即席查询：

浅析即席查询　在数据仓库领域有一个概念叫Ad hoc queries，中文一般翻译为“即席查询”。即席查询是指那些用户在使用系统时，根据自己当时的需求定义的查询。即席查询生成的方式很多，最常见的就是使用即席查询工具。一般的数据展现工具都会提供即席查询的功能。通常的方式是，将数据仓库中的维度表和事实表映射到语义层，用户可以通过语义层选择表，建立表间的关联，最终生成SQL语句。即席查询与通常查询从SQL语句上来说，并没有本质的差别。它们之间的差别在于，通常的查询在系统设计和实施时是已知的，所有我们可以在系统实施时通过建立索引、分区等技术来优化这些查询，使这些查询的效率很高。而即席查询是用户在使用时临时生产的，系统无法预先优化这些查询，所以即席查询也是评估数据仓库的一个重要指标。即席查询的位置通常是在关系型的数据仓库中，即在EDW或者ROLAP中。多维数据库有自己的存储方式，对即席查询和通常查询没有区别。在一个数据仓库系统中，即席查询使用的越多，对数据仓库的要求就越高，对数据模型的对称性的要求也越高。对称性的数据模型对所有的查询都是相同的，这也是维度建模的一个优点。

Kylin and Druid, Presto

数据仓库（Data Warehouse）是一个面向主题的（Subject Oriented）、集成的（Integrate）、相对稳定的（Non-Volatile）、反映历史变化（Time Variant）的数据集合，用于支持管理和决策。OLAP: On-Line Analytical Processing 使分析人员、管理人员能够从多种角度对从原始数据中转化出来的、能够真正为用户所理解的、并真实反映数据维特性的信息，进行快速、一致、交互地访问，从而获得对数据的更深入了解的一类软件技术。 (OLAP委员会的定义)。Data Mining是通过数学模型发现隐藏的、潜在的规律，以辅助决策。

短文本分类

Word2Vec :搜索引擎查询，APP聚类

阿里比赛项目：

关键点：

数据拉取上，队列缓冲，控制数据流速度，队列支持多线程访问

Spout节点的反压处理和恢复正常

Bolt结点的统计逻辑，发送统计值的时机，不统计直接发和统计一部分再发

结果写入上，定时写入和一直写入记录变化

今日头条面试总结：

一面：

Storm：Storm结构，ACK机制，事务ACK机制，Storm原理，拓扑提交执行过程，Storm源码结构

Redis：Redis数据结构，数据过期处理策略，主从备份原理，故障恢复机制

Hadoop：HDFS文件系统的结构，写文件的流程，备份方式，几个备份，如何备份

京东实习：工作流构建原理，项目组成，实现方式

JAVA：Java内存分布，GC原理（见深入理解Java虚拟机）

阿里比赛：系统结构，处理流程，最终效果，数据库索引有哪些，B树索引，哈希索引，聚集索引

做题：实现LRU，还有一道使用全排列思想

二面：

Storm：结构原理，使用情况，熟练度

京东实习：任务调度系统的组成部分和流程，任务调度策略实现（面试官提出了他觉得更好的策略，但是被我找到漏洞反驳回去了，然后就呵呵了）

阿里比赛：非常详细，每一步的实现，试过哪些策略，如何提高性能，调过哪些参数，缓冲队列原理，HashMap和concurrentHashMap原理和区别

做题：实现一个监控文件改变的程序，把每次文件添加的内容打印在屏幕上，要求程序可执行bugfree

其他：除了Java，还会什么其他语言，Linux问了一个命令，文件处理的，没用过也没听过，忘记了

问博客内容，平时兴趣爱好，问比赛含金量

三面：

京东实习：实习内容，完成了什么，对任务调度系统做了什么改造，为什么要改造，改造完有哪些好处，改造前有哪些缺陷，

阿里比赛：系统内容，结构，比赛的关键点是什么，比赛难点在哪，遇到什么难题，如何解决，独立解决还是与人商量，如何沟通

JAVA：Java多线程通信机制有哪些，锁有几种，平时用过哪些锁，写一个生产者消费者的实例（我选择偷懒拿项目里的实例来说的），有没有用过Java nio，怎么用的，了解到什么程度，有没有用过atmoic编程，了解到什么程度，对Java8新特性了解多少用过什么

其他：除了Java，还会什么其他语言，熟悉到什么程度

四面（hr）：问背景，问性格，问上一份实习离职原因，问实验室

JDK8新特性，

Java并发：cas和aqs

数据库，网络， 框架