



问WD鼎 飞黄腾达

2018-09-27 09:06 Java

+ 关注

大数据面经好少啊，我来写点

面过阿里（拿到意向书）、网易（无情拒绝）、海康，问题都总结一块了分不出来哪个公司问的了（见谅）。。。

- (1) spark运行流程、源码架构
- (2) Hbase主键设计、hbase为啥比mysql快、为什么项目选用hbase
- (3) Hbase读写流程，数据compact流程
- (4) Hadoop mapreduce流程
- (5) Spark standalone模型、yarn架构模型（画出来架构图）
- (6) Spark算子（map、flatMap、reducebykey和reduce、groupbykey和reducebykey、join、distinct）原理
- (7) Spark stage的切分、task资源分配、任务调度、master计算资源分配
- (8) Sparksql自定义函数、怎么创建dataframe
- (9) Sparkstreaming项目多久一个批次数据
- (10) Kafka复制机制、分区多副本机制
- (11) Hdfs读写流程，数据checkpoint流程
- (12) Sparkshuffle和hadoopshuffle原理、对比
- (13) Hivesql怎么转化为MapReduce任务
- (14) Spark调优
- (15) Spark数据倾斜解决方案
- (16) Yarn工作流程、组成架构
- (17) Zookeeper首领选取、节点类型、zookeeper实现原理
- (18) hbase的ha，zookeeper在其中的作用
- (19) spark的内存管理机制，spark1.6前后对比分析
- (21) spark rdd、dataframe、dataset区别
- (22) spark里面有哪些参数可以设置，有什么用
- (23) hashpartitioner与rangePartitioner的实现
- (24) spark有哪几种join
- (25) spark jdbc(mysql)读取并发度优化
- (26) Spark join算子可以用什么替代
- (27) HBase region切分后数据是怎么分的
- (28) 项目集群结构（spark和hadoop集群）
- (29) 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（直连和receiver方式）
- (30) kafka如何保证高吞吐的，kafka零拷贝，具体怎么做的
- (31) hdfs的容错机制
- (32) zookeeper怎么保证高可用性，怎么实现分布式锁

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 2023春招实习面（已接阿里offer
- 6 阿里笔试419 451解
- 7 请教如何刷题、i
- 8 Java-2022暑期2023春招部分面
- 9 4月19日 0offer
- 10 京东JD Young实（后端开发）

正在热议

- #2023届毁约公司名 3645次浏览
- #滴！实习打卡# 3062次浏览
- #五一调休情绪发泄# 2531次浏览
- #阿里巴巴信息集散地# 2690次浏览
- #通信硬件2023笔面# 3713次浏览
- #现在前端的就业环境# 17678次浏览
- #如何拒绝/IA# 1078次浏览
- #没有实习经# 9718次浏览
- #要钱还是要# 1398次浏览
- #我的成功项# 7058次浏览
- #你收到了团子# 6639次浏览
- #如何一边实习一边# 5997次浏览

(34) Zookeeper脑裂问题

Scala

(1) 隐式转换

(2) 柯理化

#面经##秋招##阿里巴巴##网易#

👍 13 ☆ 226 💬 32 ➦ 分享 ...

浏览 1w

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(32)

