

作业 1: Amazon 云计算平台与微软 Windows Azure 平台技术分析, 描述上述关键技术的基本功能、体系架构和家族体系。

Amazon 云计算平台与微软 Windows Azure 平台技术分析

云计算是指计算机系统资源, 特别是数据存储(云存储)和计算能力的按需提供, 而不需要客户直接主动管理。目前占主导地位的大型云, 通常有来自中心服务器的功能分布在多个地点; 如果与客户的连接比较近, 也可以指定相应的边缘服务器。云计算使企业能够避免或最大限度地降低前期的 IT 基础设施成本: 企业能够更快地启动和运行他们的应用程序, 提高可管理性和减少维护, 使 IT 团队能够更迅速地调整资源, 以满足波动和不可预测的需求, 在某些需求高峰期提供高计算能力。

Amazon Web Services 技术分析

Amazon 云计算平台提供的一项服务是亚马逊弹性计算云 (EC2), 它允许客户通过互联网拥有一个虚拟的计算机集群, 随时可以使用。AWS 版本的虚拟计算机模拟了真实计算机的大部分属性, 包括用于处理的硬件中央处理单元 (CPU) 和图形处理单元 (GPU); 本地/RAM 内存; 硬盘/SSD 存储; 操作系统的选择; 网络; 以及预装的应用软件, 如 Web 服务器、数据库和客户关系管理 (CRM)。

EC2 将基础设施作为服务 (即 IaaS, Infrastructure as a Service), 将硬件设备等基础资源封装成服务供客户使用。如图 1-1 所示, 传统的 IT 系统的管理自下而上包括网络、存储、服务器、虚拟化、操作系统、中间件、内存管理、数据和应用程序九个部分, 而 IaaS 服务的管理从中间件开始, 包括内存管理、数据和应用程序, 以及部分操作系统的内容, 大幅方便了客户的部署及使用。

