**第1章**

**1.** **简述云计算的发展历程？**

云计算的发展历程可由下表格进行概况

|  |  |
| --- | --- |
| 时间 | 相关的技术或者概念 |
| 20世纪60年代 | 虚拟化概念 |
| 1967年 | IBM公司研发出的虚拟机系统CP-40/CMS |
| 1969年 | 美国国防部高级研究计划署研究出计算机网络ARPNET |
| 1970年 | 贝尔实验室研发出Unix系统 |
| 1997年 | Ramnath K. Chellappa首次对云计算做出了学术性的定义 |
| 2002年 | 亚马逊推出了AWS平台 |
| 2006年 | 亚马逊提出了简单存储服务和弹性云计算 |
| 2006年至今 | 云计算的发展逐渐趋于成熟，并由此衍生出了虚拟化、云原生、云安全等多项技术 |

**2.云计算可以看成是一系列技术的集合，请简要叙述云计算中的主要技术，以及这些技术所起的作用。**

①虚拟化技术：虚拟化技术是云计算的核心，它主要是指某些计算机部件在虚拟的而非真实场景下运行。虚拟化技术一方面扩大了硬件的容量，简化了软件的配置过程，另一方面又提高了计算机的计算效率，打破了因硬件性能限制带来的桎梏，虚拟化技术的应用使得单CPU能够模拟多CPU的运行，并且还能在一台计算机上运行多个操作系统，应用程序可以在通过虚拟化技术所划分的空间中独立运行而不受其他因素所影响。

②分布式存储技术：分布式存储技术指的是将数据分散存储至多台独立的设备之中。分布式存储提高了数据存储的安全性、可拓展性以及高效性。

③分布式并行编程模式：分布式并行编程模式的核心思想是指将一个程序分解为若干个相互独立的子程序模块，这些子模块可以分布在多台计算机上同时执行。

④大规模数据管理：大规模数据管理技术的产生一方面使得云计算能够对数据进行高效的处理，另一方面，该技术的应用也保证了云计算能够对数据进行访问以及特定的检索与分析操作。

⑤分布式资源管理：分布式资源管理技术是管理调度数量庞大、分布于不同的地理位置中的机器资源的技术。分布式资源管理技术能够保证在多节点的并发执行环境中，各个节点的状态能够同步，并且在某一节点出现故障时，其他节点能够不受影响继续工作。此外，分布式资源管理还能够解决配置文件的同步更新问题，并为资源的统一管理、分配调度提供便利。

**3.简述云计算的优势。**

①降低成本：云计算的使用一方面使节省了相应的硬件以及机房租赁成本。另一方面又降低了相关技术的使用门槛，降低了运维人员以及网络方面的成本。

②减少相关的运营问题：对于平台中机器维护、数据安全性等问题，云平台的提供商往往会有相关的技术人员负责解决，使用云计算的企业只需处理好与用户相关的业务问题即可。

③促进团队协作：云计算的使用使得信息资源的共享能够在网络上进行，团队中的各个成员无需进行线下的接触即可进行团队协作。

④提升安全性：一方面对于数据存储的服务器，它们大多都处于供应商所提供的防火墙的保护之下，另一方面，数据在被访问时会进行相应的身份认证。

**4.云计算是否完全依赖于物联网，请简述云计算与物联网之间存在怎样的关系。**

云计算和物联网便常常同时出现，二者之间存在着密切的联系。首先对于物联网而言，它更像是一种载体或者平台，物体与物体之间的连接通过物联网这个媒介而实现，并且相关的通信与数据信息交换也在物联网上进行。在云计算的帮助之下，物联网可以更好地实现数据的存储、处理以及计算机资源的分配问题。如果云计算没有产生，物联网的出现可能则必然会推迟一段时间，并且即使出现，物联网的工作效率也会不尽人意，因此可以认为物联网对云计算存在一定的依赖关系。而另一方面对于云计算，它本质上是一种技术，并且在物联网出现之前，云计算的相关技术就已经普遍应用于计算机领域中，可以认为即使是离开了物联网，云计算仍能够持续发展。