推荐 最新 楼层



牛客小王 🟢 青出于蓝

Java

请问大佬有木有女朋友

👍 2 💬 回复 ➦ 分享 发布于 2018-09-27 09:34



Cecilia_Hu 💜 小白牛

重庆邮电大学

请问大佬喜不喜欢师妹

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-27 10:11



clutchbear 🟡 大橘已定 回复

直系学妹啊

👍 点赞 💬 回复 发布于 2019-04-02 11:37



激光豆 💜 小白牛

Java

楼主是校招还是社招啊？笔试中不会考算法的吗？还有数据结构的内容？

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-30 14:11



bupzyc 🟡 大橘已定

北京邮电大学

懂得这么多。。。

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-27 09:22

问WD鼎 🟢 楼主 回复

有些没回答上来的。。。主要是spark和hadoop

👍 点赞 💬 回复 发布于 2018-09-27 09:23

叶子在这儿

Java

膜拜

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-27 09:26

叶子在这儿

Java

请问大佬是本科生还是

👍 点赞 💬 回复 ➦ 分享 发布于 2018-09-27 09:27

3 ☆ 226 💬 32 ➦ 分享 ⋮

#机械制造2023笔面
3646次浏览

#23届找工作求助阵
122274次浏览

#京东信息集散地#
1631次浏览

#国企/银行/研究所公
12643次浏览

#0offer是寒冬太冷
58850次浏览

#为什么那么多公司
6452次浏览

#工作丧失热情的瞬
8142次浏览

#你觉得今年春招回
33161次浏览



渣硕。。

👍 点赞 💬 回复 发布于 2018-09-27 09:29

叶子在这儿 回复

都到阿里咯，是真的大佬

👍 点赞 💬 回复 发布于 2018-09-27 10:16

—— 展开3条回复 ▾

RussellHoward

Java

我面的几家数据研发岗，这部分占的比重真不大，全是java基础，数据库，感觉我面的是后台开发不是大数据

👍 点赞 💬 回复 🔄 分享 发布于 2018-09-27 09:38

问WD鼎 楼主 回复

网易、海康全程大数据问题，阿里还是问了很多java

👍 点赞 💬 回复 发布于 2018-09-27 09:39

RussellHoward 回复 问WD鼎 楼主

可能我不是科班出身的吧，问我基础比较多，都能答上来一些，但面试官似乎不太认可，搞得我都怀疑自己了，你能进阿里，真厉害，能否加个好友了？

👍 点赞 💬 回复 发布于 2018-09-27 09:57

—— 展开1条回复 ▾

天凉凉地凉凉

Java

大佬能讲讲spark中***的话会把这个rdd存到哪几个executor上吗

👍 点赞 💬 回复 🔄 分享 发布于 2018-09-27 09:59

问WD鼎 楼主 回复

不是大佬，这个不会啊。。。

👍 点赞 💬 回复 发布于 2018-09-27 10:09

牛客6313490号 回复

在哪里计算就会存在哪个executor上啊

👍 点赞 💬 回复 发布于 2018-09-27 11:29

—— 展开2条回复 ▾

Mge.

安卓

大佬

👍 点赞 💬 回复 🔄 分享 发布于 2018-09-27 10:02

🌸娟

前端工程师

708发来贺电

👍 点赞 💬 回复 🔄 分享 发布于 2018-09-27 10:04

nancy1024

Java

不是大数据方向，但是不明觉厉，帮顶👍

👍 点赞 💬 回复 🔄 分享 发布于 2018-09-27 10:12

Cecilia_Hu

重庆邮电大学

偏方三八面体

Java

本人男 需求男 大佬有兴趣没

👍 点赞 🗨️ 回复 🔄 分享 发布于 2018-09-27 10:21

13579

山西大学

我擦，我面大数据，就不问我这些

👍 点赞 🗨️ 回复 🔄 分享 发布于 2018-09-27 10:25

牛客124531号

Java

big old

👍 点赞 🗨️ 回复 🔄 分享 发布于 2018-09-28 09:59

七彩祥云吞天

京东 数据开发

大佬暑假有实习的经历嘛 🤔

👍 点赞 🗨️ 回复 🔄 分享 发布于 2019-04-17 15:26

别问，问就是稳了再拒

算法工程师

大佬，是阿里数据研发工程师吗，进去具体是干嘛呀

👍 点赞 🗨️ 回复 🔄 分享 发布于 2019-04-17 15:29

BallOffer

阿里巴巴 数据

大佬都看的什么书啊？求推荐

👍 点赞 🗨️ 回复 🔄 分享 发布于 2020-03-19 00:20

相关推荐

香农派

04-13 09:58 阿里巴巴_闲鱼技术部_JAVA开发工程师

【闲鱼】2024届实习内推-JAVA (base杭州) [内部员工回复](#)

【团队简介】：闲鱼APP是中国最大的的闲置交易平台，是阿里正在催生的第三个万亿级平台，拥有极大的创新空间。本团队核心工作内容：闲鱼C2C交易核心 & C2X交易服务【内推对象】：国内外院校2024届毕业生（... [查看更多](#)

👍 9 🗨️ 8 ☆ 7 🔄 转发 ...

小染啊

04-11 00:05 贵州商学院 计算机类

还有CPP暑期坑位吗，求推荐

[牛泪] [#c++](#) [#2024找工作](#)

👍 点赞 🗨️ 6 ☆ 收藏 🔄 转发 ...

xJXx嵌入式笔试面试手册

04-15 15:47 华南理工大学 计算机类

2-4小时突击C++(4)

多线程多线程是多任务处理的一种特殊形式，多任务处理允许让电脑同时运行两个或两个以上的程序。一般情况下，两种类型的多任务处理：基于进程和基于线程。有时候在面试过程中，面试官会问你“有没有写过多线程 ... [查看更多](#)

📄 嵌入式软件开发

3 ☆ 226 🗨️ 32 🔄 分享 :