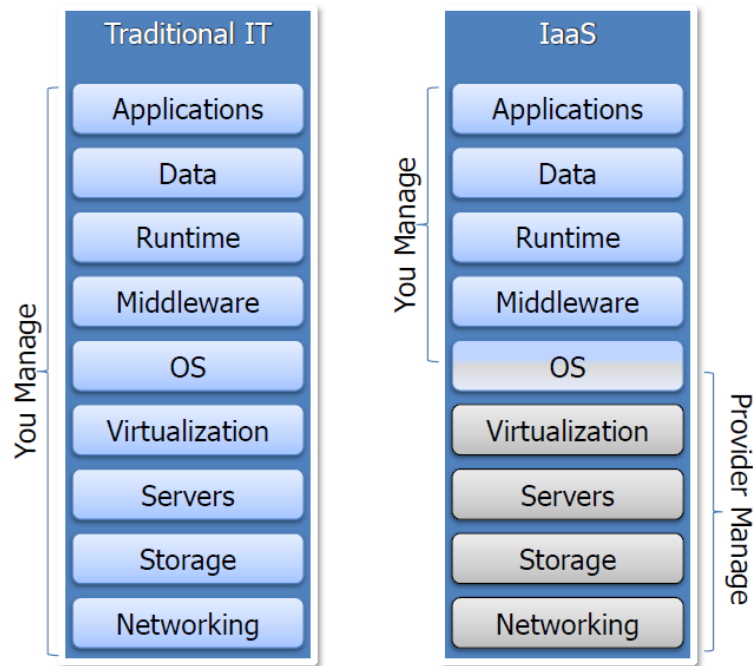


图 1-1 IaaS 概览

EC2 提供了十八种功能服务，包括但不限于以下功能：

1. 裸机实例。为应用程序提供对底层服务器的处理器和内存的直接访问。
2. 使用 Amazon EC2 队列优化计算性能和成本。只需调用一次 API，即可跨 EC2 实例类型、可用区和购买模型预置计算容量，从而帮助优化规模、性能和成本。
3. 暂停和恢复实例。可以休眠由 Amazon EBS 支持的 Amazon EC2 实例，并在稍后从此状态恢复它们。
4. 多个位置。将实例放在多个位置。Amazon EC2 位置由区域和可用区构成。可用区是专用于隔离其他可用区内故障的不同位置，可向相同地区中的其他可用区提供低延迟的廉价网络连接。通过启动独立可用区内的实例，客户可以保护客户的应用程序不受单一位置故障的影响。
5. 高性能计算（HPC）集群。可以实现自定义基础设施提供的同样卓越的计算和网络性能，同时享有 Amazon EC2 提供的可伸缩、灵活和经济实惠的优势……

AWS 提供的另一项重要服务是 S3，一款高性能的云存储服务。S3 提供了多种功能，

可供客户以各种方式组织和管理数据，从而支持特定使用案例、实现成本效率、实施安全性并满足合规要求。S3 提供了七大功能支持：

1. 存储管理和监控。具备平面的未分层结构并有多种管理功能，帮助各种规模和行业的客户按照能够为其业务和团队带来最大价值的方式组织数据。
2. 存储分析和见解。提供对对象存储使用情况和活动趋势的组织范围可见性，并提出了可行的建议，以提高成本效益并应用数据保护最佳实践；存储类分析功能将分析存储访问模式，以帮助客户决定何时将相应数据传输给正确的存储类。
3. 存储类。在多种不同的 S3 存储类中存储数据。
4. 访问管理与安全性。
5. 随时查询。内置的功能和免费服务，可以查询数据，无需复制并将数据加载到单独的分析平台或数据仓库。
6. 数据传输。提供数据传输服务组合，从而为任何数据迁移项目提供适当解决方案。
7. 性能。提供行业领先的云对象存储性能。

EC2 的基本架构如图 1-2 所示。客户可以通过密钥，以 SSH 的方式访问 EC2 实例。在内部，Amazon 机器映像提供若干 EC2 实例，实例本身可以访问自己的存储模块，也可以访问弹性块存储即与 S3 进行数据交换。Internet 访问公有 IP 地址，经过内部的防火墙，抵达对应私有 IP 地址的 EC2 实例服务器，实现服务器访问。