**第2章**

**1.简述云计算按技术进行分类时能分为哪些技术。**

①虚拟化技术：虚拟化技术简单来说值得就是将一台物理计算机模拟为多台逻辑上的计算机，并且这些计算机之间在运行上相互独立，互不干扰。虚拟化技术又可以分为全虚拟化、准虚拟化、桌面虚拟以及操作系统虚拟化。

②分布式存储技术：分布式存储作为云计算中最常用的存储技术，通过将不同机器上的存储资源进行整合之后存放各类数据信息。分布式存储技术又可以分为Ceph、Swift以及Lustre。

③数据管理技术：数据管理是利用计算机硬件以及相关的软件技术对数据进行收集、存储、处理和应用的过程。数据管理技术可分为GFS以及HDFS。

**2.简述PaaS的基础结构。**

平台即服务（Platform as a Service，PaaS）实现了硬件和应用软件平台可由第三方提供商来提供和管理，而用户只需负责处理实际的应用和数据。PaaS主要由以下几个模块组成：

①路由模块：该模块的主要功能是将客户端用户的访问请求路由到对应的服务器实例，并提供应用动态注册等功能。

②服务器管理模块：该模块主要为开发人员和管理人员工作，其基本功能包括创建应用实例、配置应用运行参数、发布应用程序、存储空间扩容或缩容等。
 ③应用容器模块：应用容器是PaaS的核心，其主要功能是管理应用实例的生命周期，并定时汇报应用的运行状态等。

④应用部署模块：应用部署模块可以将应用程序打包成为可直接部署的发布包。

⑤块存储模块：该模块主要用于存储应用的发布包，从而保证程序包的长久存储。

⑥数据存储模块：该模块需要保存应用和服务的基本信息，可以基于任何现有的数据库技术实现。

⑦监控模块：该模块的作用是持续监控应用的运行状态。

⑧用户认证模块：该模块用于保证应用程序的安全性。

⑨系统总线模块：用于系统之间通信以及数据交换。

**3.简述SaaS的优势和特点。**

①服务性：SaaS使得软件成为以互联网为载体的服务形式，考虑了服务合约的签订、服务质量的保证、服务费用的收取等问题，并且SaaS在使用过程中，省去了如安装软件、配置软件运行环境等繁琐的步骤，进一步突出了软件的服务属性。

②降低成本：传统软件行业的产品在维护反面占据了一定的成本，同时由于不同客户遍布各地，员工不得不到处奔波，耗费了大量工作时间精力。此外在设备维护以及售后方面，维护人员只需要维护部署SaaS的软硬件环境即可，并且软件服务的售后问题也无需工作人员出差解决。

③简洁性：SaaS的部署十分简单，不需要购买专业的硬件设备。无需再配备IT方面的专业技术人员就能得到最新的技术应用，用户也无需为了使用最新的技术而耗费大量时间进行学习。

**4.简述混合云的工作方式。**

首先混合云会在局域网、广域网、虚拟专用网以及API的帮助下将多台计算机连接在一起，随后会使用虚拟化、容器或软件定义存储等技术提取计算机资源，并将它们存入数据池中，最后会通过一些管理软件将数据池中的资源分配到实际的环境中，并利用身份验证服务按需配置。

当环境连接完成并进行了一定的配置后，一个个独立的云就变成混合云。这种互联的方式是混合云工作的基础，并且一定程度上实现了工作负载的移动、统一管理，以及流程的编排规划。多台计算机之间连接的强度将直接影响着混合云的工作效果。

**第3章**

**1.简述分布式系统的工作方式。**

首先在任务的划分方面，分布式系统采取的分片操作。分片操作的思想很容易理解，也即分而治之，通过将大任务划分一个个小任务，再将这些小任务分配给多台计算机进行处理，最后将处理结果进行汇总。分片操作除了应用在任务的划分上，还体现了数据的存储方面。对于海量的数据，单一计算机节点所能提供的存储空间往往十分有限，因此在分布式系统通常将大体量的数据进行划分，并存储在不同的节点上。此外，出于对安全性方面的考虑，也由于分布式系统中存储空间的富足，对于重要的数据分布式系统会进行冗余或复制集操作，在多个节点上复制存储同一份数据，从而提升系统的容错能力，数据的存放位置还会在不同程度上对计算性能起到提升作用。

