



优秀的张先生 大橘已定

2018-09-14 20:18 Java

+ 关注

十余家公司大数据开发面经

背景介绍

楼主5月初入职滴滴实习，7月底得知今年滴滴不好转正，遂离职复习巩固基础+项目备战秋招。因为准备的时间比较早，有幸赶上了各大厂的提前批面试，面试岗位是大数据开发工程师，因为牛客上关于大数据的面经一直很少，所以写一篇大数据面经造福后来人，也记录一下这段难忘的时光。我本科是水利水电专业，研究生转入软件工程，编程语言用Java，做过一个大数据开发的项目，以Kafka作为数据总线，所以Kafka的问题会涉及的比较多，实习期间主要是写hive sql，所以会有比较难的sql题，大家可以结合自己的项目和实习经历进行针对性复习。下面的面经只涉及技术，hr面就不写了，因为不同公司会有重复的问题，重复的问题也不写了。个人觉得不错的大数据学习网站：<http://dmlab.xmu.edu.cn/blog/>

腾讯IEG (offer)

一面 (全程问基础)：

- 1、介绍项目
- 2、String、StringBuffer、StringBuilder的区别，怎么理解String不变性
- 3、==和equals的区别，如果重写了equals()不重写hashCode()会发生什么
- 4、volatile怎么保证可见性，synchronized和lock的区别，synchronized的底层实现
- 5、sleep和wait的区别，sleep会不会释放锁，notify和notifyAll的区别
- 6、不了解线程的局部变量，讲讲线程池参数
- 7、什么情况会发生死锁，死锁的处理方法
- 8、Cookie和Session的区别，怎么防止Cookie欺骗
- 9、从用户在浏览器输入域名，到浏览器显示出页面的过程

二面 (全程问项目，压力面)：

- 1、看你写过UDF，谈谈对UDF的理解，写UDF的目的，代码怎么写的
- 2、改造hive表后怎么进行数据一致性校验的，有没有自动化流程
- 3、看你读过kafka源码，讲讲kafka broker的源码里面你最熟悉的类，以及这个类的主要方法，用的什么设计模式
- 4、项目里面从数据采集到最终的数据可视化，每个环节都有可能丢数据，怎么判断数据有没有丢，如果丢了如何定位到在哪个环节丢的
- 5、项目里面为什么要用kafka stream做实时计算，而不是用spark或者flink，kafka sql和spark sql了解过吗
- 6、项目里面用到了时序数据库opentsdb，为什么要用这个，有没有跟其它的时序数据库对比过
- 7、平时逛不逛社区，有没有参与过开源项目

三面 (接着问项目)：

- 1、看你写了实时计算的程序，你怎么保证计算的结果肯定是对的
- 2、数据接入的时候，怎么往kafka topic里面发的，用的什么方式，起了几个线程，producer是线程安全的吗
- 3、kafka集群有几台机器，怎么确定你们项目需要用几台机器，有评估过吗，吞吐量测过吗
- 4、spark streaming是怎么跟kafka交互的，具体代码怎么写的，程序执行流程是怎样的，这个过程中怎么确保数据不丢
- 5、kafka监控是怎么做的，kafka中能彻底删除数据吗，怎么做的

面委会 (全程聊天)：

平时是怎么学习的，爱看哪些博客，怎么看待加班，有没有成为leader的潜力

网易考拉 (offer)

一面：

- 1、sql题：学生成绩表，把每科最高分前三名统计出来
- 2、算法题：二维数组中的查找
- 3、kafka如何保证高吞吐的，不了解kafka零拷贝，具体怎么做的
- 4、sql有几种join，map join了解过没
- 5、hbase中row key该怎么设计
- 6、hdfs文件上传流程，hdfs的容错机制
- 7、怎么解决hive数据倾斜问题

二面 (全程写写写)：

全站热榜

- 1 关于计算机专业网?
- 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、
- 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 2672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 6638次浏览
- #如何一边实习一边 5990次浏览

