Technical note | 000796

Cloud computing

Demystifying the cloud

Author
Brian Alliston

Product Marketing Manager
Chromatography Software,
Thermo Fisher Scientific

Keywords
Cloud computing, Software as a Service, SaaS, Infrastructure as a Service, IaaS, Cloud security, GxP compliance, validation, CSV, Chromatography Data System, CDS, Chromeleon CDS

Goal

To give a basic understanding of the cloud, explain some of the jargon, and give an overview of cloud infrastructure. This technical note will help you learn the difference between SaaS and IaaS, VPC and VPN.

Introduction

Everyone has heard of the cloud and cloud computing; we use them every day in our personal and work lives. Stream videos and music, send and receive emails, or talk to your smart speaker and you use cloud computing. Business software, such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and Human Resources Management (HRM), have become common cloud-based applications.

You have used the cloud, but what is it?

The origin of the term “cloud computing” is shrouded in the mists of time (early 1990s), with many claims and counter claims as to who used it first and in what context. Oxford Languages defines cloud computing as “the practice of using a network of remote servers, hosted on the internet to store, manage, and process data, rather than a local server or a personal computer.”

The term “cloud” is quite appropriate. A cloud in the sky consists of millions of water droplets, whereas the computing cloud is a network of thousands of secure server farms containing millions of servers. The word cloud also gives an ethereal, otherworldly feeling, which is appropriate as cloud processing and storage happens elsewhere.

You may have seen the meme, “There is no cloud, it’s just someone else’s computer.” It’s not quite as simple as that, and it involves lots of acronyms and initialisms.

Virtualization

A key element of cloud computing is hardware virtualization; this is where virtual machines (VMs) are created and run inside software called a hypervisor. Whether you are using a single computer or a large cloud data center, a hypervisor can be used to create individual VMs within physical hardware.

In a physical world, the application and OS run on physical hardware, PCs, or servers, as shown in Figure 1. In the virtual world, software applications run inside a VM, which is a software entity running inside the hypervisor, all of which are running on physical hardware. Hardware resources can be distributed between VMs; for example, RAM, number of CPU cores, and data storage can be divided up depending on the total available and the application's requirements.

As discussed, VMs can be created on a single PC or within on-premises servers, but it is the virtualization of cloud service providers' (CSP) infrastructure that allows them to deliver virtual private servers (VPS) also known as virtual dedicated servers (VDS). Each VPS can run a different OS and multiple applications, or they can be networked to run larger applications.

Virtual servers, either in the cloud or on-premises, can be used when running applications such as Thermo Scientific™ Chromeleon™ Chromatography Data System (CDS), which has been designed to run in physical and virtual environments.

You can configure a VPS to meet the operating requirements for your application. The servers that may be required for Chromeleon CDS (as seen in Figure 2) can be provisioned as multiple VPS, with the required computing capacity and OS. Any other software applications, for example, Microsoft™ SQL Server™ required for the database server, can also be installed.

Of course, these virtual servers must be able to communicate with each other for the system to work, so we need a virtual network.

Private, public, or hybrid clouds

Private clouds are infrastructure that is deployed, maintained, and operated specifically for one company. This may be entirely on-premises or hosted externally.

Public clouds are infrastructure made available to the public on a commercial basis by a CSP. Scale of economies means that public cloud applications can have a significant cost benefit over private cloud systems.

Hybrid clouds combine on-premises and cloud-based infrastructure, allowing data and applications to be shared between them.

CPU – Central processing unit

RAM – Random access memory

I/O – Input/output
(keyboard, mouse, display, etc.)