**2.简述分布式计算与并行计算之间的区别与联系。**

分布式计算与并行计算在作用、地位上具有一定相似性的同时，它们之间也存在一些差异。一方面分布式计算系统拥有处理并行计算任务的能力，而并行计算系统却并不具备处理分布式任务的能力，另一方面分布式计算的处理范围较之并行计算要更为广泛。

**3.简述分布式存储中的主要技术。**

①元数据管理：常见的元数据管理可以分为集中式和分布式元数据管理架构两大类。集中式元数据管理架构主要面向集中式系统，它采用了单一的元数据服务器，构造方面非常简单。分布式元数据管理架构则将元数据分散存放在多个结点上，进而解决了元数据服务器的性能瓶颈等问题，并提高了元数据管理架构的可扩展性。此外，还有一种无元数据服务器的分布式架构，通过在线算法组织数据，不需要专用的元数据服务器用于存放数据。

②系统弹性扩展技术：目前正处于大数据环境下，数据规模和复杂度的增加往往非常迅速，数据的存储需求也时常发生变化，系统弹性扩展技术满足了变化的存储需求。

③存储层级内的优化技术：存储层级内的优化技术根据实际使用情况对存储系统的构造进行调整，一方面采用多层不同性价比的存储器件组成存储层次结构以降低成本，另一方面通过分析应用特征的方法，识别热点数据并对其进行缓存或预取，通过高效的缓存预取算法和合理的缓存容量配比，以提高访问性能，从降低成本的角度。

④针对应用和负载的存储优化技术：该技术针对应用和负载来优化存储，将数据存储与应用耦合，简化或扩展分布式文件系统的功能，根据特定应用、特定负载、特定的计算模型对文件系统进行定制和深度优化，使应用达到最佳性能。

**4.简述Hadoop、Spark、Flink之间的异同。**

①性能：Hadoop的性能最低，Spark的性能略低于Flink，Flink的性能最高。

②内存管理：Hadoop、Spark都提供了可配置的内存管理，Flink提供自动内存管理。

③容错能力：Hadoop具有高度的容错能力，故发生时障，无需重新启动应用程序，Spark可进行故障恢复，并且无需任何额外的代码或配置，Flink遵循的容错机制基于分布式快照算法。该机制在维持高吞吐率并同时提供强大的一致性保证。

④语言支持：Hadoop支持Java、C、C ++、Ruby、groovy、Perl、Python，Spark支持Java、Scala、Python、R，Flink支持Java、Scala、Python、R。

**第4章**

**1.** **简述一下二层交换机的原理？**

二层交换机通过配置或自动学习生成端口与接入设备的MAC地址之间的端口映射表来实现接入设备与端口之间的绑定，从而避免了多个接入设备之间的单播通讯冲突，提高了通讯效率和可靠性。

**2.** **简述一下NAS和SAN的区别？**

SAN和NAS都是基于网络的外置存储解决方案。SAN可以使用光纤通道连接，拥有较高的传输速度和可靠性，而NAS通常使用标准以太网连接到网络。SAN提供块级别的数据存取，对于客服端表现为磁盘设备。NAS提供文件形式的数据存储服务，对于客服端表现为文件服务器。

**3.** **简述一下VLAN的作用和原理？**

VLAN用于将一个物理LAN在逻辑上划分为多个独立的广播域，从而可以在业务上划分网络并解决单个广播域内设备过多造成广播风暴导致通讯效率下降的问题。VLAN的原理是在链路层数据帧中增加数据所属的VLAN标签，通过VLAN标签确认数据帧所属的VLAN，从而使得广播数据只在相应的VLAN范围内广播。

**第5章**

**1.** **简述CPU全虚拟化、半虚拟化、硬件辅助虚拟化的区别和各自的优缺点？**

全虚拟化是通过软件的方式模拟CPU，在WMM的角度来看，GuestOS在虚拟CPU上执行Ring0指令时会产生异常，从而被WMM捕获异常。为了使GuestOS顺利执行Ring0指令，WMM还需要将GuestOS的指令通过二进制翻译转换为WMM中相同的功能，这个过程称为捕获-模拟。全虚拟化需要WMM频繁的进行捕获-模拟以及二进制翻译，这个过程需要大量的系统开销，因此全虚拟化CPU的运行效率较低。全虚拟化的优点是可以通过软件模拟各种类型的CPU，并且不需要对GuestOS进行修改。

