# 常见面试题

## 一、Scala相关

1. var和val的区别
2. scala中的class，object，trait有何区别？与java中的class，interface有何异同？
3. [scala和java都是运行在jvm上的语言，它们有哪些异同？](http://wiki.nonobank.com/bin/view/A.%E6%8A%80%E6%9C%AF%E4%B8%AD%E5%BF%83/%E3%80%90%E9%9D%A2%E8%AF%95%E9%A2%98%E3%80%91scala%E5%92%8Cjava%E9%83%BD%E6%98%AF%E8%BF%90%E8%A1%8C%E5%9C%A8jvm%E4%B8%8A%E7%9A%84%E8%AF%AD%E8%A8%80%EF%BC%8C%E5%AE%83%E4%BB%AC%E6%9C%89%E5%93%AA%E4%BA%9B%E5%BC%82%E5%90%8C%EF%BC%9F)
4. [Case Class与普通class有何异同？](http://www.jianshu.com/p/deb8ca125f6c)
5. 伴生对象与普通对象有何异同？
6. [方法(method)和函数(function)有何异同？](http://blog.csdn.net/u012302488/article/details/50543661)
7. [scala中下划线(\_)有多少中应用场景？举例说明](http://ananthakumaran.in/2010/03/29/scala-underscore-magic.html)
8. [对scala中lazy关键字的理解和使用](http://www.jianshu.com/p/61b2546ebc54)

## 二、Spark相关

1. Spark生态圈介绍(spark streaming/sql/mllib/xgrah)
2. Spark Driver和Executor的功能是什么
3. Spark master 和 worker的主要工作是什么
4. [Spark有几种部署模式(client, cluster)，每种模式特点](http://www.jianshu.com/p/65a3476757a5)
5. [Spark应用中application, job,stage,taskset,task之间是什么关系？划分job/stage/task的依据是什么?](http://blog.csdn.net/shadow_mi/article/details/51821613)
6. Spark中有哪些常见的算子？transformation型和action型算子有什么区别？如何理解“惰性计算”？
7. [cache和persist的区别？](http://blog.csdn.net/shen_jz2012/article/details/48417037)
8. 谈谈对RDD的理解？
9. 累加器有什么作用？有哪些应用场景？
10. Spark广播的原理
11. [Spark应用程序的执行过程是什么](http://www.cnblogs.com/bonelee/p/6039386.html)

12. 介绍一下你对Unified Memory Management内存管理模型的理解

13. Spark中资源调度算法有哪几种？分别简述下FIFO和FAIR的区别及使用场景？

## 三、Kafka相关

1. kafka集群有哪些组件(producer/borker/consumer/consumer group)及各自的作用？
2. kafka是否保证消息有序？
3. kafka如何保证消息不被重复消费？
4. 什么是ISR？如何通过ISR保证partition的高可用性？
5. 阐述consumer与partition的消费对应关系？（分析consumer增加／减少，partition增加／减少等场景下的消费再均衡）。

## 四、HBase相关

1. hbase集群各组件的作用（master/region server)？
2. 列式存储和行式存储有何不同？hbase相对于传统关系型数据库有何优劣？
3. 介绍get/put/delete数据的流程。delete是否物理删除记录？delete的记录能否再次插入？
4. minor compaction和major compaction的区别？介绍两种compaction的触发条件。
5. 常用filter介绍(RowFilter/FamilyFilter/ColumnFilter/TimestampFilter/ValueFilter等)？filter与get/scan有啥区别？
6. coprocessor有几种？endpoint和observer有啥区别（各自的应用场景举例）？如何部署coprocessor？

## 五、数据结构与算法

1. 分页计算公式，如果有N条记录，每页M条记录，总页数是多少？ PageCnt ＝ （N-1）／M + 1
2. 有两个变量a, b，如何在不定义新变量的前提下将这两个变量的值调换，比如a=2, b=5, 如何转换为a＝5，b＝2

## 六、Canal相关

1. Canal Server和Canal Client各自的作用？

2. Canal HA架构如何部署？

3. 要使用Canal,MySQL端要配合做哪些修改？

## 七、Storm相关

1. Storm任务执行过程？

2. Strom并行度由哪些参数确定？

3. Storm ack机制

4. 如何提高storm处理能力？

5．Storm与Spark streaming的比较？

## 八、综合题

1、如何设计一个实时监控系统（分别面向系统监控、应用监控和业务监控）？

2、如何搭建一个实时大数据平台（采集、存储、计算等技术选型）？

3、当前实时计算面临的挑战和难题？

4、对比spark streaming、storm和flink的相同及不同点？