1 各公司面试题

### 1 触宝科技

#### 1 进程间通信

https://www.cnblogs.com/liugh-wait/p/8533003.html

应用场景：数据传输、共享数据、通知事件、资源共享、进程控制

通信方式：

管道：

普通管道

流管道

2 快速排序

### 2 国双科技

对项目的提问：

spark streaming 窗口

spark streaming结果写进impla，mysql等应该不是Hbase

Hbase存储原始数据而不是sparkstreaming处理后的数据

Hbase rowkey设计

Hbase处理数据热点（类似于数据倾斜）

触宝面试

写了个字符串转整数的代码

topk

堆排序

https://blog.csdn.net/weixin\_37874869/article/details/82988462

unity面试总结

1）数据结构搜索方法总结

2）各个算法的时间复杂度要清楚

复杂度为Olog(N)的算法

3）二叉树的补充

4）十六进制转换成二进制

编程题三道题

触宝面试

1）字典树

2）进程间的通信：socket等

socket编程实践

3）比快速排序更快的排序方式：基数排序-->了解下

4）新版本spark的使用

5）hive数据倾斜

6）多线程Thread和Runnable的区别

7）index索引，常见的索引

hive操作查询，取中位数

拼多多面试

1）写一个死锁的代码实现；实现 Thread

2）NoSQL hive 分区的原理

3）spark数据倾斜的处理方式

1）spark streaming实践

2）java网络编程

3）java多线程

4）java虚拟机

5）巩固hadoop 和 spark笔记

平安普惠面试

1. 知识库系统的设计思路

MVC角度

zookeeper的功能、和区块链的了解、zookeeper和区块链的联系

作者：Skye\_kh  
链接：[https://www.nowcoder.com/discuss/144976](https://www.nowcoder.com/discuss/144976" \t "_blank)  
来源：牛客网

# 秋招总结

秋招结束了，从六七月份开始到十一月，从艰难的准备阶段，到忙碌的面试奔波，感慨良多。我面的基本上都是杭州的公司，目前拿到的offer有，网易互娱、雷火、华为、海康、拼多多、滴滴、51，大部分是数据研发岗位。写个经验总结,回馈牛客。

## 岗位选择

在准备复习之前首先应该确定的是自己的岗位方向，然后再针对性地去寻找面经和资料复习。我当时可选的是数据挖掘算法和数据研发两个方向，因为不同的公司对岗位的具体要求有一些差别，导致我有些迷惑。我自身因为工程和算法比赛都接触过，但是在算法原理方面了解的还是比较少的，所以当时有一些退缩了，选择了数据研发方向。其实机器学习的薪资在今年来看更诱人，从我的经历来看我觉得也能找到不错的工作。当然这都是后话，确定岗位之后就是怎么准备了。

## 复习准备

我大概是从六月开始准备，刚开始最大的困难就是不知道该如何准备效果好。看到其它实验室的同学年后就开始准备了，并且秋招的提前批七月左右就开始，很着急，其实现在想来，**踏实复习**才是关键，不能因为着急而导致知识点一知半解。大厂对基础的要求相对来说更高，其次是自身的亮点。

面试中第一关一般都是手撕代码，其实代码题把**剑指Offer**过两遍就能应付大多数面试算法了，因为面试的时候不可能让你写很长的代码。学有余力去刷 leetcode。

其它知识点可以先把相应岗位的面经都保存下来，再去总结重要的知识点。知识点形成体系比较重要，可以画个体系图不断回顾。比如，Java 的JVM部分，集合类部分，并发部分，分开整理形成知识体系。把自己保存的面经上的问题都准备一遍，其实面试就不虚了。

## 其它

* 暑期实习的经历挺重要的，但是我们实验室不给实习比较可惜。
* 提前批的机会要把握好，也就是多一次机会。大厂其实十一月份可能还有补招。
* 早点准备，将自身的亮点表现给面试官。

## 面经

分享自己面的几个比较典型的面经。

### 蚂蚁金服 数据研发 五面

蚂蚁是我第一个面的，也是最可惜的一个，很后悔当时自己没有准备好。大概面了五面，但是因为当时复习不够好，面试表现不够好，没能拿到 offer。 阿里都是电面，前后时间拉得也比较长。

一面

1. 自我介绍
2. 项目
3. 设计支付宝 APP 滑动条的推荐系统，算法，架构实现
4. 用户轨迹如何获取特征，点击条目 7-5-2-4 这种， word2vec 怎么构建句子特征
5. GC
6. Linux 10个进程10个线程，系统如何调度
7. 线程池，poolsize 5 , 进来第六个线程，会发生什么
8. Java 锁机制，悲观锁、乐观锁
9. 数据结构题，青蛙跳台阶，二叉树层序遍历