Disk – Data storage

App – Software application

OS – Operating system

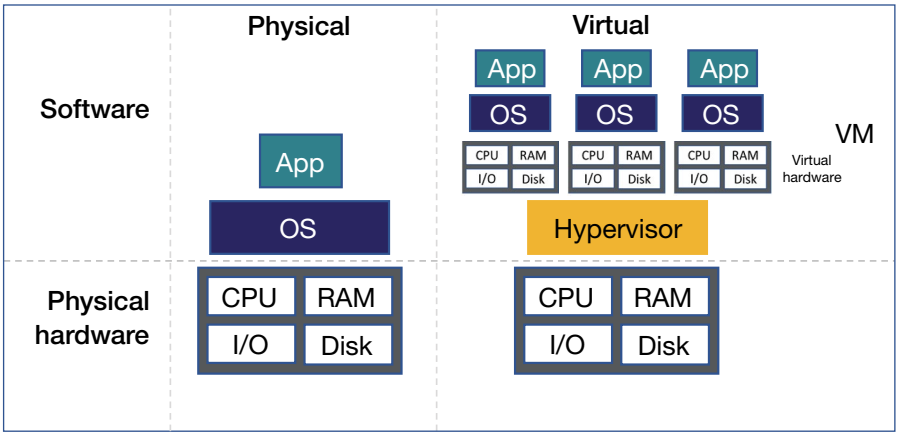

Figure 1. Comparison of physical and virtual computer systems

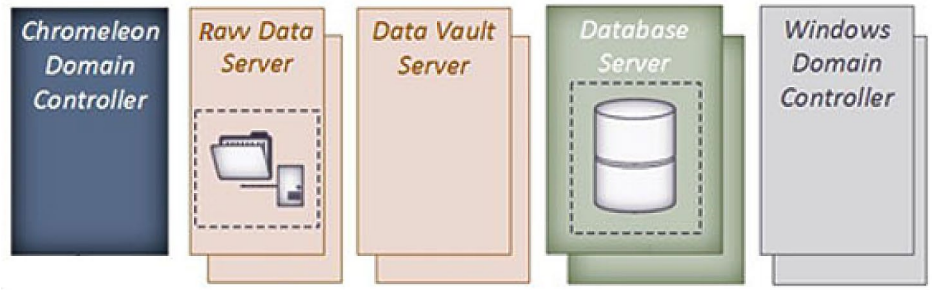

Figure 2. Required Windows and Chromeleon CDS components

Virtual private cloud (VPC)

Where a VPS is just a single VM, a VPC is a network of VPSs specifically provided to an individual organization. VPSs are networked within a VPC, which in turn is inside the public cloud. VPC is the division of the public cloud into secure private networks using technologies such as encryption, private Internet Protocol (IP) addresses and virtual local area networks (VLAN), see Figure 3.

Think of a VPC as a virtual server room, where you are separated from other cloud customers. It provides a secure, configurable virtual environment where you can network virtual servers, control who has access, and configure computing power and data storage. VPCs can be configured to span multiple data centers providing failover should something crash, secure data backups to protect data, and rapid disaster recovery facilitation.

You can connect your VPC to your site via a virtual private gateway, and communications can be kept secure by using a virtual private network (VPN) or a direct connection to the cloud.

Direct connection

Simply put, a direct connection is a dedicated connection from a private, on-premises network into the public cloud, not using the internet. They provide additional security, low latency, and dedicated capacity, but of course they come with additional costs.

Internet gateway

A virtual private gateway is a "guardian" at the end of your internet connection, allowing correctly encrypted data to enter your VPC. This enables you to create a Virtual Private Network (VPN) connection between your on-premises network and the cloud over the internet. The VPN secures your data in transit; it encrypts data and provides a secure "tunnel" where it cannot be accessed by other parties while it is on route.

Connection bandwidth and latency can affect the performance of any cloud-based system. It is worth remembering that while your data is segregated and secure, there are many other internet users and that can slow things down. You can even slow your own systems down if you are sharing internet connections across other business systems.

Bandwidth

Bandwidth is defined as the maximum amount of data that can be transmitted, via the internet, in a given timeframe. It is measured in megabits per second (Mbps). This is not internet speed but capacity to transmit data. Think of your internet connection as a pipe, the wider the pipe the more data can pass through it unobstructed. Try to funnel too much into the pipe and the flow rate is reduced.

Latency

Latency is the "flow rate", the time taken to transmit data from source to destination, and is measured in milliseconds (ms). Latency can be affected by the distance of the connection but also by insufficient bandwidth. If your pipe isn't big enough, then the flow is reduced, meaning your speed from A to B is decreased. Latency can be minimized by locating your application in the closest CSP data center to your site and with careful consideration to the bandwidth your organization requires for all systems operating in the cloud. Thermo Fisher Scientific can provide details of recommended latency and bandwidth for Chromeleon CDS operations.