AWS 整体的架构如图 1-3 所示。客户的应用基于包括亚马逊支付平台、亚马逊 SQS 查询系统、亚马逊云前端等在内的系统之中。

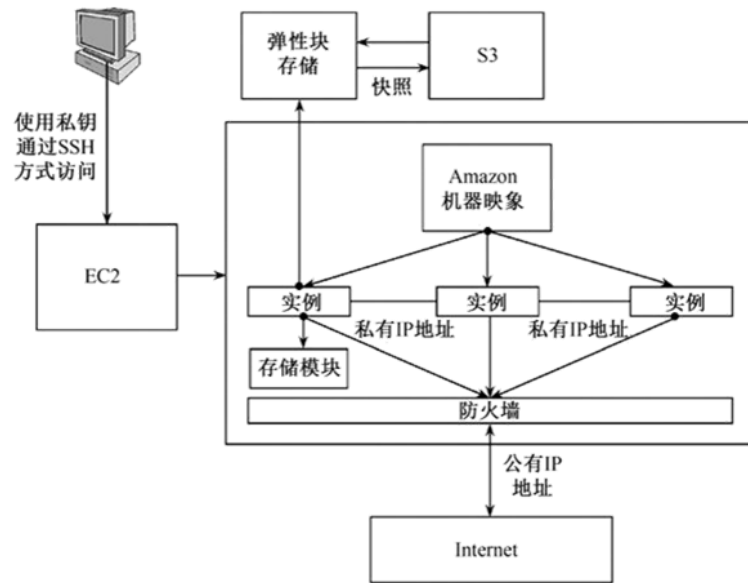


图 1-2 EC2 基本架构

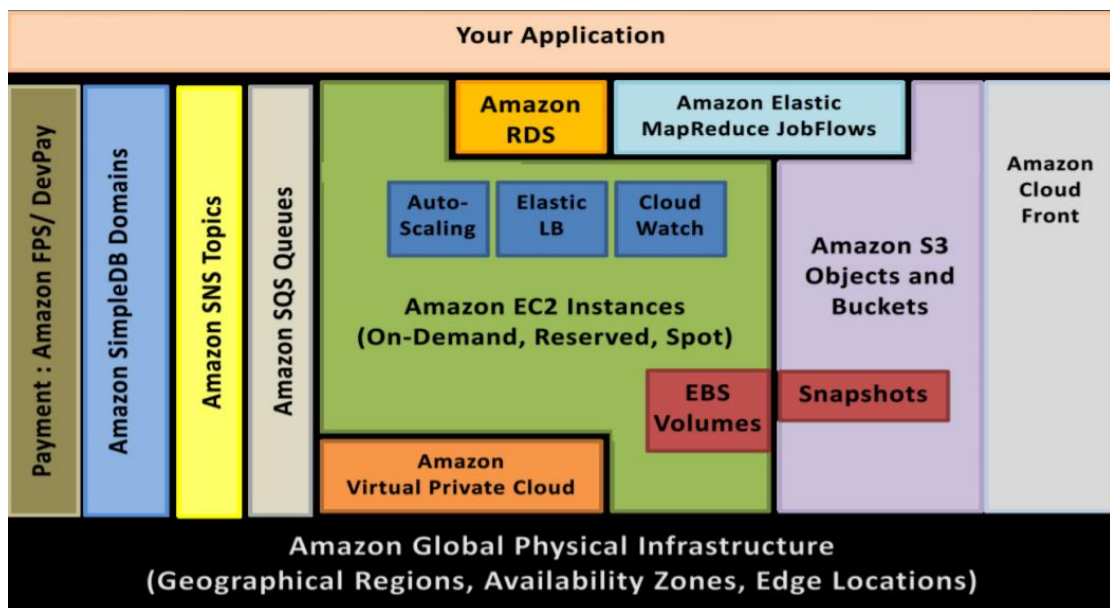


图 1-3 AWS 整体架构设计

Windows Azure 技术分析

Windows Azure，现已更名为 Microsoft Azure，是微软创建的一种基于云计算服务的操作系统，用于通过微软管理的数据中心构建、测试、部署和管理应用程序和服务。它提供软件即服务（SaaS）、平台即服务（PaaS）和基础设施即服务（IaaS），并支持许多不

同的编程语言、工具和框架，包括微软特有的和第三方的软件和系统。

Microsoft Azure 提供了 600 多项服务，包括但不限于：

1. 计算机服务。虚拟机，基础设施即服务 (IaaS)，允许用户启动通用的微软 Windows 和 Linux 虚拟机，以及流行软件包的预配置机器镜像。应用服务，平台即服务 (PaaS) 环境让开发者轻松发布和管理网站。WebJobs，可以部署到 App Service 环境中的应用程序，以实现可以按计划、按需或连续运行的后台处理等。
2. 身份认证。Azure 活动目录用于同步预置目录并实现单点登录。Azure Active Directory B2C 允许在云中消费者身份和访问管理。Azure Active Directory 域服务用于将 Azure 虚拟机加入到没有域控制器的域中。Azure 信息保护可用于保护敏感信息。
3. 移动服务。移动参与收集实时分析，突出用户的行为。它还向移动设备提供推送通知。HockeyApp 可以用来开发、发布和测试移动应用。
4. 存储服务。存储服务提供 REST 和 SDK API，用于存储和访问云上的数据。表服务使程序可以将结构化文本存储在分区实体集合中，这些实体通过分区键和主键进行访问。Azure Table Service 是一个 NoSQL 非关系型数据库。Blob 服务允许程序将非结构化文本和二进制数据存储为可通过 HTTP (S) 路径访问的 blobs。Blob 服务还提供安全机制来控制对数据的访问。队列服务允许程序使用队列通过消息进行异步通信。文件服务允许使用 REST API 或 SMB 协议在云端存储和访问数据。
5. 机器学习。Microsoft Azure 机器学习服务是 Cortana 智能套件的一部分，通过 Cortana 使用自然语言和语音实现预测分析和数据交互。Cognitive Services 是