- 3、写一下mysql binlog的数据格式，怎么进行数据清洗的
- 4、写一个正则表达式进行手机号匹配
- 5、讲一下数据库库层级的划分，每层的作用

美团新到店 (offer)

去了北京美团公司里面面试，一上午面完，第二天通知高分通过

一面（简单的聊了聊，10min）：

- 1、介绍项目，以及滴滴的实习经历
- 2、JVM内存的划分
- 3、垃圾收集算法
- 4、数据建模，星型模型和雪花模型
- 5、数仓层级的划分，怎么对接到mysql拿数据

二面：

- 1、sql题：写一条sql删除订单表中重复的记录
- 2、sql题：一张网页浏览信息表，有两列，一列是网页ip，一列是浏览网页的用户（比如a或者b、c、d直到z），求这些网页被a和b或者a和c或者b和c两两组合访问的次数
- 3、hive数据倾斜产生的原因，怎么解决
- 4、设计学生成绩管理系统，符合第三范式要求，并绘出UML图
- 5、算法题：斐波那契数列
- 6、spark程序的运行流程
- 7、spark streaming从kafka中读数据的两种方式
- 8、讲讲数据库索引，B树和B+树
- 9、Elasticsearch的索引，单field索引和多field的联合索引
- 10、linux查看某文件的大小，vim中怎么替换内容
- 11、海量数据的Count问题（单机），如果把大文件hash成不同的小文件，此时小文件装不下某个key对应的数据，该怎么办
- 12、智力题：8升水，有一个5L的杯子和3L的杯子，怎么得到4升水

三面：

- 1、osi七层模型，三次握手和四次挥手，为什么两次握手不行
- 2、kafka怎么保证高吞吐量，项目中有测过吞吐量吗，相比于其它MQ，为什么会选择kafka，kafka怎么保证exactly once语义
- 3、了解hbase吗，hbase为什么查询速度快
- 4、hive sql怎么转换成底层的MapReduce程序，以及shuffle的过程
- 5、算法题：被围绕的区域，leetcode第130题原题
- 6、智力题：一头母牛每年生一头小母牛，每头小母牛从第四年开始，每年也会生一头小母牛，写个公式求第n年会有多少头牛

小米 (offer)

一面：

- 1、java和python的区别，对面向对象的理解，和面向过程相比有什么区别
- 2、java为什么不能多继承
- 3、讲一下java抽象类和接口
- 4、java中为什么要写非static方法
- 5、volatile和synchronized的区别
- 6、算法题：跳台阶问题
- 7、算法题：树的非递归后序遍历
- 8、设计题：一个停车场有一些大车位和小车位，大车只能停大车位，小车既能停大车位又能停小车位，实现这种场景下的调度系统

二面：

- 1、算法题：输入一个字符串，输出该字符串中字符的所有排列

贝壳 (offer)

一面：

- 1、synchronized的底层实现

#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

#0offer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



4、算法题：二分查找

5、快排的时间复杂度和空间复杂度，最优情况和最差情况分别是多少，是稳定排序吗，快排为什么快

二面：

1、介绍项目，项目中涉及到了一些算法，介绍一下

2、两道算法题：路径问题，leetcode上63题和64题原题

3、写正则表达式匹配电话号码

4、智力题：一张圆桌子，我和面试官轮流往桌子上放硬币（随便放），直到桌子放不下为止，最后一个放硬币的人赢，如果我先放，怎么保证我肯定赢

华为（offer）

一轮玄学面：

面试官是做安卓的，瞧不起大数据，觉得大数据很虚，我跟他bb了一堆。然后问我有没有女朋友，我说以前有，现在分了；问我什么时候谈的，什么时候分的，我说本科谈的，毕业分了；问我为什么要分，此处省略一万字.....问我现在想没想过再谈，我说毕竟转专业过来的，想趁在校期间利用好短暂的时光提升自己的技术水平（其实因为找不到）；然后面试官说以后工作了就不好找咯，我说您说的有道理.....