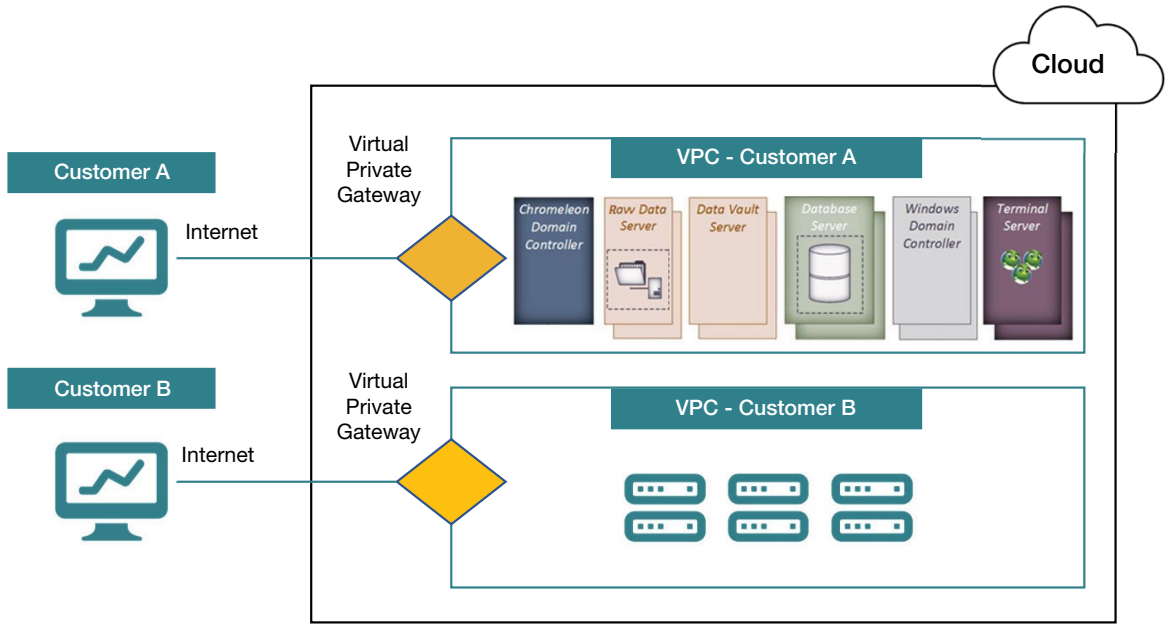

Figure 3. VPC overview

Cloud compute

As we have already discussed, the internet is made up of huge physical server farms located in highly secure warehouses, and this hardware is partitioned into individual VMs. These are the basic building blocks of the cloud. You can provision and configure cloud compute instances, with CSPs providing a variety to choose from. CSPs use the word compute when talking about cloud processing power, memory, networking, storage, and any other resources required for your application; an instance is a VM. Selecting cloud compute instances is just like purchasing a new PC—you decide on the CPU, the amount of RAM, and the size and type of the hard drive. There is quite a selection to choose from, and an example is Amazon Elastic Compute Cloud™ (EC2) available in Amazon Web Services™ (AWS) cloud. These come in five main types: general purpose, compute optimized, memory optimized, high performance/accelerated computing, and storage optimized. Each type also comes in multiple "flavors"; essentially you get to choose the type of processor, number of processors, amount of memory, internal network bandwidth, type of data storage, and size of data storage. CSP websites provide more detail on the types of computing instances available.

Data storage

Basic compute instances come with a temporary instance data store. Think of this as the data storage on your mobile phone. Terminating an instance is like hitting the factory reset—if you haven't saved the data somewhere else, it will be lost. This may be tolerable for some types of data but is unacceptable when operating a CDS in a regulated environment.

Since instance storage is physically attached to the instance, stopping or terminating an instance resets it to its original "factory" state and all data will be lost. You will also lose your data if the instance hard drive fails. Instance storage is only suitable for information that changes frequently, such as data buffers and caches, and should not be used for the long-term storage of valuable data.

