



Commando20180403011197 飞黄腾达

2018-11-15 12:27 字节跳动 研发工程师

+ 关注

分享一个大数据的面经 精华

面的公司并不多。拿到了头条和阿里的意向书，如果不出意外应该就是这俩选一个了。

没有实习，没有项目经历，什么都没有，上去就是硬怼。本科不是相关专业，研究生转方向但是是留学生，也不是特别厉害的学校。各方面都比较劣势。

优势就是特别能背。

头条：

一面：面岗位相关的基础，上来面试官直接发现我什么经历都没有，我直接说熟悉hadoop。挑了HDFS让我讲一下。NN和DN。

HA的实现

zookeeper的原理，zk是如何保证一致性的，zk是如何判断session超时，connection超时的。如何触发回调。

client和HDFS文件的读写过程，延迟太高，怎么解决。我当时脑子一蒙，直接说了设计的就是高吞吐量的文件系统。

yarn的结构，RM和NM的交互，如何分配任务的。

yarn在什么层面调度，和k8s和mesos有什么区别，内存调度是什么怎么调度的，如果考虑CPU怎么调度的。如何实现隔离的，Control group 和Namespace是怎么回事。

手撕算法，链表排序。

人生中的第一次面试。。。面的一团糟，多亏面试官使劲高抬贵手。。。

二面：面的计算机基础：

上来手撕，剑指原题，两个栈实现一个队列。

问了一下转专业的经历。

讲个简历的项目。

操作系统中的进程和线程的区别。

JVM的内存模型说一下，堆，栈，永久区，GC。

Java的多线程。

Hadoop的MR和Spark有什么区别，为什么Spark有优势。

TCP

用过什么数据库啊，然后就没往深了问了。

手推神经网络上的反向传播，把链式求导写出来，激活函数sigmoid。

介绍一下LSTM，和GRU的区别。

过。

三面：技术深度，问得特别特别细

看过源码对吧，hadoop的HDFS，Yarn，MapReduce源码讲讲，精确到方法和类。

客户端和NN的通信过程，和DN的通信过程。

HA中的***过程，如何防止脑裂，为什么需要fencing，SSH连不上怎么办。

checkpoint的过程

yarn的状态机和epoll

yarn的公平调度器和容量调度器和FIFO

yarn抢占

MapReduce全过程，分片怎么读的，为什么用快排，换别的行不行，多路归并怎么实现的，环形缓冲区怎么实现的。

场景，很多牛逼的电脑，大CPU大内存，网卡超烂，如何优化？强制本地化，选高压压缩的序列化格式，核心就是尽量减少网络IO。

场景，一个加载在内存里的HashMap，Key和Value全是int，从硬盘读进来只做查询不做修改，不考虑查询效率，尽可能提高空间效率，稍微考虑一下时间效率。从JVM调优到数据结构到GC到。。。。

手撕，实现一个线程池。

在spark和hadoop中如何处理数据倾斜。

spark的调度过程，DAGScheduler如何划分stage，TaskScheduler如何调度任务的。Spark的shuffle是什么样的，怎么优化的。

评价一下Hadoop和Spark的代码风格。我曹。我随便说了说。怕送命题。幸好我说完了他就自己说了。疯狂点头。

7

☆ 315

💬 29

🔗 分享

⋮

回复帖子

全站热榜

- 1 关于计算机专业法网?
- 2 腾讯软件后台开发面 40分钟 挂
- 3 美团面试 (已OC
- 4 0419阿里笔试代
- 5 阿里笔试419 451解
- 6 请教如何刷题、i
- 7 2023春招实习面 (已接阿里offer
- 8 4月19日 0offer
- 9 Java-2022暑期2023春招部分面
- 10 央企我感觉很推荐人去试试国家电网

正在热议

- #2023届毁约公司名 3515次浏览
- #滴! 实习打卡# 3033次浏览
- #五一调休情绪发泄 2523次浏览
- #阿里巴巴信息集散 2666次浏览
- #通信硬件2023笔面 3711次浏览
- #现在前端的就业环 17654次浏览
- #如何拒绝/ 1071次浏览
- #没有实习经 9708次浏览
- #要钱还是要 1396次浏览
- #我的成功项 7055次浏览
- #你收到了团子 6638次浏览
- #如何一边实习一边 5990次浏览

