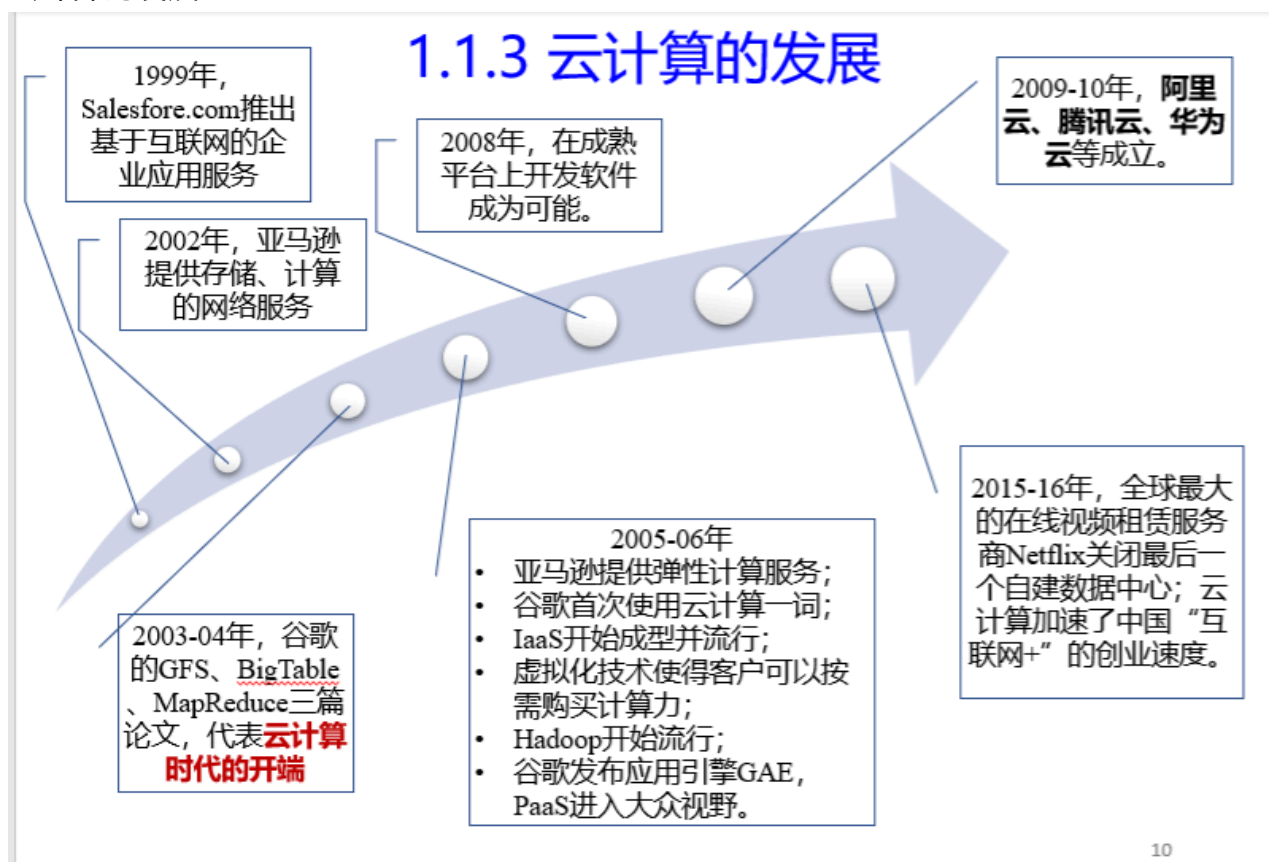


第一章 云计算概述

1.1 云计算的产生与发展

1. 产生背景：企业（工业、商业、中小企业）需要（计算、存储、软件、网络）；个人需要（网盘、宽带）
2. 特点：利用云计算技术，可以**管理、调度、整合、优化**分布在网络上的各种资源，并以统一界面为用户提供各类计算服务。
 1. 按需服务
 2. 通用性
 3. 超大规模计算能力
 4. 实时在线
 5. 高可靠性
 6. 虚拟化
 7. 低成本
 8. 动态调度
 9. 负载均衡
3. 云计算的发展



10

4. 知名云服务提供商：阿里云，腾讯云，华为云，百度智能云
5. **定义**：云计算是分布式计算的一种特殊形式，它引入**效用模型**来远程供给**可扩展和可测量**的资源。
6. **商业驱动力**：