一套 API、SDK 和服务，可供开发人员使用，使他们的应用更加智能化、更具吸引力和可发现性。这些服务包括人脸识别和验证、名人识别、计算机视觉、视觉特征标记和剪贴画识别……

利用 Azure 数据平台，实现云存储解决方案，以实现大型机和中型机数据现代化为例，其体系架构的设计如图 2-1 所示。成功的 modernizations 使用数据优先策略。利用此方法，组织将重点放在数据上，而不是新系统。数据管理不再只是现代化清单上的一项。构建可持续性的系统结果是，作为面向质量的数据解决方案，可取代零散的、不太受管制的数据库解决方案。具体而言，该计划涉及：对象转换，将源数据存储中的对象定义转换为目标数据存储区中的相应对象；数据引入，连接到源数据存储区并提取数据；数据转换，将提取的数据转换为适当的目标数据存储结构；数据存储，最初和连续将数据从源数据存储加载到目标数据存储。

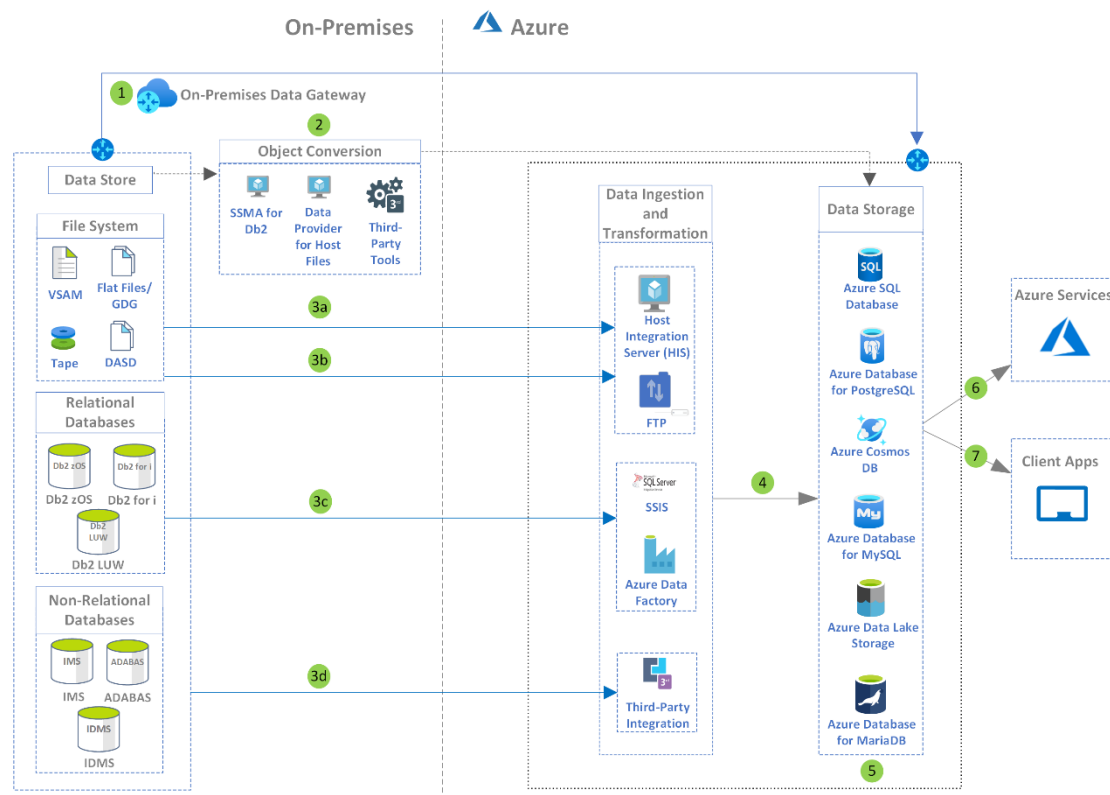


图 2-1 实现大型机和中型机数据现代化体系架构