Java的Object谈谈

Java的容器谈谈

Java的NIO用得咋样，不咋样。。。那就不聊了。

MPI聊聊

你Scala用得怎么样呀，Python怎么样啊

设计，设计一个RPC框架，我看你简历上做过，谈谈思路，观察者模式讲一下，并发容器。

设计，现在有一个RPC框架，需要使用线程池，多次复用socket，TCP，怎么传递命令。变相考hadoop，使用操作码。

结束。这一面的面试官超强，被压着打。

四面：Um.....和我这个岗没有一毛钱关系。

手写一个leetcode hard，就是两个数组，找全局中间数那个。

问问搜索。

如何构建倒排索引

BM25

如何加速查询，WAND算法细节。

倒排索引的log时间合并

如何评价一次query，设计一个机器学习模型，如何设计这样的模型，如何提取feature，用什么模型

如何做查询补全 Trie

设计一个查询缓存

如何对倒排索引进行空间压缩，我先说了OptPForDelta

如何选择最优分裂点，抽样，然后算。

手撕个代码实现一下

如果不抽样，你这样是全局最优吗？我说是啊，他说不是。。。对每一个算一个。好吧，他说得对。

还有别的优化方式吗 Variable Byte

聊家常。。。出了门感觉人生都灰暗了，跪

美团：数据开发，就是数据仓库开发，好多重复的我不写了。已经拒了

一面基础：

讲讲项目

java基础

HTTP和TCP

写俩算法，可以用IDE。都是超简单的，有一个台阶，还有一个m的n次幂，自己写个测试试一下

写个sql

mysql的索引，如何实现的，何时失效，聚簇索引和非聚簇索引，B+树

写个单例，懒汉饿汉。

处理数据倾斜

Spring系列会吗？不会

二面

hadoop和spark简单讲讲，只说结构原理不说源码

spark的调优

JMM，全程不打断，我一口气说到了底。。。

数据仓库的星型模型和雪花模型，上卷下钻

Hive

还有啥忘了

spark sql，窝不会

三面

讲项目

HTTP，TCP

操作系统的进程与线程

Java多线程，线程池

爬虫讲讲原理（我项目里用了）

场景：实现一下chrome里那个历史记录这个功能，用啥数据结构怎么实现

python，MPI，数据仓库，你都会点什么都说说

问了项目里的数据清洗，补全和NLP，然而我的NLP一点深度学习没用，感觉对面可能有点失望。

问了推荐系统

#机械制造2023笔面
3645次浏览

#23届找工作求助阵
122246次浏览

#京东信息集散地#
1631次浏览

#国企/银行/研究所
12635次浏览

#Offer是寒冬太冷
58843次浏览

#为什么那么多公司
6449次浏览

#工作丧失热情的瞬
8142次浏览

#你觉得今年春招回
33155次浏览



阿里：阿里在我投简历两个月之后捞我，而且面试简单得出奇，只考广度不考深度。。。我还是犯了好多错误，提醒大家，如果面试这种面试官不深问的，一定要自己使劲说，不要说两句就停了，他觉得你只会这么多。

一面：这个面试官是个做算法的

说个项目

Java基础，说说容器

多线程，线程池

java的序列化方式，hadoop的序列化方式，Avro，parquet，transient关键字

spark的原理，我就说了个架子等他问，他不问了！

设计模式，观察者，hadoop用了啥，我说用了组合，等他问我第一第二关系的事，他不问了！

业务题，拿spark实现个业务逻辑，拿嘴说就行。map然后reducebykey然后。。。

一个白板代码，是他工作业务逻辑的抽象。

还问啥我给忘了

二面三面开始集中面试，连着面了我2小时，我一起写了。

说个项目，问我项目实现细节。工程细节，你这怎么加速，Geohash怎么用的。并对我的辣鸡项目表示了鄙视。

spark streaming和storm，flink的区别

各种分布式计算框架的区别

storm如何处理反压，如何保证流的可靠性的

flink会吗，不会

hbase的row key设计，我现在就要范围查询value怎么设计。

yarn里面机器崩了，怎么让任务接着算。。。我说了重启nodemanager重算这个分片，如果这个分片也不想重算呢，接着算，没这功能没说出来。回来想想其实还是应该能说个思路的。

说了AppMaster的故障怎么实现的HA

zookeeper是如何***新的Active Namenode，没仔细说，说了个原理他又不继续问了，后悔。

大数据算法：10亿个int查重，讲思路，布隆过滤，mapreduce，bitmap，算法的复杂度说说

机器学习竞赛跟我说说，讲了特征工程，提了GBDT，然而我觉得他并不知道GBDT是啥，讲完和我说你这就是java的hello world啊。主要也是自己讲的不好，我也很是哽咽

HR面：

阿里这个HR面是真的犀利，我也没经验，第一次面这种话里带枪的HR面。

中间委婉谈及了我的学历，我的水课，我的没实习没经历。。。被贬得一无是处，我也不敢顶她，只能避重就轻解释一下。

现在回想可能表现得太没自信了更减分。

拿到了意向书。。。没有谈薪资，没有加面，感觉要拿劝退价了。

#面经##秋招##大数据开发工程师#

👍 27 ☆ 315 💬 29 📄 分享 ...