- 1. **容量规划**（过度配置、配置不足），将IT资源容量与其需求之间的差异最小化，以获得预期的**效率和性能**。
 - 1. 领先策略：(Lead Strategy)，根据预期增加IT资源的容量。
 - 2. 滞后策略：(Lag Strategy)，当IT资源达到其最大容量时增加资源容量。
 - 3. 匹配策略：(Match Strategy)，当需求增加时，小幅增加IT资源容量。
 - 4. 挑战性：若按峰值负载配置IT资源，会出现不合理的资金投入；有限的资金投入，会导致因使用限度降低而出现交易损失和使用受限。
- 2. **降低成本**（获得新基础设施、保有其所有权的成本）
 - 1. IT成本与业务性能之间的平衡很难保持。
 - 2. IT环境的扩展总是与对其最大使用需求的评估相对应，这可以让不断增加的投资自动支持新的、扩展的业务。
 - 3. 大部分所需资金都注入到基础设施的扩建中，因为给定的自动化解决方案的使用潜力，总是受限于底层基础设施的处理能力。
- 3. **组织灵活性**（组织对变化响应程度的衡量）
 - 1. IT企业常常需要应对**行业变化**，通常采用的措施是在原来预期或者计划的IT资源规模上进行扩展。
 - 2. **变化的业务需求和优先级**也会要求IT资源具备更高的可用性和可靠性。

7. **技术创新：**

- 1. **集群化**：集群是一组互联的独立IT资源，以整体形式工作。
 - 1. 具有冗余和容错特性（可用性和可靠性提高，故障率降低）
 - 2. **内置冗余和故障转移是云平台的核心概念。**
 - 3. 硬件集群的组件系统由基本相同的硬件和操作系统构成；构成集群的组件设备通过专用的高速通信链路来保持同步。
- 2. **网络计算：**
 - 1. 网络为计算资源提供了一个平台，使其能组织成一个或多个逻辑池。这些逻辑池统一协调为一个高性能分布式网络，有时也称为“超级虚拟计算机”。
 - 2. 网络比集群更加松耦合，更加分散，网络可以包含异构的、且处于不同地理位置的计算资源。
- 3. **虚拟化**（传统虚拟化技术）：是一个技术平台，用于创建IT资源的虚拟实例。

性能	虚拟化前	虚拟化后
资源利用率	每台主机一个操作系统，系统的资源利用率低。	主机与操作系统不一一对应，按需分配使用，系统的资源利用率高。
独立性	软硬件紧密结合，硬件成本高昂且不够灵活。	操作系统和硬件不相互依赖，虚拟机独立于硬件，能在任何硬件上运行。
程序运行效率	同一台主机上同时运行多个程序容易产生冲突，运行效率较低。	管理操作系统和应用程序被封装成单一个体，不同个体间不冲突。同一台机器上运行同一个程序，效率高。
安全性	安全性较差。	强大的安全和故障隔离。

1. 允许物理IT资源提供自身的多个虚拟映像，多个用户可以**共享**它们的底层处理能力。
2. 虚拟化环境中运行的仿真软件可以模拟对硬件的需求，**打断了软硬件之间静态的依赖性**。
3. 云计算的现代虚拟化技术，克服了传统虚拟化平台在性能、可靠性和可扩展性等方面的局限性。
4. 其他： 宽带网络和Internet架构，数据中心技术，（现代）虚拟化技术，Web技术，多租户技术，服务技术。

1.2 云计算的基础知识

1. 概念

1. **用户视角**：云计算是由第三方通过互联网提供的计算服务，用户只需关心云所提供的服务。
 1. 云计算平台提供**硬件服务、软件服务、数据资源服务、可配置的平台服务**。
 2. 用户按需向平台提交自己的**硬件配置、软件安装、数据访问及其他计算需求**，支付一定的**费用**后即可使用。
2. **云计算提供商视角**：解决三个实质问题
 1. **大规模问题**：（1）云计算由数据中心支持，其聚集大量计算资源；（2）云计算也能够支撑大规模的、互联网级别的数据和应用。
 2. **低成本问题**：给客户带来成本上的优势。
 3. **服务运营问题**：通过无差别的存储计算能力提供公共基础服务，服务运营的本质是**按量付费、弹性扩展**，并解决多租户环境的**安全问题**。
3. **平台技术角度**：
 1. **计算**：云计算是由一组内部互联的物理服务器组成的并行和分布式计算系统，该系统能够根据服务提供商和客户之间协商好的服务等级协议**动态提供计算资源**。
 2. **服务**：云计算通过互联网提供弹性的硬件、软件和数据服务，它以社会化服务的形式呈现。
 1. **基础设施即服务/平台即服务(IaaS/PaaS)** 提供商对客户**提供数据中心的硬件、平台和软件服务**。
 2. **软件即服务(SaaS)** 提供商通过互联网以服务的形式给终端用户交付应用软件。
 3. **存储**：云计算将信息**永久**存储在云上的服务器中，客户端**使用时缓存**。
 4. **配置**：云是以**付费使用**的形式向用户提供各种服务的分布式计算系统，其本质是一个对虚拟化的计算和存储资源池进行**动态部署、分配、实时监控**，从而向**用户提供满足要求的计算服务、数据存储服务、平台服务**的系统。

