《分布式系统》非填空题大纲

**1.** 中间件在分布式系统中扮演着什么角色？

**2.** 什么是开放的分布式系统？开放性带来哪些好处？

**3.** 分布式系统的类型。

**4.** 计算机系统的硬件异构性、软件异构性主要表现在哪几方面？（操作系统异构性）

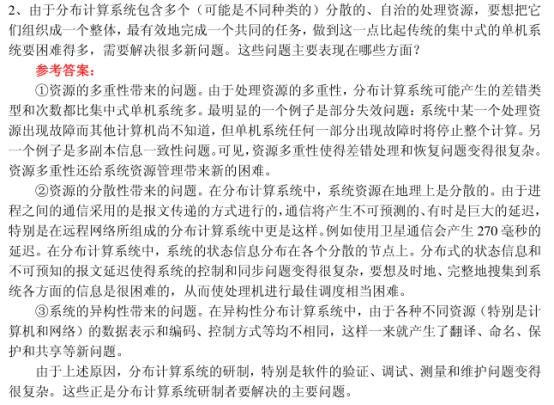
**5.** 对服务器进程中的线程数目进行限制有意义吗？

6. 请描述客户和服务器进程间使用套接字时如何进行无连接通信。

7. 移动计算和普适计算的区别

8. 简述分布式系统中异构性的解决方案

9. 简要描绘全局唯一标识符的一个有效实现

**10.** 

**11.** 分布式系统具有透明性时，系统有什么优点。

**12.** 什么是幂等操作？说说幂等的含义以及其真正的价值。判断集合并、文件记录的插入、缓冲区的读取、**HTTP** **GET**方法、**POST**和**PUT**是否是幂等的。

**13.** **For** **each** **of** **the** **following** **use** **cases,** **identify** **the** **weakest** **RPC** **semantics** **that** **can** **be** **used.** **In** **each** **case** **explain** **your** **answer.**

**14.** 什么是间接通信？简述间接通信的基本策略。

**15.** 发布—订阅系统的几种已定义的常见模式：

**16.** 列出出版订购系统中事件路由的常见技术？

**17.** 用组通信方式时，举例说明消息顺序的重要性，并说明解决方法说明。

**18. 在深度为k的分层定位服务中，当移动实体改变它的位置时，最多需要更新多少条位置记录？**

**19.** **分层定位服务**中的根结点可能是一个潜在的瓶颈。如何能有效地避免这个问题？

**20.** 要使用 **Lamport** 时间戳实现全序多播，是不是每个消息都必须要被严格地确认？

**21.** **IBM** **MQSeries** 以及许多其他消息队列系统中的路由表是人工配置的。描述一种自动完 成配置工作的简单方法。

**22.** 许多分布式算法需要使用协调进程。简单讨论一下，这样的算法实际上可以在什么程 度上被看作为分布式的？

**23.** 有三个进程 **P1**，**P2** 和 **P3** 并发工作。进程 **P1** 需用资源 **S3** 和 **S1**；进程 **P2** 需用资源 **S1** 和 **S2**；进程 **P3** 需用资源 **S2** 和 **S3** 。回答：