二面

1. 天池比赛项目及思路，为什么选择 XGBoost
2. 推荐系统是什么，给完全不懂的人讲一讲 （按分类讲各种算法）
3. 协同过滤的缺点，现在流行怎么做 （协同过滤目前更多用于召回）
4. 讲一讲逻辑回归，原理及推导
5. Synchronized。 锁在操作系统级别的实现
6. Java 中的并发包
7. Linux 进程调度
8. 网络相关协议
9. Zookeeper 选举算法，Paxos
10. 用命令行敲命令存储一个文件到 HDFS 上，源码级别是怎么样一个流程
11. Linux 的基本命令，如何查看内存使用等
12. 动态规划和分治法的区别

三面

1. 项目自己发挥，追问细节
2. 觉得自己是个怎样的人，别人觉得自己怎样
3. 有过比较大的挫折吗
4. 自己做过最有成就感的事情

四面

1. 讲一个项目
2. 项目中的细节追问（线上 A/BTest 怎么做，对结果怎么优化）
3. 项目中自己做的亮点
4. 一个环形链表里面存数字，如何最快速度找到指定数字
5. 一个文件，里面是英文小说，统计词频最高的两个词
6. notify() 和 sleep() 的区别（应该问错了，wait() 和 sleep() ），在源码中哪里看到过 (阻塞队列)
7. 线程池，corePoolsize 5 ,max 200 ，什么时候线程池中有10个线程
8. 阻塞队列满了，线程也超过 200 了 ，会发生什么
9. 线程池实际怎么用
10. 你有什么要问我的吗  
    11.为什么没去实习，是不是从现在到毕业都不能出去

五面

1. 自我介绍
2. Github 库做了什么
3. 各个项目做了什么
4. Mapreduce Join
5. Mapreduce shuffle
6. 远程代码题， 输入一个 n ， 求 1~n 中每个数字二进制位中 1 的个数

### 网易杭研 大数据开发

一面

1. 自我介绍
2. Mapreduce 原理
3. 数据倾斜怎么办
4. 比如统计广东省的记录数据倾斜，怎么自定义 partition ，代码？（我觉得我说得挺清楚了，一定要代码写出来）
5. 推荐系统项目细节，角色，做了什么
6. 华为软件精英挑战赛，如何调参
7. 代码，两个集合取交集，怎么最快
8. 字符串反向输出，用递归实现
9. 知道什么ETL工具
10. dataX架构

二面

1. 自我介绍
2. 项目细节
3. Mapreduce 原理
4. splitsize 如何确定，算法
5. 数据倾斜怎么办
6. 无 shuffle 的 过程叫什么
7. 大数据组件都知道哪些
8. 数据仓库和数据库的区别
9. 数据库原理懂吗
10. 数据库范式知道吗
11. sql 题，按某列排序，并增加排序编号（说了 mysql 怎么做，hive 有点忘了row\_number没说明白，可惜了)
12. sql 题，一列从1到n顺序编号，一列是各种类别，输出某一类别编号是连续四个以上的数据行
13. 代码题，两个集合取交集....，有哪些做法
14. 平时有什么爱好

### 网易雷火 数据挖掘

一面

1. a += 4 和 a =a+4的区别，a是byte类型
2. Integer 数据比较
3. python 两个列表合并去重排序（python 很久不敲，基本语法忘记很多了）说了下怎么做。
4. Linux 两个文本不同的统计，删除目录文件，删除重复行
5. 代码 java 字符串反转

二面

1. 推荐项目
2. 一条sql 外表是 gz格式压缩，查询会有几个 map 几个 reduce
3. 什么压缩格式解决gz不可分片问题
4. 集群 mr任务分配到某个节点一直 timeout 什么问题。 （网络，防火墙问题）
5. Github

三面

1. 推荐项目
2. 代码 输出连续数字和等于某个值的所有情况
3. 知道的算法
4. LR 原理，为什么用极大似然
5. 随机森林原理
6. Bagging原理
7. GBDT和XGBOOST区别
8. 比赛为什么选用XGBOOST
9. Boosting

四面

1. 介绍自己，别说技术
2. 自己会给雷火带来什么收益
3. 自己最不能接受的状态，会考虑离职的
4. 推荐系统的应用一堆扯

### 51 大数据开发

* 项目细节，为什么用kylin，知道其他etl工具吗
* sqoop和datax区别
* 推荐项目数据量多大，日增多大
* 搭建集群碰到过什么问题
* mapreduce 二次排序怎么做
* mapreduce原理，调优做了什么
* mapreduce比较慢，你认为哪些地方可以优化
* hadoop 和spark的区别
* Hbase 知道什么，和mysql的区别
* Hbase 支持事务吗
* mysql事务隔离级别
* java常用集合
* hashmap 原理，resize过程
* hashmap 为什么会出现并发问题
* AQS
* java 的线程是用户态还是内核态
* 栈里保存了什么
* 什么情况栈溢出
* thread 栈多大
* jvm其他组成部分
* gc了解吗 自己调优过吗
* 直接看代码的？有没有dump出来看日志
* cms什么时候发生stop the word
* 垃圾收集算法
* Hadoop 生态还了解哪些组件
* two sum 问题，说怎么做，不写代码