2. 基本术语：

1. **IT资源 (IT Resource)**：指一个与IT相关的**物理或虚拟**的事物。
 1. 基于**软件**：虚拟服务器或定制软件程序等。
 2. 基于**硬件**：物理服务器或网络设备等。

- 2. 企业内部的 (on-premise)：指不基于云的可控的IT环境内部的。
 - 1. 可以**访问**一个基于云的IT资源，并与之交互。
 - 2. 可以被**迁移到云中**，从而成为一个基于云的IT资源。
 - 3. IT资源既可以**冗余部署在内部环境中**，也可以在**云环境中**。
 - 3. 云用户：**使用**基于云的IT资源的一方。
 - 4. 云提供者：**提供**基于云的IT资源的一方。
 - 5. 可扩展性：IT资源可以**处理增加或减少的使用需求**的能力。
 - 1. **水平扩展** (horizontal scaling)：指**分配和释放**IT资源，其中**分配**资源称为**向外扩展**(scaling out)，**释放**资源称为**向内扩展**(scaling in)。
 - 2. **垂直扩展** (vertical scaling)：指现有IT资源被具有**更大或更小**容量的资源代替，分别称为**向上扩展**(scaling up)和**向下扩展**(scaling down)。
- | 维度 | 水平扩展 | 垂直扩展 |
|--------|-----------------|-------------------|
| 扩展方向 | 横向增加节点 | 纵向升级单节点硬件 |
| 可用性 | 天然高可用 (节点冗余) | 存在单点故障风险 |
| 成本曲线 | 按需线性增长 | 呈指数增长 (高端硬件溢价) |
| 实施复杂度 | 需分布式架构改造 | 停机时间短，操作简单 |
| 典型适用场景 | Web应用、微服务、无状态服务 | 数据库、科学计算、GPU密集型任务 |
| 物理极限 | 理论上无限扩展 | 受芯片制程/散热等物理限制 |
- 6. 云服务 (cloud service)：指任何可以**通过云远程访问**的IT资源
 - 1. **并非云中所有**IT资源都可以被远程访问，比如云中的**数据库或物理服务器**只能被本云的其它IT资源访问。
 - 2. 公开发布的API可以专门部署为允许远程客户访问。(API，全称为**应用程序编程接口** (Application Programming Interface)，是一组预定义的规则和协议，允许不同的软件应用程序相互通信和交互。API就像是数字世界中的**万能翻译官**，让不同的软件系统能够无缝对话和协作。)
 - 7. 云服务用户 (cloud service consumer)：是一个**临时的运行时角色**，由访问云服务的软件程序承担。

1.3 目标与收益

- 1. 云IT**投资理念**：**减少或者彻底消除**前期的IT投资，也就是软硬件的采购和拥有成本。(云环境由相当广泛的基础设施组成，提供了“**按使用付费**”模式租赁的IT资源池，即IT资源仅根据实际使用情况付费。与相同的企业内部环境相比，云具备**减少初期投资**，以及**与可测使用情况成正比**的运营成本的能力。)

1. **成比例的成本**(proportional cost): 使用可测的运营支出, 代替预期资本投入。

思考: 云计算中, 100台服务器使用1小时, 与1台服务器使用100小时, 耗费是否相同?