半虚拟化是通过WMM对上层运行的GuestOS提供特定的接口，从而使得GuestOS可以通过特定的接口调用来执行Ring0特权指令。半虚拟化虚拟机执行效率较高，但需要对GuestOS进行修改。

硬件辅助的CPU虚拟化是在CPU中增加非根模式，虚拟机运行在非根模式，GuestOS的特权指令可以运行在非根模式的Ring0上。硬件辅助虚拟化能够让GuestOS以接近物理机的性能来运行，同时不需要对GuestOS进行修改。硬件辅助虚拟化需要CPU在硬件层面的支持。

**2.** **简述一下内存全虚拟化和内存半虚拟化的工作原理？**

内存全虚拟化中WMM为每个虚拟机生成一张影子页表，这张影子页表中维护了虚拟机虚拟地址到机器物理地址间的转换关系，当虚拟机页表修改时，WMM会负责通过影子页表中的映射关系找到虚拟机物理地址和机器地址之间对应，从而实现内存虚拟化。

内存半虚拟化是将虚拟地址和机器地址的映射通过WMM直接写入虚拟机的MMU中。当GuestOS创建了一个新的页表时，会向VMM注册该页表，之后GuestOS对该页表的写操作都会陷入到VMM，VMM会检查页表中的每一项，确保它们只映射了属于该GuestOS的MA，之后VMM会根据自己所维护的映射关系，将页表中的客户机物理地址转换为相应的机器地址，最后再把修改过的页表载入到MMU中，MMU就可以根据修改过的页表直接完成GuestOS虚拟地址到机器地址的转换。由于内存页由WMM而不是GuestOS自身来维护，因此，使用内存半虚拟化时，首先需要修改GuestOS的内存管理方式，禁止GuestOS对自己的页表的修改操作。

**3.** **简述I/O透传技术的主要缺陷？**

I/O透传就是WMM将一个物理设备，如网卡、硬盘、USB接口等直接指定给一个虚拟机使用，一个虚拟机就需要独占一个物理设备。在云计算中，理想的模型是所有的硬件资源都可以被资源池化，因此这种硬件资源和虚拟机一对一绑定的模式很少会使用，同时设备绑定也会在虚拟机迁移时带来困难。

**第6章**

**1.** **介绍VPN的原理以及使用场景**

虚拟专用网络（VPN）在企业网络中有着广泛的应用，集团公司、分布在不同地区的子公司可以互相访问各自局域网资源。出差在外的人员也可通过VPN技术访问公司内部局域网资源，也可跨境访问外部资源。

VPN的原理如下：①通常情况下，VPN网关采取双网卡结构，外网卡使用公网IP接入Internet。②网络一（假设为公网Internet）的终端A访问网络二（假设为公司内网）的终端B，其发出的访问数据包的目标地址为终端B的内部IP地址。③网络一的VPN网关在接收到终端A发出的访问数据包时对其目标地址进行检查，如果目标地址属于网络二的地址，则将该数据包进行封装，封装的方式根据所采用的VPN技术不同而不同，同时VPN网关会构造一个新VPN数据包，并将封装后的原数据包作为VPN数据包的负载，VPN数据包的目标地址为网络二的VPN网关的外部地址。④网络一的VPN网关将VPN数据包发送到Internet，由于VPN数据包的目标地址是网络二的VPN网关的外部地址，所以该数据包将Internet中的路由正确地发送到网络二的VPN网关。⑤网络二的VPN网关对接收到的数据包进行检查，如果发现该数据包是从网络一的VPN网关发出，即可判定该数据包为VPN数据包，并对该数据包进行解包处理。解包的过程中会先将VPN数据包的包头剥离，再将数据包反向处理还原成原始的数据包。⑥网络二的VPN网关将还原后的原始数据包发送至目标终端B，由于原始数据包的目标地址是终端B的IP，所以该数据包能够被正确地发送到终端B。在终端B看来，它收到的数据包能够与从终端A处发送的数据包保存一致。⑦从终端B返回终端A的数据包处理过程和上述过程一样，通过上述方式两个网络内的终端就可以实现相互通讯。