Extra data storage volumes can be added to your instances. These additional services are like buying external hard drives for your PC. If you turn off the PC or if the onboard hard drive fails, you still have your data. This is a more permanent solution to storing data. If you terminate the instance, any data in a storage volume is safe.

Both the instance data store and the extra storage volumes are block-level storage, also known as block storage. The hard drive in your PC is a block-level storage device. With block-level storage, data files are broken up into even sized blocks. This type of data storage is primarily used for data requiring frequent updates, such as a database application.

CSPs also provide object storage solutions, for example Amazon S3™ and Azure™ Blob Storage, in which data files are broken into pieces called objects. These are clumps of data and metadata stored in buckets, which are stored in a single repository, not files or folders. Object storage has the advantage of being scalable and can cope with extremely large data volumes, while being cheaper to run than block storage. Object level storage is more suitable for long term storage of large volumes of data that do not need to be repeatedly processed, such as data backups or data archives.

Operation of Chromeleon CDS requires a relational database, such as Microsoft SQL Server or Oracle™ Database, which can easily be installed and run on an appropriate compute instance. Relational databases use tables to store data items and the relationship between data items. This means that data is stored and categorized in a way that can be queried and filtered without the need to reorganize the data. In comparison, flat file databases store data in lists and have no structure for indexing data or the relationship between data.

CSPs also provide managed services where the database is delivered as a Platform as a Service (PaaS), for example, AWS Relational Database Service™ (RDS) and Microsoft Azure SQL Managed Instance.

Cloud delivery models

Cloud services can be delivered in several ways.

- Infrastructure as a Service (IaaS) offers compute, storage, and networking resources on demand, on a pay-as-you-go basis. Basically, you are renting computing infrastructure, where you specify the components, upon which you install and manage your own applications and data.
- Platform as a Service (PaaS). Here, like IaaS, the infrastructure is provided, but includes middleware such as database management services, development tools, etc.
- Software as a Service (SaaS) is a complete solution. You rent the software and hardware infrastructure; the service provider manages both. A service level agreement (SLA) sets out expected availability and security.

Responsibilities are always shared but vary depending on the model and SLA (Figure 4). However, cloud customers always retain responsibility for the data and user access.

Chromeleon CDS has been installed using IaaS with AWS, Azure, and Google Cloud Services, and is currently in operation with several science-based organizations working, for example, in the biopharma, pharma, and oil and gas sectors.

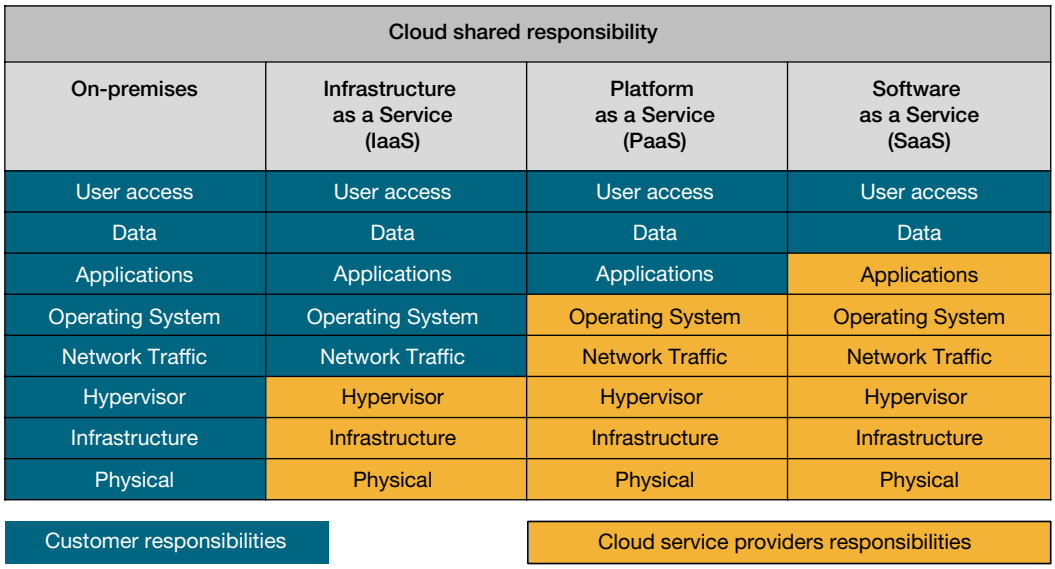

Figure 4. Cloud Shared Responsibility Model

Infrastructure as Code (IaC)

Setting up server infrastructure in the physical world requires physical activity. You must purchase the hardware, wait for it to be delivered, have an IT infrastructure engineer install the hardware into the server room, configure, test, and install applications, etc. Cloud hardware is already set up and ready to go. It just needs to be configured depending upon your requirements. Cloud infrastructure is configured and managed using code instead of manual processes.