相同



不同

2. 企业从小规模开始, 根据需求相应的增加IT资源配置。
3. 减少前期投资可以使资本用于核心业务投资。

2. 云用户能获得的可测**收益**:

1. 可以**短期**按需访问按使用付费的计算资源, 并在不需要时释放这些资源。
2. 感觉上在需要时可以获得**无限**的计算资源, 因此减少了资源供给的需求。
3. 可以**细粒度的增加或者删除**IT资源, 比如按照 1 G 的幅度增减存储空间。
4. 基础设施抽象化, 应用不会与设备或位置绑定, 可以在需要时**方便的迁移**。

3. 提高的**可扩展性**:

1. 云用户可以根据处理需求的波动和峰值, 自动或手动地扩展其云IT资源。
2. 处理需求减少时, 自动或手动地释放其云IT资源。
3. 提供可灵活扩展的IT资源是云**固有的、天生的特性**, **该特性与成比例的成本收益直接相关**。
4. IT资源总是可以**满足和实现不可预知的用户需求**, 避免在使用需求达到阈值时可能出现的损失。

4. 提高的可用性和可靠性:

1. IT资源的可用性与可靠性**与企业利益直接相关**:

1. 停机限制了IT资源为用户服务的时间, 也限制了其产生收益的能力。
2. 使用高峰期, 没有立即纠错的运行故障, 不仅无法响应用户需求, 还会降低用户信心。

2. 云计算提高可用性和可靠性:

1. 可用性更高的IT资源具有**更长的可访问时间**, 云提供者通常提供**可恢复的IT资源**, 以便能够保证**高水平的可用性**。

2. 具有**更强可靠性**的IT资源能**更好的避免意外情况**，或从中更快恢复。

以下属于云计算收益的是：

- A** 降低成本
- B** 提高安全性
- C** 提高可扩展性
- D** 提高可靠性

1.4 风险与挑战

1. 增加的安全漏洞：

1. 将业务迁移到云中，意味着**云提供者要分担数据安全**的责任。
 1. 远程使用IT资源需要**云用户将信任边界扩展到外部云**。
 2. 建立包含这样信任边界的安全架构，同时又不引入安全漏洞是非常困难的，除非云用户和云提供者碰巧支持相同或兼容的安全架构。
2. **云提供者有访问用户数据的特权**，会造成**重叠信任边界**。
 1. 云用户和云提供者**双方采用的安全控制和策略**决定着数据安全的程度。
 2. 云IT资源通常是**共享的**，基于这一事实，**不同云用户的信任边界可能重叠**。
 3. **重叠信任边界**和不断增加的**数据曝光**，给恶意云用户（人和自动化工具）提供了更多**攻击IT资源、偷窃或破坏企业数据**的机会。
3. 对云提供者而言，提供可以满足两个云服务用户安全需求的安全机制是一项挑战。

2. 降低的运营管理控制：云用户对云资源的管理控制通常**低于**对企业内部IT资源的管理控制，因此**云提供者如何操作云**，以及**云用户之间进行通信所需的外部链接**都可能引入风险。

1. 一个**不可靠的云提供者**可能不会遵守对它的云服务发布的SLA保证，对于使用这些云服务的用户来说，这将**威胁到它们的解决方案的质量**。（SLA是Service-Level Agreement的缩写，意思是服务等级协议。服务等级协议最根本的形式是协议双方（服务提供者和用户）签订的一个合约或协议，这个合约规范了双方的商务关系或部分商务关系。）
2. 云用户和云提供者之间**较长的地理距离**，可能需要**更多的网络跳数**，这导致了**延迟波动和可能的带宽受限**。
3. 云提供者之间**有限的可移植性**：由于云计算行业内**没有建立工业标准**，因此，公有云存在**不同程度的私有化**。当云用户定制的解决方案要依赖于这些私有环境时，在**云提供者之间进行迁移**就成为了挑战。

4. 多地区**法规遵循和法律**问题：

1. 第三方云提供者常常在可负担的或是方便的地理位置建立数据中心，云用户往往意识不到**资源位置**，但对某些组织来说，这可能造成严重的法律问题。比如英国法律规定，英国公民的个人数据只能留在英国境内。
2. 另一个潜在的问题涉及**数据的获得和公开**。有些国家的法律规定，某些类型的数据必须向某些政府机构或者数据主体公开。比如美国的爱国者法案，使得位于美国的欧洲云用户数据，比位于欧盟国家的数据更容易被政府机构访问到。

如下属于云计算特有风险的是：

- ☐ A 疫情引起的IT资源无法人工维护。
- ☒ B 不同云用户信任边界重叠引起的安全风险。
- ☒ C 云用户无法对使用的IT资源完全控制。
- ☒ D 云的私有特征可能会抑制云IT资源的移植。
- ☒ E 数据存储的地理位置引起的法律法规问题。
- ☐ F 地震、台风等引起的数据破坏。