**2. 简述Overlay网络和Underlay网络的关系**

Overlay网络和Underlay网络是一组相对概念，Overlay网络是建立在Underlay网络上的逻辑网络。而Overlay网络是通过网络虚拟化技术，在同一张Underlay网络上构建出的一张或者多张虚拟的逻辑网络。不同的Overlay网络虽然共享Underlay网络中的设备和线路，但是Overlay网络中的业务与Underlay网络中的物理组网和互联技术相互解耦。Overlay网络的多实例化，既可以服务于同一租户的不同业务（如多个部门），也可以服务于不同租户，是SD-WAN以及数据中心等解决方案使用的核心组网技术。

**3.** **简述一下存储虚拟化相对于传统存储方式有哪些优点？**

相比于传统的存储设备使用方式，存储虚拟化模型的核心是在存储设备之上添加一个逻辑层，把多个存储介质比如多个硬盘或者多个磁盘阵列集中起来，组成一个存储资源池并进行统一的管理。存储虚拟化可以实现灵活配置存储资源，提高存储资源利用率，同时还可以实现快照、链接克隆等传统存储方式难以实现的功能。

**4.** **简述一下数据快照ROW方式实现和COW方式实现的区别以及各自适用的场景**

ROW在创建快照后，原磁盘文件会被设置为只读状态，此时新写入的数据会记录到差分卷中。在读取数据时，读取的数据在创建快照后没有更新，则直接读取原磁盘文件。如果读取的数据在创建快照后有变化，则读取差分卷中的数据。ROW适用于创建快照后有大量随机写入的场景。

COW在创建快照后读写操作仍然会在原磁盘文件中进行，在创建快照后第一次写入新的数据时，会先将需要写入的磁盘位置的原数据复制到差分卷中，然后再将新数据写入原磁盘文件，即写时复制。COW适用于创建快照后以读取为主或者写入操作集中在热点区域的系统。

**5.** **简述一下基于主机的存储虚拟化和基于存储设备的存储虚拟化之间的区别和各自的优缺点**

基于主机的存储虚拟化需要在服务器上安装进行存储虚拟化的软件。基于主机的存储虚拟化可以使用异构的存储系统，缺点是需要在服务器主机上运行额外的软件，会消耗一定的服务器计算资源。

基于设备的存储虚拟化中，各台服务器可以直接使用设备上划分的存储资源，不需要在服务器上安装额外的软件，不占用服务器资源。缺点是不同厂商和类型的磁盘阵列设备的虚拟化实现方式可能会不同，不能跨越各设备间的限制。

**第7章**

**1.简述什么是容器**

容器是一种沙盒技术，主要目的是为了将应用运行在其中，与外界隔离；及方便这个沙盒可以被转移到其它宿主机器。本质上，它是一个特殊的进程。通过名称空间（Namespace）、控制组（Control groups）、切根（chroot）技术把资源、文件、设备、状态和配置划分到一个独立的空间。通俗点的理解就是一个装应用软件的箱子，箱子里面有软件运行所需的依赖库和配置。开发人员可以把这个箱子搬到任何机器上，且不影响里面软件的运行。

**2.简述k8s的作用是什么**

kubernetes，简称K8s，是用8代替名字中间的8个字符“ubernete”而成的缩写，是一个开源的，用于管理云平台中多个主机上的容器化的应用。K8s的目标是让部署容器化的应用简单并且高效，k8s提供了应用部署，规划，更新，维护的一种机制。简单来说，k8s就是一个编排容器的工具，一个可以管理应用全生命周期的工具，从创建应用，应用的部署，应用提供服务，扩容缩容应用，应用更新，都非常的方便，而且可以做到故障自愈。