浏览 2.8w

在这里输入你的评论吧！

🖼 图片 @ 话题 😄 表情 ☐ 同时转发到我的动态

评论

全部评论(29)

推荐 最新 楼层



NUIBoss 进阶牛

西安邮电大学

你是真滴强

👍 3 💬 回复 📄 分享 发布于 2018-11-15 13:05



buptzyc 大佬已定

北京邮电大学

大佬

👍 2 💬 回复 📄 分享 发布于 2018-11-15 12:44

7 ☆ 315 💬 29 📄 分享 ...





offer快到碗里来

Java

强的一p

2 发布于 2018-11-15 12:49



惟jie无二

算法工程师

这也太难了吧

2 发布于 2018-11-15 13:04



牛客2347492号

蚂蚁集团 选品平台 开发

巨佬tql

1 发布于 2018-11-21 19:24



凶猛肥宅在线面试

算法工程师

出来看神仙

1 发布于 2018-11-21 19:34



哈哈哈哈哈同学

测试工程师

*** 太强了

1 发布于 2022-06-16 21:30



SneakyCrab

阿里巴巴 开发工程师

这是算法岗还是后台岗?

点赞 发布于 2018-11-15 13:20



Commando20180403011197 楼主 回复

大数据开发

点赞 发布于 2018-11-15 13:21



柠檬阿婆主

Java

阿里是刚发的意向么

点赞 发布于 2018-11-15 13:24

Commando20180403011197 楼主 回复

前天吧还是大前天。。。最近

点赞 发布于 2018-11-15 13:29

柠檬阿婆主 回复 Commando20180403011197 楼主

方便问下是甚么部门么

点赞 发布于 2018-11-15 15:02

—— 展开7条回复 ▾

我是小变态。

华为上研 终端 大数据研发(准入职)

大佬还是大佬

点赞 发布于 2018-11-15 14:17

Y0001

Java

大佬太谦虚了.....



牛客331725号

大数据开发工程师

请问楼主 背的哪里的东西

👍 点赞 🗨 回复 🔄 分享 发布于 2018-11-21 19:36

Commando20180403011197 楼主 回复

基础部分是上课的教材。。。hadoop有hadoop权威指南和源码。剩下的随便看看就行了

👍 点赞 🗨 回复 发布于 2018-11-21 21:19

牛客331725号 回复

上课的教材是啥😂

👍 点赞 🗨 回复 发布于 2018-11-22 00:13

爷爷泡de茶

数据分析师

厉害

👍 点赞 🗨 回复 🔄 分享 发布于 2018-11-21 19:42

东北28

Java

大佬牛逼.....

👍 点赞 🗨 回复 🔄 分享 发布于 2018-11-21 22:43

User_19

算法工程师

这都能过大佬是真的6.....

👍 点赞 🗨 回复 🔄 分享 发布于 2019-04-17 08:20

你好逗言莫发呆

字节跳动 data 数据开发

大佬是真的牛逼啊

👍 点赞 🗨 回复 🔄 分享 发布于 2019-08-21 23:06

法国吴彦祖

门头沟学院 机械类

源码牛逼。。。

👍 点赞 🗨 回复 🔄 分享 发布于 2022-06-28 23:27 来自Android客户端

相关推荐

如常r

03-18 21:40 工商银行_产品经理

银行打工3年，干啥啥不行，反PUA第一名

今年是我在银行打工的第三年，刚工作的时候我不仅很爱吃老板画得饼，而且还很擅长自己PUA自己，工作挨了批评全从自己身上找原因，就连生病了也会埋怨自己不该这个时候生病，耽误项目进度，但后来想明白这三 ... [查看更多](#)

你认为工作的意义是什么

👍 45 🗨 3 ☆ 46 🔄 1 ...

枫煦h

03-23 16:50 中国农业银行_研发中心_研发工程师(准入职工)

最终还是违约蚂蚁，选择了农行 [内部员工回复](#) 烫

作为一名普通的23届应届生，经历了上届师兄们的就业狂欢（实验室很多师兄拿了bat全套offer，sp，ssp），也感受过了自己秋招的寒冬。在最终的反复思考之后，选择了违约蚂蚁，签约农行研发中心成都研发部。回顾 ... [查看更多](#)

7 ☆ 315 🗨 29 🔄 分享