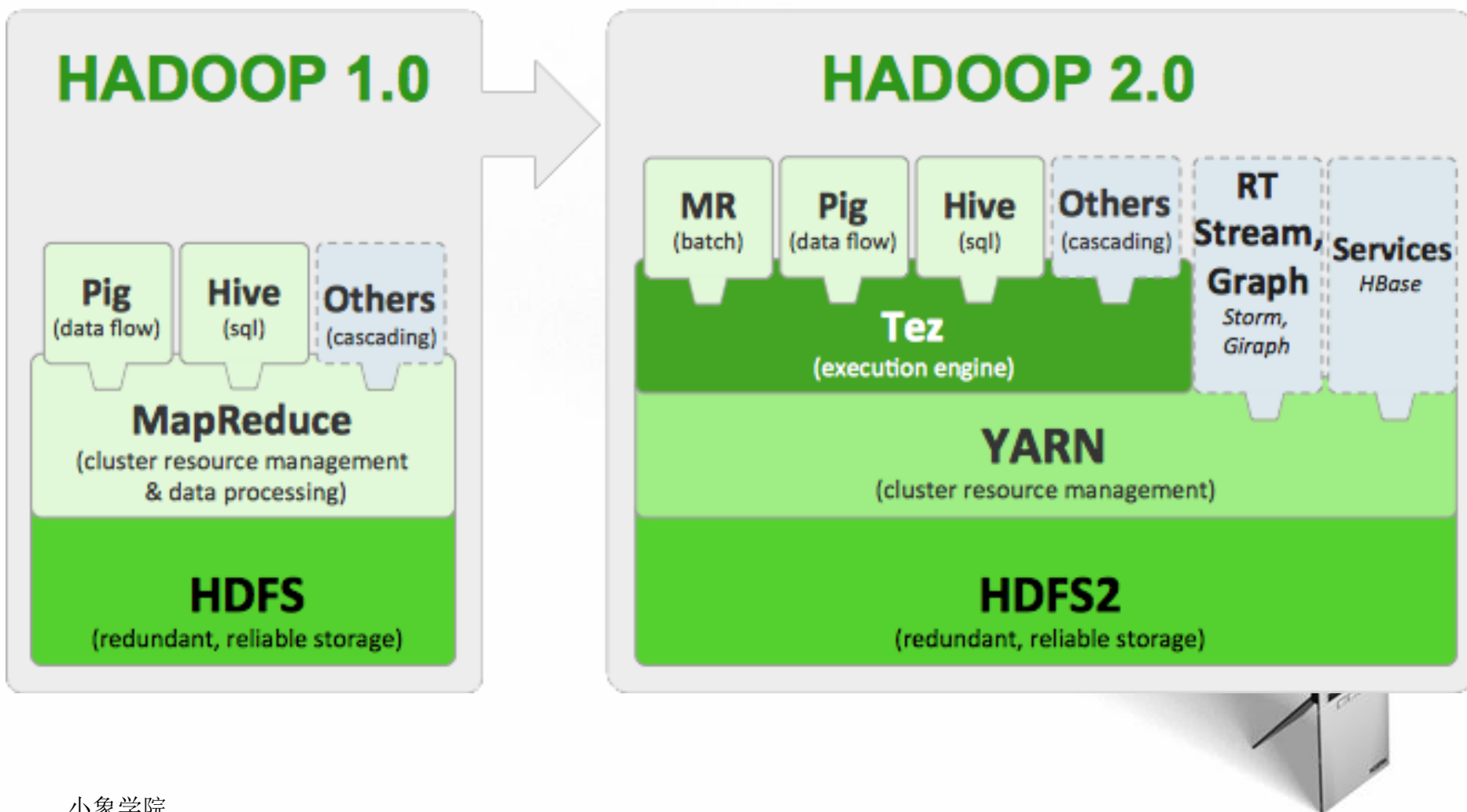
- 2014 新星: Spark



大数据相关技术 - Hadoop Ecosystem

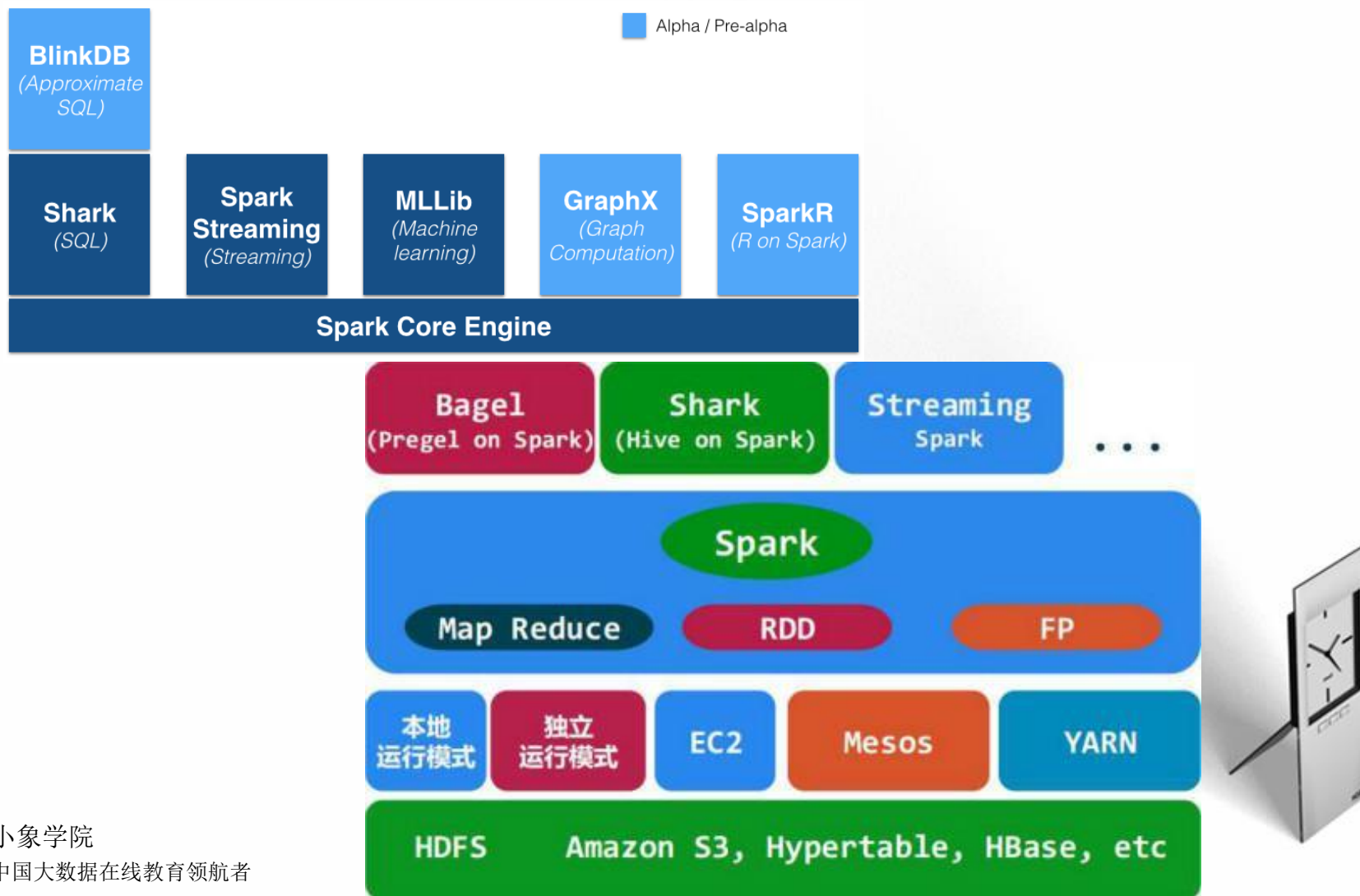


大数据相关技术 - Hadoop2.0

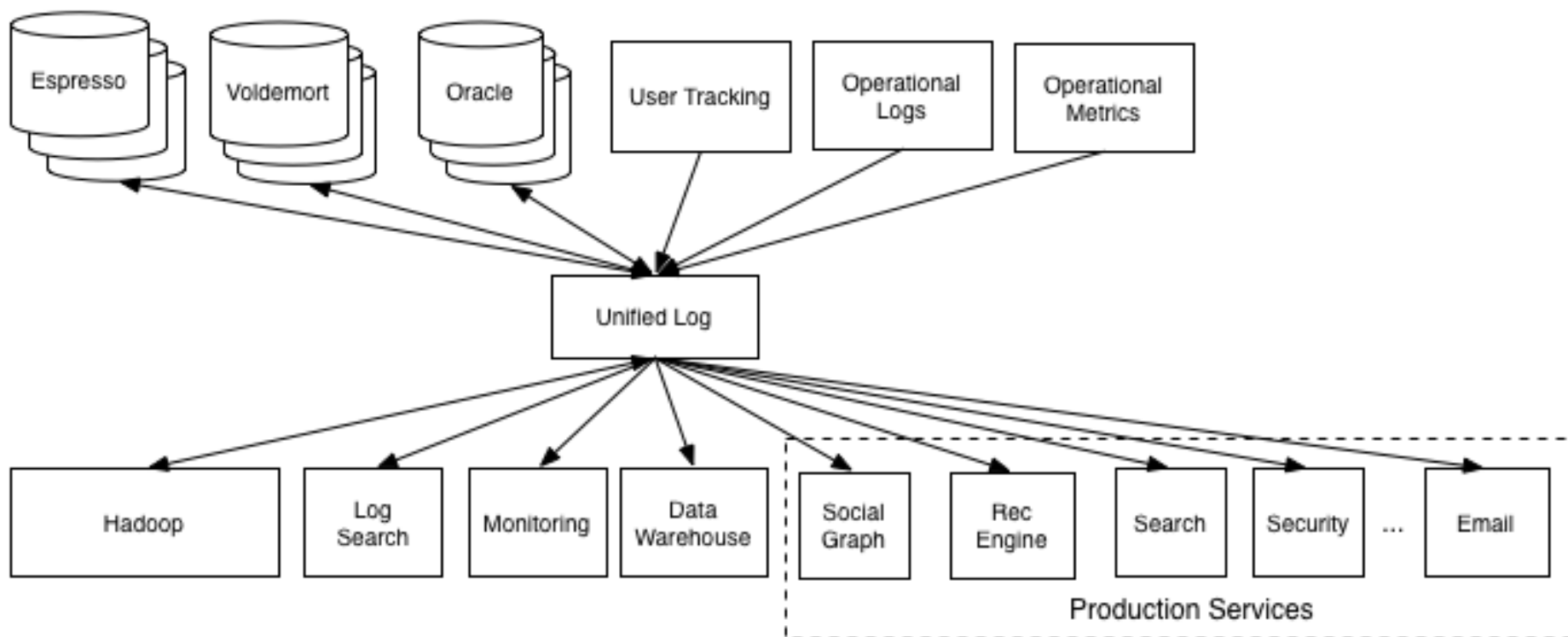


大数据相关技术 - Spark

Apache Spark: killer or savior of Apache Hadoop?



Central Data Pipeline: Kafka



北美求职经验

- 学历<学校<实习<工作经验
- 简历修改，技能匹配
- 面试准备，专业知识
- NETWORK，语言，文化



面试流程

- Phone Interview (1-2 round)
- On-site interview (4-5 rounds)
- Technical Questions
- Communication skills
- Behavior Questions



面试准备1

- Coding
 - 能否写出无bug的Strstr()
- 算法，数据结构
 - Hashtable, Heap, Trie
- 系统设计
 - 数据库，网络，分布式
- Misc
 - 概率，排列组合，Linux



- 知道MapReduce大致流程
- 知道Combiner, partition作用
- 搭建Hadoop集群, Master/Slave 都运行那些服务
- Pig, Hive 语法, UDF写法
- Hadoop 参数调优, 性能优化
- ...

<http://www.zhihu.com/question/24965053/answer/29612377>



高频题:

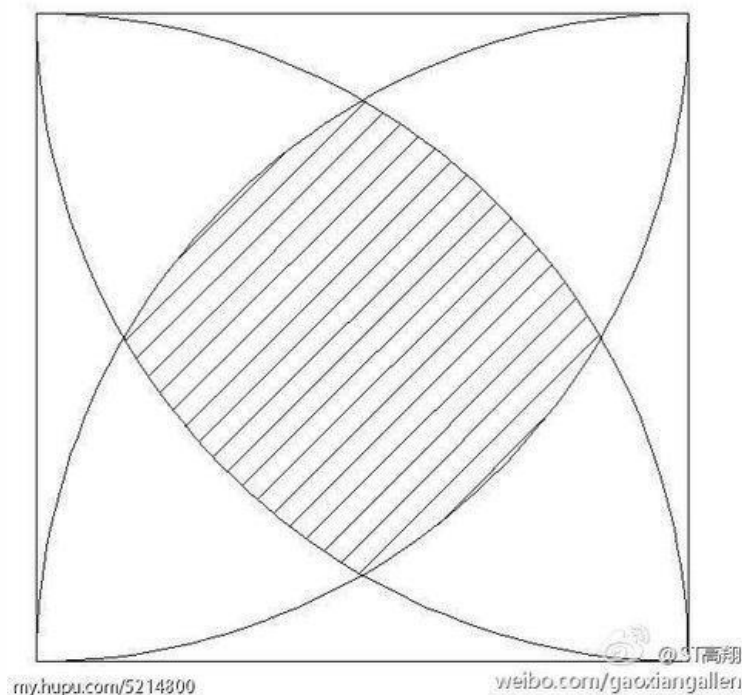
- String: strstr, wordcount
- Tree: height, lca, balance tree
- Heap: 查找最大的k个数
- DP: 最大连续子串和
- Array: find a key in rotated array, 去除重复字符
- Linkedlist: 是否有环, 插入结点, 删除重复结点

<http://www.zhihu.com/question/24964987/answer/29611950>



有趣面试题

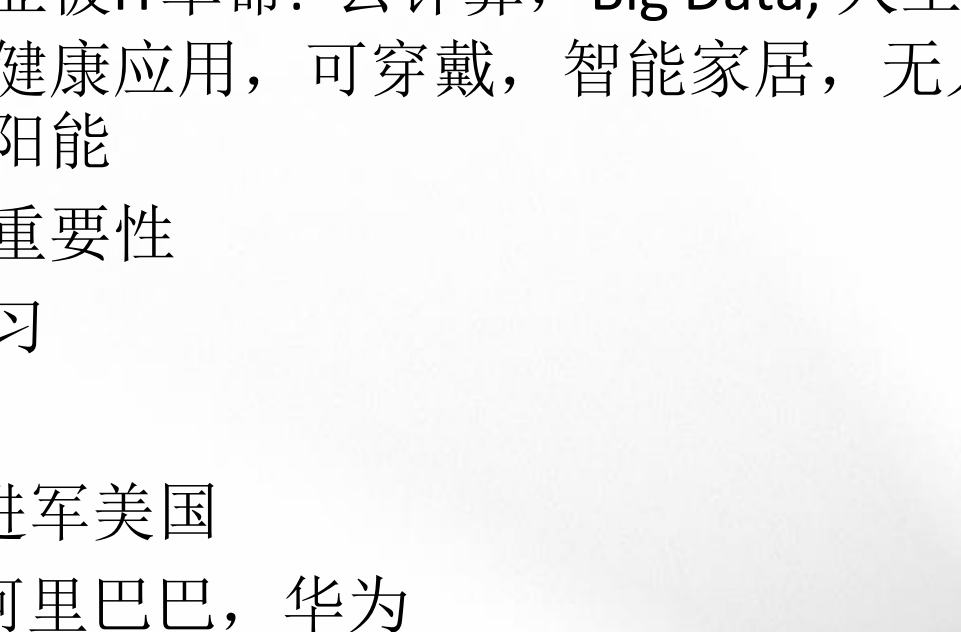
小明拿着两个相同的转基因鸡蛋站在100层的大楼上。鸡蛋可能很结实，从楼顶掉下也不会摔破；可能很易碎，在一楼摔下就破碎。那么，请问小明最少试验多少次可以找出鸡蛋不会被摔碎的最高楼层



选择工作

- 公司口碑
 - Glassdoor 员工评价排名
- 上市大公司，Pre IPO，技术型，面向企业型，大众消费，移动
- 待遇福利
 - Base, 奖金，股票，期权
- 兴趣，成长平台，压力



- 硅谷创新一直在延续。。。 
 - 传统行业被IT革命：云计算，Big Data, 人工智能，在线教育，健康应用，可穿戴，智能家居，无人车，电动车，太阳能
 - 人才的重要性
 - 不断学习
- 中国公司进军美国
 - 百度，阿里巴巴，华为



联系方式

- dongfeiwww@gmail.com
- www.linkedin.com/in/dongfei
- <http://www.zhihu.com/people/dongfei>
- Weixin: chasedream0



谢谢

Q&A