快手（offer）

一面：

1、jvm类加载机制，类加载器，双亲委派模型

2、java实现多线程的方式

3、spark怎么划分stage，宽窄依赖，各包括哪些***作

4、zookeeper怎么保证原子性，怎么实现分布式锁

5、写个快排，为什么要用三数取中法，好处是什么

二面：

1、sql题：找出单科成绩高于该科平均成绩的同学名单（无论该学生有多少科，只要有一科满足即可）

2、sql题：找出单科成绩高于该科平均成绩的同学名单（该学生所有科都必须满足）

3、算法题：求数组中连续子数组的最大和

4、算法题：使用最小花费爬楼梯，leetcode746题原题

三面：

1、讲一下java IO

2、算法题：输入n个整数，找出其中最大的k个数

3、算法题：给一个整数数组和一个目标值，找出数组中和为目标值的两个数

完美世界（offer）

京东广告部（四面完没了消息）

阿里菜鸟（三面完已回绝）

阿里的面试还是比较重视基础的，应该是bat里面问基础问的最多的

一面：

1、HashMap和HashTable的区别，HashMap怎么解决hash冲突，jdk1.8后对HashMap的改进

2、讲讲ConcurrentHashMap，ConcurrentHashMap怎么保证线程安全，HashTable怎么保证线程安全

3、HashSet的底层实现，是不是线程安全的

4、ArrayList和LinkedList的区别，是不是线程安全的

5、讲讲设计模式，最常用哪种设计模式，单例模式的实现方式

6、进程和线程，Java实现多线程的方式，什么是线程安全，怎么保证多线程线程安全

7、可重入锁的可重入性是什么意思，哪些是可重入锁

8、为什么要用线程池，线程池的好处

9、JVM垃圾处理方法，对象什么时候进入老年代，什么时候进行FullGC

10、Java堆溢出问题怎么处理，内存泄漏和内存溢出的区别

11、智力题：50个红球和50个黑球往两个桶里放，然后自己去抽，怎么样才能使抽到红球的概率最高

二面：

1、讲讲数据库存储引擎

2、介绍一下索引。索引设置的规则。聚簇索引和非聚簇索引的区别。索引的最左前缀原则

- 5、jvm调优做过没，-Xms和-Xmx分别指什么
- 6、算法题：输入两个字符串，输出它们合并排序后的结果
- 三面：
- 1、讲讲数据库的范式
 - 2、Linux进程通信和线程通信
 - 3、线程池的参数
 - 4、什么是内部类，什么是匿名内部类
- 5、设计题：一个市有9个消防站，现在要新增3个消防站，这3个消防站应该放在哪里

另：

欢迎探讨相关技术；欢迎探讨人生规划；欢迎在逆境中昂起头颅、继续阔步向前！

#腾讯##网易##美团##面经##小米##贝壳找房#

👍 100 ☆ 1109 💬 169 ➦ 分享 ... 浏览 5.5w

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(169)

👤 内部员工回复(共1人)

