面试笔试题

系统抖动：

在请求分页存储管理中，从主存(DRAM)中刚刚换出(Swap Out)某一页面后(换出到Disk)，根据请求马上又换入(Swap In)该页，这种反复换出换入的现象，称为系统颠簸，也叫系统抖动。产生该现象的主要原因是置换算法选择不当。即对刚被替换出去的页，立即又要被访问。需要将它调入，因无空闲内存又要替换另一页，而后者又是即将被访问的页，于是造成了系统需花费大量的时间忙于进行这种频繁的页面交换，致使系统的实际效率很低，严重导致系统瘫痪，这种现象称为抖动现象。解决方案运用局部性原理优化置换算法。

磁盘的读写单位：

扇区是磁盘存储信息的最小物理单位，大小是硬件生产商固定的。

操作系统操作的基本单位是磁盘块，大小由操作系统决定，是虚拟的单位

ICMP协议属于网络层 IP协议的一种

多播IP用的是哪类地址：多播地址(multicast address)即组播地址，是一组主机的标示符，它已经加入到一个多播组中。在以太网中，多播地址是一个48位的标示符，命名了一组应该在这个网络中应用接收到一个分组的站点。

http协议的八种请求类型

10.170.31.1/25的同网段的地址：前25位不变 所以最大的是10.170.31.127 第25位位0 要是128 第25位是1

容器事务类型？Java事务类型：JDBC事务、容器事务、JTA事务

UDP要管对方的状态吗：无连接 不可靠 可靠性由上层应用层保证

数据倾斜，如何解决

Hdfs架构的基本原理，各个模块的基本作用

Wordcount

Rdd:

conf = new SparkConf()

sc = new SparkContext(conf)

line = sc.textFile(“”)

line.flatMap(\_.split(" ")).map((\_, 1)).reduceByKey(\_+\_).collect().foreach(println)

sc.stop()

dataframe:

data = sparkSession.read.text("src/main/resources/data.txt").as[String]

words = data.flatMap(value => value.split("\\s+"))

groupedWords = words.groupByKey(\_.toLowerCase)

counts = groupedWords.count()

counts.show()

sparksql:

from pyspark.sql import SparkSession

spark= SparkSession.builder()

.appName("WordCount")

.master("local")

.getOrCreate()

Data=spark.read.load(‘ ’),format(‘text’)

data.createOrReplaceTempView(‘data’)

result=spark.sql(“select word ,count(\*) from data group by word”)

result.show()