Cloud systems can be set up by directly controlling infrastructure using a command-line interface (CLI). For example, a VM can be provisioned and controlled by entering code, line by line, into the CLI. However, for a large complex system, the management of infrastructure could be a laborious task prone to environmental drift, where small errors introduced each time infrastructure is provisioned, add up to a significant deviation from the original design.

IaC is the process of managing and provisioning cloud infrastructure using machine-readable definition files. Using IaC services and tools such as AWS CloudFormation™, Azure Resources Manager™, and Google Cloud Deployment Manager™, infrastructure and applications can be templated. This allows infrastructure version control, and systems can be provisioned in a safe, repeatable manner. Templates can include infrastructure, operating systems, and software applications.

Remote access

With all your data and servers in the cloud, wouldn't it make sense to also have virtual PCs in the cloud too? That would allow you to run applications right next to the data, ensuring optimum performance and end user experience.

Software applications can be installed and run on VMs in the cloud. Using remote desktop services, users can remotely access a desktop and run the software application, all without the need to install the software on their local devices. There are several advantages to having the application running next to the data:

- Latency is reduced to local area network (LAN) levels
- VMs can be configured to provide the computing power required by the application, but the user's device can be less powerful as the application isn't running on their device.
- The user's device is connected to the remote desktop via the internet, but no data is transmitted back to the user's device for processing. Only keyboard and mouse inputs and the display output are sent via the internet, minimizing the impact of latency. It's like watching a streamed movie, it's never on your device.

Chromeleon Clients (workstations) can be installed using desktop services in the cloud. Users can access these clients and process data, removing the need for data to be repeatedly transmitted to and from the cloud. Downloads over the internet only occur when a sequence is started or restarted. Sequences are downloaded to local Data Vaults, where the data is recorded and transmitted back to the central Chromeleon Data Vault in the cloud. Once a sequence is completed, any further processing occurs inside the cloud.

Cloud regions, availability zones, and failover

CSPs operate data centers within geographical regions, and each region consists of data center clusters called availability zones. These data centers are logically and physically separated and connected through dedicated, low latency networks (typically 1–2 ms round trip).

Cloud infrastructure can be deployed in multiple data centers within an availability zone. Applications can be run across these distributed computing resources. In the event of a disaster at one data center, other separate data centers will continue to operate or can be activated so your application can keep running. These distributed resources may all be active as part of routine operations, and if there is an outage in one center, the remaining resources take up the slack. This is a type of failover known as round-robin failover or load balancing, as seen in Figure 5A. All three servers are part of normal operations. If there is a disaster in the East availability zone and server A stops working, servers B and C, in separate zones, take up the load and the application continues to work.

Alternatively, failover may be configured using a secondary, standby server. This server is not routinely in use, but is a backup, reserve server that can automatically switch on in the event of an outage as seen in Figure 5B. If there is a problem with server A in the East availability zone, then server B in the West zone becomes active and the application can continue to operate.

Failover is not configured in the cloud by default. It must be part of your cloud design and based upon your business continuity plan. Once configured, failover must be periodically tested to ensure if the worst does happen, it will minimize disruption to your business processes.

Data and application backup

One of the unwritten rules of information technology is that you haven't got a copy of your data unless you have two. The purpose of any backup is that in the event of a disaster, the system and data can be recovered to an operational state as quickly as possible. CSPs offer cloud backup services, where both data and applications can be backed up to a remote server. These services can also be used to backup on-premises systems and data and has the advantage of keeping the copies geographically separate. With cloud-based systems, it is good practice to store backups in separate availability zones. Organizations' disaster recovery plans should define how business critical systems and data are recovered. A Recovery Time Objective (RTO) defines how long recovery is expected to take. A Recovery Point Objective (RPO) sets the frequency of data backups. How many days, hours, minutes of data can you afford to lose?