**3.简述什么是桌面虚拟化**

桌面虚拟化是指将计算机的桌面进行虚拟化，以达到桌面使用的安全性和灵活性。可以通过任何设备，在任何地点，任何时间访问在网络上的属于我们个人的桌面系统。桌面虚拟化是指支持企业级实现桌面系统的远程动态访问与数据中心统一托管的技术。一个形象的类比，就是今天，我们可以通过任何设备、在任何地点，任何时间访问在网络上的我们的邮件系统，或者网盘；而未来我们可以通过任何设备，在任何地点，任何时间访问在网络上的属于我们个人的桌面系统。

**4.概述桌面虚拟化和云桌面的区别**

很多人总是认为桌面虚拟化就是云桌面，这对非业内人士来说是没错的。桌面虚拟化技术可以在服务器中虚拟出多个可供用户使用的桌面，这些虚拟出来的桌面就是云桌面，因为这些桌面不是在本地电脑中，而是在服务器中，可以把服务器看成云端，而存在云端的桌面自然就是云桌面。自此桌面虚拟化和云桌面的区别已经得出了结论：桌面虚拟化是一种计算机技术，而云桌面是这种技术衍生出来的结果。

**第8章**

**1.OpenStack是什么？有什么作用？**

OpenStack是一个[开源](https://baike.baidu.com/item/%E5%BC%80%E6%BA%90/20720669?fromModule=lemma_inlink)的云计算管理平台项目，是一系列[软件](https://baike.baidu.com/item/%E8%BD%AF%E4%BB%B6/12053?fromModule=lemma_inlink)开源[项目](https://baike.baidu.com/item/%E9%A1%B9%E7%9B%AE/477803?fromModule=lemma_inlink)的[组合](https://baike.baidu.com/item/%E7%BB%84%E5%90%88/1218690?fromModule=lemma_inlink)，它提供了一个部署云的操作平台或工具集。由美国航空航天局和RackSpace公司合作研发完成的OpenStack是一个灵活的、能够整合多个系统，构建公有云、私有云、混合云的IaaS云平台的组件集合，它通过统一的管理接口对云平台中的资源（如存储、虚拟机、网络等）进行管理，使用OpenStack能够搭建包括公有云、私有云、混合云的IaaS云平台。

OpenStack主要用于公有云和私有云平台的搭建和管理。它为云计算提供存储空间、计算能力等资源，并提供相应的建设和管理工具。可以认为，OpenStack是一个云计算操作系统，是一个由多个组件组合而成的云计算管理平台。

**2.OpenStack有什么优点？**

OpenStack作为一个开放的系统无厂商绑定。可以随时选择新的硬件供应商。

OpenStack具有良好的开源社区氛围。

OpenStack可扩展性强。

OpenStack为用户提供多种使用方式。

OpenStack使用和配置灵活。

OpenStack兼容性强大。

**3.OpenStack的核心服务有哪些？**

用于处理计算的Nova，用于存储的Storing、Cinder和Swift，用于网络通信的Neutron，用于身份认证的Keystone。

**4.OpenStack各组件功能和协作关系？**

Nova负责提供计算，Neutron负责提供网络，Glance负责提供镜像，Cinder提供存储。此外，Horizon为各组件提供可视化管理功能，Ceilometer为各组件提供监控服务，Keystone为各组件提供验证服务，Swift则负责对象存储。

协作关系参考图8-4。

**第9章**

**1.简述云计算分布式应用的开发思路。**

（1）向云计算服务提供商申请大数据处理服务；

（2）向云平台上传数据；

（3）向云计算平台提交作业。

**2.云计算分布式应用开发相对于传统分布式应用开发有哪些优势？**

使用云计算平台进行分布式计算应用开发，省去了传统分布式应用开发步骤中对分布式平台的搭建、测试和维护，降低了开发的时间成本和基础设施成本，这些工作都由云计算服务提供商完成，用户可以专注于分布式应用业务逻辑的开发。此外，使用云计算平台开发也更加灵活，如果业务对数据处理平台的要求较高，开发者可以向服务提供商申请更多的资源，然后在上面部署自己的软件即可。后期如果需求发生变化，开发者可以根据需要随时增加资源或者减少资源。

**3.对于本章所用到的数据，还能提出什么计算需求？**

开放性问题。如求每年的最低温度，计算每年的平均降水量等。