**(1)**若对资源分配不加限制，会发生什么情况**?**为什么**?**

**(2)**为保证进程正确工作， 应采用怎样的资源分配策略**?**为什么**?**

**24.** 在分布式系统中，为什么需要代码迁移？代码迁移可以分为 **Sender-initiated** 和 **Receiver-initiated** ，解释其中的含义，并举例说明。

**25.** 代码迁移的动机有哪些？

**26.** 简述三模冗余的基本思想，并举例说明三模冗余能否处理 **Byzabtine** 故障。

**27.** 说明 **RPC** 的主要思想及 **RPC** 调用的主要步骤。

**28.** 在**RPC**调用时，如果服务器或客户机崩溃了，各有哪些解决方法。

**29.** 在**RPC**中，如果客户机在发送请求后在服务器应答消息到来之前崩溃了，将会发生什么问题？如何解决

**30.** 一个影响 **RPC** 执行时间的问题是消息的拷贝问题，试说明在那些环节需要拷贝，并说明减少拷贝次 数的方法。

**31.** **RPC**通信与通常的自定义协议的 **Socket** 通信有哪些区别和联系

**33.**简述远程方法调用(**Remote** **Method** **Invocation，RMI**)的基本通信原理。

**34.** 试举例说明没有统一时钟的分布式系统会发生什么问题？

**35.** 什么是 **RPC**？试简述 **RPC** 的执行步骤。

**36.** 分布式系统中的体系结构样式有那几种？并简述之。

**37.** 什么是 **Quorum** 机制。

**38.** 分布式令牌环算法存在令牌丢失的问题，如果令牌丢失，会导致算法失败，请将该算法改进一下，使该算法既能检测到令牌丢失，也能进行补救。

**39.** 简述分布式系统设计所面临的问题及遇到的挑战。

**40.** 由于分布计算系统包含多个(可能是不同种类的) 分散的、自治的处理资源，要想把 它们组织成一个整体，最有效地完成一个共同的任务，做到这一点比起传统的集中式的单机系统要困难得多， 需要解决很多新问题。这些问题主要表现在哪些方面？

41. 挑战：设计与实现一个对用户来说是透明的且具有容错能力的分布式系统。

**42.** 说明客户**/**服务器模式的主要思想，并说明在采用了阻塞的、有缓存的、可靠的发送和接收原语的情况下，系统是如何工作的。

**43.** 对于接收消息**Receive** 原语，为什么需要缓存**,** 缓存的作用是什么**?**

**44.** 一个最完备的分布式体系由以下模块组成。请说明各模块的功能？



**45.** 集群“脑裂”怎么产生，如何解决。

**46.** 为什么分布式系统需要用到 **ID** 生成系统

**47.** 应用哪些技术可以使得一个分布式系统具有可伸缩性？

**48. 简述利用时间戳预防死锁的不同方法。如果进程 P1、P2、P3分别有时间戳 5、10、15 ，在下列情况下，应该怎样处理？**

**49.** 请介绍一下分布式两阶段提交协议？

**50.** 描述可能发生在因特网上的儿类主要的安全威胁**(**对进程的威胁、对信道的威胁、拒绝 服务**)**可能的发生情况。

**51.** **Describe** **a** **strong** **cache-consistency** **mechanism** **for** **a** **distributed** **file** **system** **with** **a stateless** **server.** **Describe** **a** **weak** **cache-consistency** **mechanism** **for** **the** **same** **system.**

52. 服务器实现概要

**53.** 怎么理解分布式一致性？

**54.** 讨论在 **TCP/IP** 连接之上实现请求一应答协议时可以获得的调用语义， 该调用语义要确 保数据按发送顺序到达，既不丢失也不重复。考虑导致连接中断的所有条件。

**55.** 一个页面从输入 **URL** 到页面加载显示完成， 这个过程中都发生了什么？

**56.** 为什么要用 **ajax**

**57.** 简述 **ajax** 的过程

**58. 客户向服务器发出远程过程调用。客户花5ms 时间计算每个请求的参数，服务器花 10ms 时间处理每个请求。木地操作系统处理每次发送和接收操作的时间是 0.5ms,网络传递 每个请求或者应答消息的时间是 3ms.编码或者解码每个消息花 0.5ms 时间。**

**59.** 为什么使用分布式锁？

**60.** 什么是分布式锁？如何实现之？

**61.** 在 **DHT-based** 系统中可以采用 **finger** **table** 的方法来相应的提高查询效率， 试着说明 其工作原理以及在实际中的应用案例

**62.** 举例简述负载均衡策略

**63. CAP,BASE 以及 ACID 的关系（重点！！）**

**64.** Redis中缓存穿透、缓存击穿、缓存雪崩解决方案？

**65.** 请简述数据库反规范设计方法

**66.** 反规范设计的数据完整性维护方法。

**67.** 简述虚拟化的实现层次。

**68.** **IaaS提供了哪些虚拟化？**

**69.** 虚拟机的两种架构的本质差别表现在那些方面？

**70.** 简述什么是 **SOAP**，以及 **SOAP** 与 **XML** 之间的关系。

**71.** 选举算法中 **Bully** 算法的思想

**72.** 在分布式系统中，许多算法都需要一个进程充当协调者，因此需要协调者选举算法。 试说明霸道算法的主要思想，并说明在**8**个进程的情况下号码为**3**的进程发现协调者崩溃后的选举过程。

**73.** 无线环境下的选举算法

**74.** 描述一下客户和服务器之间使用套接字的无连接通信是如何进行的？

**75.** 试描述客户和服务器之间使用套接字的面向连接的通信是如何进行的？

**76.** 简述 **TCP** 和 **UDP** 协议在通信中的区别

**77.** 分布式系统的类型。

**78.** 移动边缘计算

**79.** 简述面向服务的体系结构(**SOA**)。

**80.** 选举算法中环算法的思想

**81.** 请求驱动式令牌传递方法中，若 **pi** 发出**request** 消息后久未获得**Token**，该怎么处理？ 若引入时钟戳，该算法应做何修改？

**82.** 设计一个能支持分布式无用单元回收和在本地与远程对象引用之间转化的远程对象表。 给出一个例子，包括在不同地址上的几个远程对象和代理，以阐述该表的使用。说明当一个调用导致创建新代理时这个表的变化。说明当一个代理不能用时这个表的变化