推荐

最新

楼层



快拿掉快拿掉 🟡 飞黄腾达

Java

华为是大厂面试中最搞笑的

👍 5 💬 回复 ➦ 分享 发布于 2019-05-07 09:34



京龙鱼 🟡 飞黄腾达

字节跳动 数据开发工程师

前排膜大佬，亲眼见证了大佬的offer收割历程，怀念与大佬一起实习合租的快乐时光

👍 2 💬 回复 ➦ 分享 发布于 2018-09-14 20:19



优秀的张先生 🟡 大橘已定 楼主 回复

写的头疼，快来抱抱我

👍 1 💬 回复 发布于 2018-09-14 20:21



SharpaySharpay 🟡 小白牛

算法工程师

求发大佬的子雨大数据资料🙏

👍 1 💬 回复 ➦ 分享 发布于 2019-04-16 15:39



放开风筝线 🟡 青出于蓝

算法工程师

终于看到一个跟我一样从水利水电转计算机专业的了😄😄

👍 点赞 💬 回复 ➦ 分享 发布于 2019-05-06 21:48



J&Y 🟡 飞黄腾达

算法工程师

没有大数据项目支持，能找到大数据的工作吗还🤔

👍 点赞 💬 回复 ➦ 分享 发布于 2019-07-28 09:08



FightForLife 🟡 青出于蓝 回复

同没有支撑。。。到网上找了一个日志流处理的项目。。

👍 点赞 💬 回复 发布于 2019-08-11 20:17



求一个offer中 青出于蓝

大数据开发工程师

这是我见过大数据面经里面最有深度的面经

1 回复 分享 发布于 2020-08-30 11:18



弱鸡的程先生 青出于蓝

大数据开发工程师

博哥你告诉我怎么才能成为你这样的大佬

点赞 回复 分享 发布于 2018-09-14 20:28

优秀的张先生 楼主 回复

这个id有点秀

点赞 回复 发布于 2018-09-14 21:37

我家的狗不咬人

运营

优秀!

点赞 回复 分享 发布于 2018-09-14 20:30

优秀的张先生 楼主 回复



点赞 回复 发布于 2018-09-14 20:40

And

Java

博哥你是如何做到如此优秀的

点赞 回复 分享 发布于 2018-09-14 20:34

优秀的张先生 楼主 回复

头像识人，哈哈，加油

点赞 回复 发布于 2018-09-14 23:12

小晨同學_Frank

Java

大佬优秀，沾沾喜气

点赞 回复 分享 发布于 2018-09-14 20:38

优秀的张先生 楼主 回复

老哥你的面经才是强，我还看过你的面经

点赞 回复 发布于 2018-09-14 20:40

小晨同學 Frank 回复 优秀的张先生 楼主

半路出家 还得好好准备才行。

点赞 回复 发布于 2018-09-14 20:42

0offer回家摊煎饼果子

Java

大佬优秀，沾沾喜气

点赞 回复 分享 发布于 2018-09-14 20:38

5234534

Java

沾沾喜气!

点赞 回复 分享 发布于 2018-09-14 20:43



请问楼主，自学大数据开发都要准备哪方面知识，谢谢！

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-14 20:45

优秀的张先生 楼主 回复

学习网站是厦大数据库实验室博客，其实网上能搜到，我感觉对于0基础想要入门大数据的同学来说比较有系统的学习价值，这层的同学都已经私了，应该没有漏的.....

👍 4 💬 回复 发布于 2018-09-15 20:49

Y0001 回复 优秀的张先生 楼主

大佬再请教个问题,面试的时候有问到Spring+SpringMVC+Mybatis这些框架么,应该不用学了吧- -

👍 点赞 💬 回复 发布于 2018-09-15 21:04

—— 展开68条回复 ▾

赛丽亚·克鲁敏

Java

快手offer能转让吗老哥 π_π...

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-14 20:48

优秀的张先生 楼主 回复

还没有快手offer.....

👍 点赞 💬 回复 发布于 2018-09-14 20:51

OneFinger 回复

大佬，能私信一下地址吗

👍 点赞 💬 回复 发布于 2018-09-23 15:59

—— 展开1条回复 ▾

OreoOli

算法工程师

大佬优秀，沾沾大佬的喜气

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-14 20:51

唔系柠檬boy

电子科技大学

厉害了，前排膜

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-14 20:54

Melody201805171233281

武汉大学

大佬大佬，沾沾大佬喜气

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-14 20:54

手淘2022实习生招聘

阿里巴巴 服务端

优秀

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-14 20:56

sslen

安卓

膜拜大佬

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-14 21:18

爱吃蔬菜的埃尔法特

安卓

陈独秀都没你秀