机器学习是近年来 Azure 重点部署的功能，以深度学习模型的分布式训练为例，其体系架构的设计如图 2-2 所示。在基于同步更新的数据并行分布式训练中，模型跨 n 个硬件设备进行复制。训练样本的小批分割成 n 个微批。每个设备对微批执行向前和向后传递。设备完成该过程后，将与其他设备共享更新。这些值用于计算整个小批的更新权重，权重将在模型之间同步。本体系结构还可用于模型并行和异步更新。在模型并行分布式训练中，模型将分割到 n 个硬件设备，每个设备包含模型的一部分。在最简单的实现中，每个设备可以包含一个网络层，在向前和向后传递期间，信息将在设备之间传递。可以这种方式训练大型神经网络，但代价是性能下降，因为一个设备需要持续等待另一个设备完成向前或向后传递。某些高级技术使用合成渐变，可在一定程度上缓解此问题。

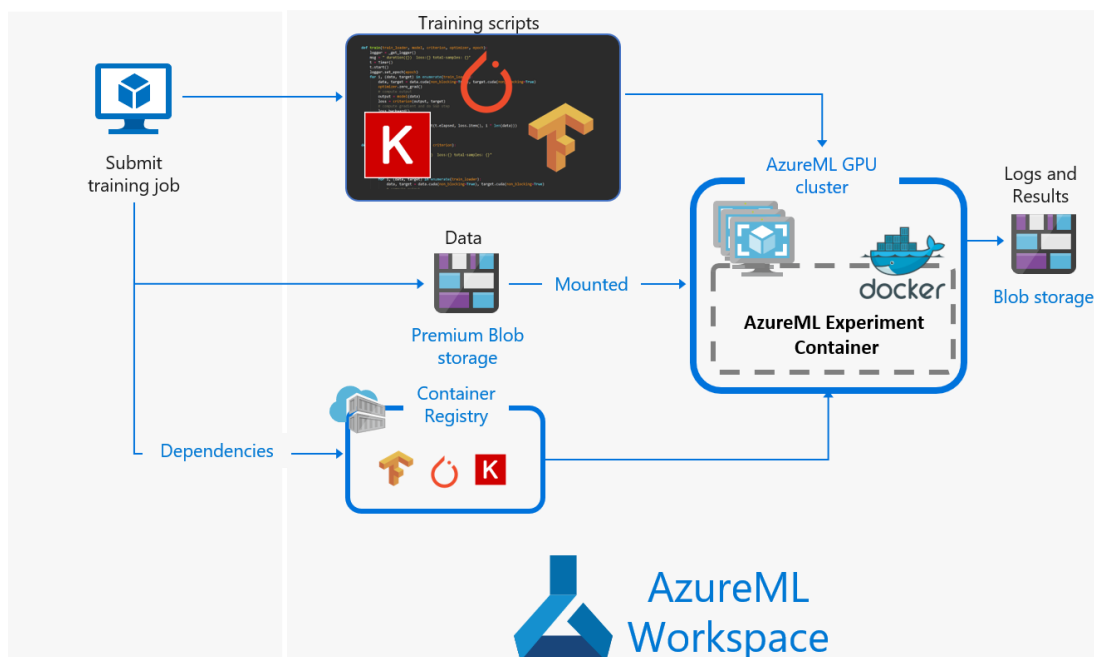


图 2-2 深度学习模型的分布式训练体系架构

作为服务器部署时，Linux 是优先选择的系统，Azure 支持运行 Linux Vm，体系架构的设计如图 2-3 所示。除 VM 本身以外，在 Azure 中预配虚拟机（VM）还需要其他一些组件，包括网络和存储资源。