**83.** **Jones** 正在运行一组进程 **p1,p2,...** ，**PN**。每个进程 **pi**包含一个变量 **vi**。她希望判定所 有变量 **v1,v2** ，**...** ，**vn**在执行中是否相等。

**84.** 修改用于互斥的中央服务器算法，使之能够处理任何客户〔在任何状态**)**的崩溃故障， 假设服务器是正确的，并且有一个可靠的故障检测器。讨论这个系统是否能够容错。如果拥有令牌的客户被错误地怀疑为出了故障，会发生什么样的情况？

**85. 分布式数据库系统的模式结构（重点！分布透明性）**

**86. 解释分片透明性、复制透明性和位置透明性等三级透明性的区别。**

**87.** 请你给出单点登录的实现方案？

**87-2.补充：分布式哈希表DHT的应用：**

在负载均衡场景中，分布式哈希表通常被用于实现一种称为一致性哈希的算法。一致性哈希将数据映射到一个哈希环上，并将节点也映射到该哈希环上。当需要存储或访问某个数据项时，先将该数据项的键值通过哈希函数计算出在哈希环上的位置，然后顺时针找到最近的节点进行处理。这样可以实现动态添加或删除节点时对数据的均衡分布，从而实现负载均衡。

在分布式锁场景中，分布式哈希表通常被用于实现一种称为基于 Redis 的分布式锁。该锁的实现需要使用 Redis 的分布式缓存系统，将锁的状态存储在分布式哈希表中。当多个客户端需要获取锁时，先通过哈希函数计算出锁所在的节点，然后在该节点上执行获取锁的操作，从而实现分布式锁的互斥性。

**88.** 实现分布式系统同步的复杂性表现在哪几个方面？说明先发生关系， 并说明在 **LAMPORT** 算法中怎样给事件分配时间。

**89.** 有三个进程分别运行在不同的机器上，每个机器都有自己的时钟并以不同且不变的速 率工作(进程**1**的时钟嘀嗒了**6**下时，进程 **2** 的时钟嘀嗒了**8**下，而进程**3**的时钟嘀嗒了**10**下)。举例说明进程之间消息传递中违反先发生关系的情况，并说明如何用 **Lamport** 方法解决。