二面

* 项目各种细节
* mapreduce 原理
* 推荐系统算法了解哪些，商品推荐如何做
* 用户协同过滤和聚类的区别
* 基础机器学习算法
* 线上场景如何评估算法优劣
* 个人优势，遇到的挫折等等

作者：十月offer会收割吗  
链接：[https://www.nowcoder.com/discuss/143458](https://www.nowcoder.com/discuss/143458" \t "_blank)  
来源：牛客网

## NLP

### I 百度NLP部

* 一面：
  + 项目介绍：CRF-loss梯度，self-attention原理公式，LR和SVM原理和异同；有向图模型标注偏执问题(n-gram);
  + 代码：k-group 链表反转；
* 二面：
  + 项目：charCNN-emb，BiLSTM-CRF结构介绍，CTC原理，word2vec和GloVec和fasttext；
  + 算法：GBDT和Xgboost，同为boost和adaboost的关注点异同，LSTM结构，为神马改善梯度消失；
  + 代码：二分、10亿数据中O(1)定位；
* 三面(经理面)：
  + 项目：NLP和CV项目(标点，实体标注、OCR，视频分类，理解，目标检测，语义分割)
  + 代码：单词翻转，不考虑空格；
* 部门四面:
  + 开放性问题，是否了解机器翻译，小语种机器翻译如何实现良好的训练；
  + 样本不平衡问题，ohem\_loss,focal\_loss; smooth\_L1\_loss公式，mAP计算；
  + 代码：从(0,0)点出发，点阵中有障碍，求到(m,n)的最短路径，BFS；

### II 好未来AI lab

* 一面：
  + 项目介绍：seq2seq-attention原理和公式，soft、hard-attention;
  + lattice-LSTM，解决char分词问题；
  + 代码：删除链表节点O(1)、3sum；
* 二面：
  + 项目介绍；
  + attention推导，RNN梯度消失的罪魁祸首，LSTM为什么改善梯度消失，梯度更新；
  + EM算法步骤，为什么Q(z)为后验概率(下界)，E-step(更新Q(z)),M-step(更新估计参数)；
  + PLSA，LDA介绍，有向图介绍(根据图方向判断生成模型)
  + gibbs采样和MCMC采样的关系，描述拒绝采样；
* 三面：
  + 项目介绍；
  + 部门做的项目介绍，交流；
  + SVM推导，Xgboost和GBDT和lightGBM区别；
  + 代码：编辑距离，最长公共子序列；

### III 网易互娱AI Lab

* 一面(电话)：
  + 项目介绍；
  + RNN为什么梯度消失，BPTT；LSTM结构，为什么改善梯度消失；
  + SVM为什么要对偶(优化复杂度转变，核化)
  + 1x1 conv介绍，运用场景；
  + 代码：Top K；
* 二面(现场)：
  + 项目介绍；
  + CTC\_loss公式推导；
  + Viterbi解码公式和简单代码；
  + self-attention原理和推导；
  + beam-search decode介绍；
  + 代码：二位矩阵中，每个点可以走四邻域，求最大连续和路径；
* 三面(leader面)：learder人很好
  + 项目介绍；
  + 互娱的基本介绍，以及nlp团队情况；
  + word2vec：层次softmax和负采样的原理和公式；GloVec的推导，解决对称问题；
  + 代码：最大正方形；

### IV 腾讯

* 一面：
  + 项目介绍；
  + viterbi解码原理，code；beamsearch；
  + 代码：一道hash的题，忘了；
* 二面：
  + 项目介绍；
  + BN介绍(为什么加速收敛，从SGD更新角度和weight scale角度)，dropout介绍，训练测试差异；
  + 代码：二分查找的题，绝对值；
* 三面(总监)：
  + 聊项目；

### V 阿里巴巴

* 一面(电话)：
  + 项目介绍，self-attention；
  + xgboost和gbdt区别，开放问题，淘宝评论；
  + 代码：买卖股票不限次数；
* 二面(电话)：
  + 项目介绍；
  + 场景问题；
  + 代码：全排列；
* 三面(现场)：
  + 项目介绍；
  + image caption中的attention；视频场景题；
  + 字典树，B-tree，B+tree；
  + 代码：字典树，share\_ptr
* 四面(交叉)：
  + 项目介绍；
  + 开放问题淘宝评分系统，LDA主题模型定价，面试官被我一波频率学派和贝叶斯学派给逗笑了；

# 知识点总结

LR和SVM原理和区别-- 公式推导

GBDT和XGBoost（boost和adaboost的关注点异同）；lightGBM

word2vec和GloVec和fasttext；

梯度消失：

LSTM结构，为神马改善梯度消失；

RNN为什么梯度消失，BPTT；LSTM结构，为什么改善梯度消失；