1. 分布式文件中的一致性是指什么？

* 分布式文件的一致性，是指在分布式文件系统中，由于文件在多个节点上具有多个备份，要求每个节点上同一个文件的数据必须是一致的，即数据一致。

1. 试分析讨论客户端和服务器端的强一致性、弱一致性、最终一致性？

* 强一致性，即在多个节点的同步达成一致性之前，该系统是不可用的，即不一致的情况下不可用。
* 弱一致性，在多个节点未达成一致性之前，仍可提供服务。
* 最终一致性，即BASE理论中，我们不要求实现完全的强一致性，但是，可以允许多节点经过一定时间之后达到一致性。

1. GFS采用的放松一致性是指什么？举例说明

* GFS的放松一致性是：GFS中，由于兼顾性能与降低实现的复杂程度，采用了更加宽松的一致性模型，主要如下：

**一致状态**：如果所有客户端无论从那个副本读取，读到的数据都一样，则认为文件region状态是一致的。 即可能数据还在同步之中，并不是最新的数据，但是，客户端从任何一个分布式节点上执行数据读取操作，读到的数据都是一样的。

1. 试分析比较HDFS和GFS的相同点和不同点。

不同点：HDFS只允许串行写入，而GFS允许并非写入；

GC上面，GFS采用惰性回收，而HDFS则是饥饿回收；

相同点：HDFS基于GFS模型实现，底层都是块存储；

1. 试分析MapReduce的过程，举例说明。

Mapreduce实际就是一个分布式计算框架，具体的过程可以分为三个：map、shuffle、reduce

Map即对数据进行映射，比如原本数据为{ k1:v1 ,k2:v2 ,k1:v3 ,k3:v4 }

那么map阶段的工作就是讲k1、k2、k3进行映射操作。

Shuffle阶段，即是混洗，因为mapreduce框架是基于磁盘的，所以，在内存中保留shuffle buffer，用于将map之后的数据进行分区的映射，并且执行落盘，分到磁盘上的不同文件中（多节点系统会冲刷到多节点的磁盘上）。同时进行排序。

Reduce就是聚合，即是用最小堆执行多路归并，提高效率。