

2022年分布式系统考试回忆版

考试时间：2023年2月24日

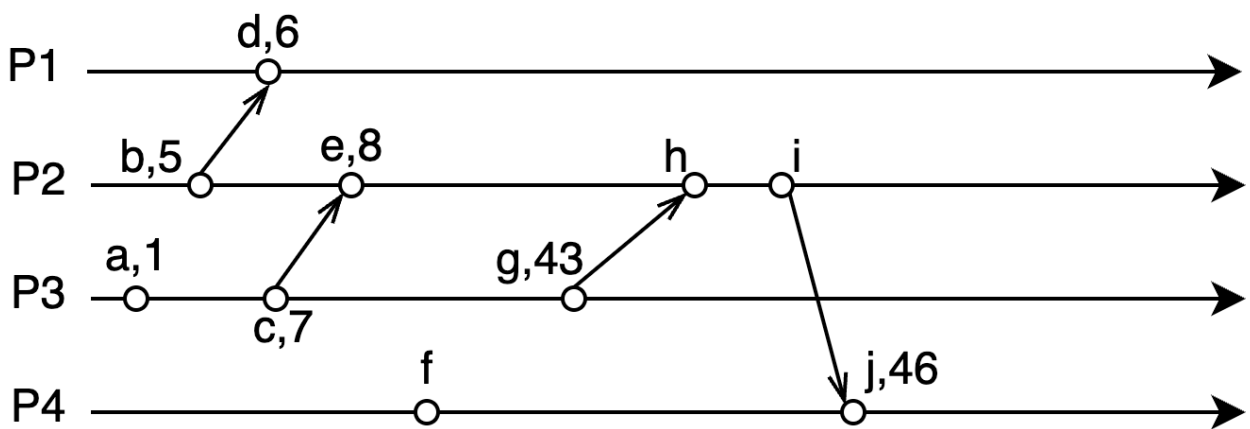
1. 分布式系统的定义，为什么要使用分布式（写出3点原因即可）？
2. 分布式系统的类型，请给每种类型举个例子。
3. TCP和UDP分别位于网络堆栈的哪一层？使用TCP和UDP进行RPC通信分别有什么好处？
4. 分布式系统的三层客户端-服务器架构是什么？
5. 判断题，并给出理由
 - 1) 文本消息程序应该使用UDP进行传输；
 - 2) 有了RPC，进程就可以使用自己本地的内存地址执行程序了；
 - 3) 集中式互斥模式是安全的，但是不公平的。
6. Quorum-based协议，有10个节点，要保证读写不冲突、写写不冲突，请给出可能的写节点数量和对应的读节点数量的范围（5种可能的组合）。
7. RPC，要求体现at-least-once语义，但是限制了提交的次数。
 - 1) 为什么没有体现at-least-once语义？如何修改？

```
1 void submit(char[] data){
2     ...
3     client.send(data);
4     ackMsg = receive(...);
5     ...
6 }
```

2)修改后会发现限制的提交次数很快就用完了，那么要使发送的包中包含哪些信息，可以告知服务器丢弃重复的子任务？

8. 一客户端使用Cristian's算法向服务器获取时钟，它在11:45:14.740时刻向服务器发起获取时钟请求，在11:45:15.810时刻收到响应，里面写着真实时钟是11:45:14.970，客户端估计从服务器到的单趟延迟至少为0.400s，那么客户端收到消息的最早真实时刻为,最晚真实时刻为。

9. lamport时钟，如图所示（记不太清了，但大意如此）。



1) b和d是什么关系，哪个是时间发生更早？

2) 某同学觉得这种方式太麻烦，想到可以在发送消息的时候携带上进程中的日志事件信息和本地时间戳，如P3在c向P2发送消息b的时候附上a发生的时刻，这样P2就可以还原出P2和P3之间所有事件发生的顺序，请问这种方式能区分出a和b哪个先发生嘛？如可以，写出到e事件发生前到所有事件发生顺序？如不能，请说明理由。

3) 你可能已经注意到h没有写上时间戳，那么h的时间戳可能是多少。

0. 使用2PC进行分布式事务T提交， N_0 为协调者， $\{N_1, N_2, N_3\}$ 为参与者节点，现在已经进行了如下信息的发送：

$T_0: N_0 \text{ to } N_1: \text{phase1-prepare}$

$T_1: N_0$ to N_2 : phase1-prepare

$T_2: N_1$ to N_0 : OK

$T_3: N_0$ to N_3 : phase1-prepare

$T_4: N_2$ to N_0 : OK

- 1) 接下来在 T_5 时刻，应该是谁发送消息了，预期的接收者又是谁？
 - 2) 假设在 T_4 时刻，由于网络原因， N_0 没有在规定时间内收到 N_2 发来的OK响应，根据2PC，接下来会发生什么，并解释一下原因。
 - 3) 现在 N_0 已经收到了所有参与者的OK响应，进入第二阶段，并向所有参与者发送了phase2-commit指令，这时 N_1 宕机了。当 N_1 重新上线的时候，它看到的事务状态应该是怎样的，解释一下原因。
1. 类似2021年第11题，问满不满足因果一致性和顺序一致性
 2. 可靠性分析。25个节点，错误节点会返回随机值
- 1) 各节点返回的正确结果相同，正确节点达成一致，允许的最大容错节点数量为多少，并用1-3句话解释一下？
 - 2) 各节点返回的正确结果不同，收集正确节点的结果再取平均，允许的最大容错节点数量为多少，并用1-3句话解释一下？
3. 代码迁移的强迁移和弱迁移的含义分别是什么？
 4. raft可分为领导人选举、日志复制、安全三个部分，请简述每个部分的过程。
 5. 集群训练大模型面临的挑战有哪些？
 6. 数据中心为何要建立容灾机制？容灾可以采用哪些方式？