90. How web service work

91.(**Remote** **Method** **Invocation**，**RMI**)的基本通信原理。

**92.** 服务器虚拟化的关键技术？

**93.** 举例说明 **Lamport** 等人提出的算法是如何解决 **Byzantine** 将军问题的。

**94.** 云计算技术体系结构可以分为哪几层？

**95.** 简述云计算服务的三个层次

**96.** 简述云计算与分布式计算的关系

**97.** 简述私有云、公用云和混合云的基本概念

**98.** 举例描述**\*aaS** 的概念。

**99.** 服务器虚拟化的特性？

**100.** 服务器虚拟化的关键技术？

**101.** 简述管理虚拟集群的四种方式。

**102.** **Multitenant** **technique**

**103.** 简述什么是 **SOAP**，以及 **SOAP** 与 **XML** 之间的关系。

**104.** 在面向消息的通信中，什么是持久通信和暂时通信？试举例说明。

**105.** 讨论下列操作是否是幂等的**:**

**106.** 如何解决幂等性问题

**107.** 举例说明时间触发和事件触发的区别。

**108.** 使用上载**/**下载模式的文件服务器系统与使用远程访问模式的文件系统之间有什么区别？

**109.** 说明保持客户高速缓存一致性的四种算法。

**110. 试分别解释严格一致性、顺序一致性、因果一致性、PRAM 一致性等几种以数据为中 心的一致性模型的含义。下图中的事件序列对上述哪几种一致性模型是有效的？**

**111.** 列出 **URL** 的 **3** 个主要成分，叙述如何表示它们的边界， 并用例子说明每个成分。在什 么程度上 **URL** 是位置透明的**?**

**112.** 简述基于中间件技术的分布式计算系统的组成结构。

**113. 下图给出在两个进程 P1 和 P2 中发生的事件。进程之间的箭头表示消息传递。从初始 状态〔0,0)开始，画出并标注一致状态(p1 的状态， p2 的状态)的网格。**

**114.** 在多计算机系统中的 **256** 个 **CPU** 组成了一个 **16** **X** **16** 的网格方阵。在最坏的情况下， 消息的延迟时间有多长(以跳**(hop)**的形式给出，跳是结点之间的逻辑距离) ？

**115.** 虚拟机的两种架构的本质差别表现在那些方面？

**116.** **SAN** 和 **NAS** 在存储虚拟化方面有哪些不同？

**117.** **GAE** 平台的应用服务架构与传统的 **Web** 应用服务架构有什么异同？

**118. 为了解决大量数据的存储和计算能力不足的问题， 我们有两种选择：**

**119.** 说明分布式系统相对于集中式系统的优点和缺点。 从长远的角度看， 推动分布式系

统发展的主要动力是什么？

**120.** **Browser/Server** 体系结构与**Client/Server** 体系结构相比不仅具有**Client/Server**体系结 构的优点，而且又有 **Client/Server** 体系结构所不具备的独特优势。

**121. 应用哪些技术可以使得一个分布式系统具有可伸缩性？**

**伸缩性：增加请求时依旧保持高性能**

**122.** 给出一个多线程客户端的例子，并给出一种构造多线程服务器的方法。

**123.** 在支持多线程的系统中，可以采用三种模型来组织多线程，详细说明这三种模型。如果在不支持多线程的系统中实现文件服务，如何构造文件服务器呢**?**

**124.** 在代码迁移时，需要迁移代码片断、资源片断和执行片断，说明在迁移资源片断时需 要考虑的主要问题。

**125.** 移动实体、宿主代理和宿主位置的理论

**126.** 什么是有状态服务器和无状态服务器，给出相应的例子，并说明有状态服务器存在的 问题。

**127.** 分布式系统中处理共享文件的四种方法(文件共享的四种语义)。

**128.** 说明无状态服务器和有状态服务器，并说明二者的区别。

**129.** 客户为了发送消息给服务器，它必须知道服务器的地址。试给出服务器进程编址的几 种方法，并说明如何定位进程。

**130. 分布式系统的类型。**

**131.** 客户端**-**服务器模型。

**132.** 论大规模分布式系统缓存设计策略

**133. 解释透明性的含义， 并举例说明不同类型的透明性。**

**134.** 进程和线程的比较。

**135.** 多线程服务器的优点？

**136.** 什么软件代理？举例说明其作用。

**137.** 代码迁移的动机有哪些？

**138. 代码迁移时进程对资源的绑定类型有哪些？**

**139.** 代码迁移时资源对机器的绑定类型有哪些？

**140. 什么是远程过程调用？远程过程调用的步骤。**

**141.** 什么是远程对象调用？

**142.** 消息同步通信与异步通信的区别？

**143.** **DNS** 名称解析的方法有哪两种？各自优缺点

**144.** **Lamport** 时间戳算法的思想

**145.** 当发现一个时钟 **4s** 时，它的读数是 **10:27:54.0(**小时**:**分钟**:**秒**)**。解释为什么在这时不愿 将时钟设成正确的时间， 并**(**用数字表示**)**给出它应该如何调整以便在 **8s** 后变成正确的 时间。

**146.** 全局状态

**147.** 删除无引用实体的方法

**148.** 设计一个分布式垃圾回收算法，简述其原理

**149.** 分布式账本与区块链技术

**150. 选举算法中 Bully 算法的思想**

**151. 选举算法中环算法的思想**

**152. 分布式互斥算法**

**153. 给出Lamport 时间戳互斥算法的主要思想**

**154.** 生产系统每天会产生一个日志文件 **F**，数据量在 **5000W** 行的级别。文件 **F** 保存了两列 数据，一列是来源渠道， 一列是来源渠道上的用户标识。文件 **F** 用来记录当日各渠道 上的所有访问用户， 每访问一次，记录一条。请问如何快速计算出各渠道上新增的用 户？

**155. 什么叫容错性？**

**156.** 掩盖故障的冗余有哪几种？

**157. 拜占庭将军问题**

**158. NFS 的共享预约**

**159. 分布式提交—两阶段提交的思想？**

**160. 分布式提交—三阶段提交的思想？**

**161.** 假如开发一个高吞吐的航空机票订单服务，需要手机端、网页端、以及其他第三方接口调用，为了保证高吞吐，如何用 **RPC+**消息队列方式来实现，说说你的设计思想和实现思路

**162.** 现要为某网上商城实现一个商品价格查询服务，该服务具有以下功能：

**163.** 给定 **a**、**b** 两个文件，各存放 **50** 亿个 **url**，每个 **url** 各占**64** 字节，内存限制是**4G**，让 你找出 **a**、**b** 文件共同的 **url?**

**164.** 使用逻辑时钟的“先于”关系在以下表格中填入校正的时间(如需要)，并标记出所有 的先于关系以及所有的并发事件