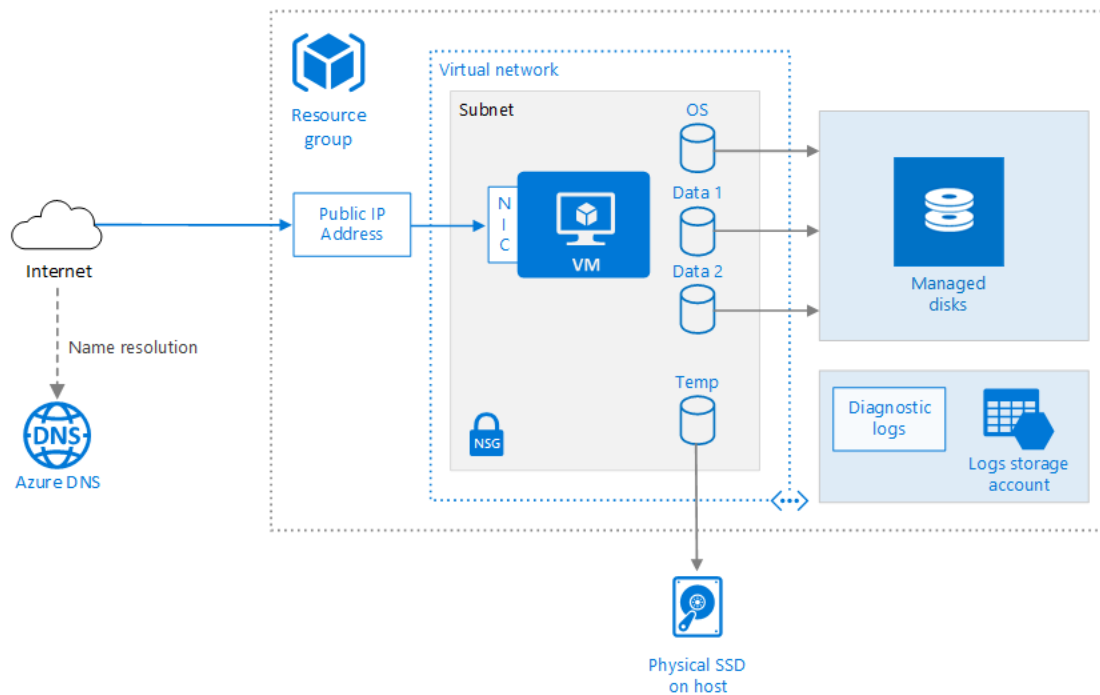


图 2-3 在 Azure 上运行 Linux VM 体系架构设计

Microsoft Azure 具有针对数据中心架构的自我管理机能，可以自动监控划分在数据中心数个不同的分区的所有服务器与存储资源，自动更新补丁，自动运行虚拟机部署与镜像备份等能力，Windows Azure 被安装在数据中心的所有服务器中，并且定时和中控软件：Microsoft Azure Fabric Controller 进行沟通，接收指令以及回传运行状态数据等等，系统管理人员只要通过 Windows Azure Fabric Controller 就能够掌握所有服务器的运行状态，Fabric Controller 本身是融合了很多微软系统管理技术的总成，包含对虚拟机的管理，对作业环境的管理，以及对软件部署的管理等，在 Fabric Controller 中被发挥得淋漓尽致，如此才能够达成通过 Fabric Controller 来管理在数据中心的服务器能力。

Microsoft Azure 环境除了各式各样的虚拟机外，它也为应用程序打造了分散式的巨量存储环境，也就是 Microsoft Azure Storage Services，应用程序可以根据不同的存储需求来选择要使用哪一种或哪几种存储的方式，以保存应用程序的数据，而微软也尽可能的提供应用程序的兼容性工具或接口，以降低应用程序移转到 Windows Azure 上的负担。

Microsoft Azure 不但是开发给外部的云应用程序使用的，它也作为微软许多云服务的

基础平台，像 Microsoft Azure SQL Database 或是 Dynamic CRM Online 这类的在线服务。