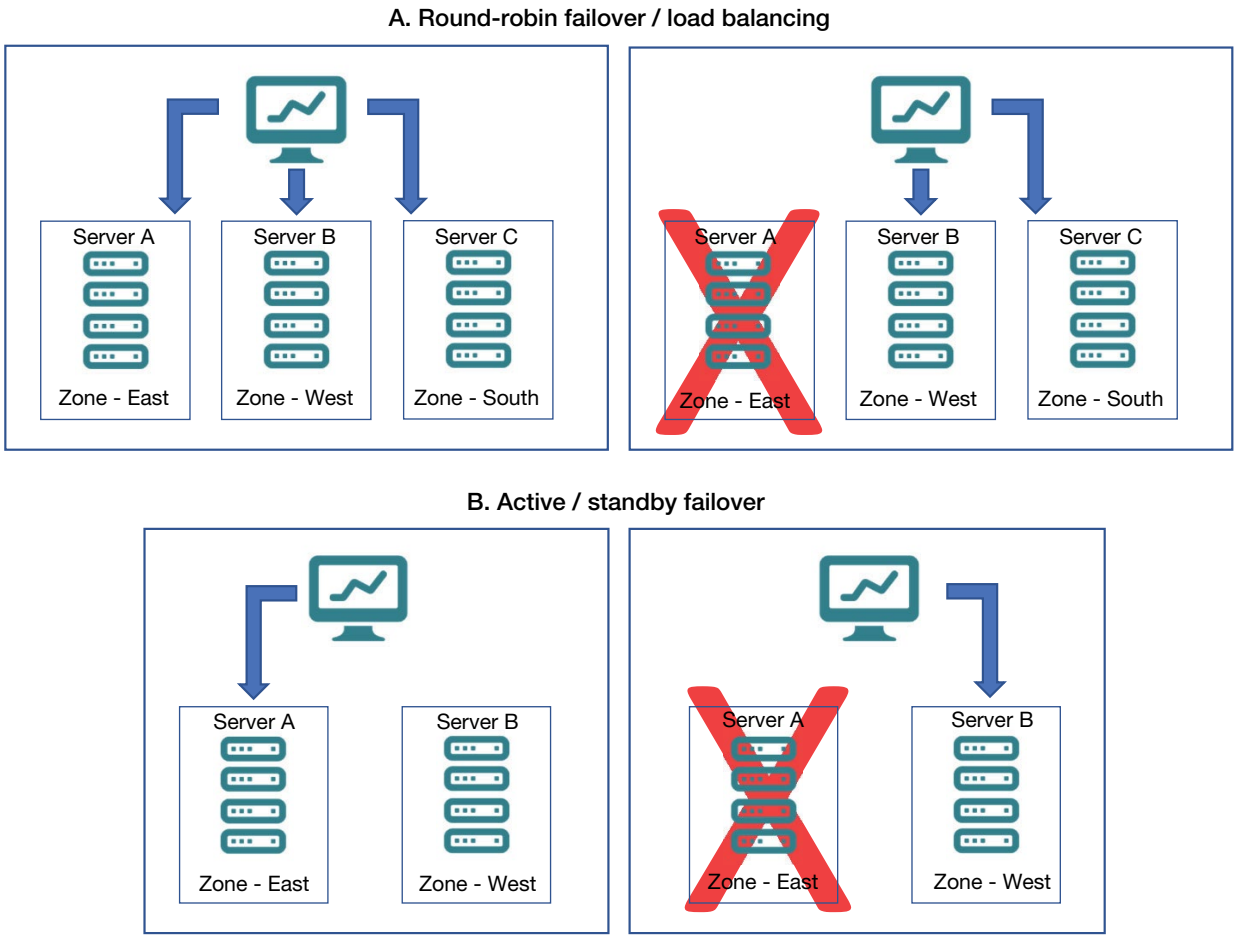

Figure 5. Failover types

Data backups are not automatic and must be part of the design of a cloud system. In regulated environments, it is not only necessary to validate the data backup process as part of system implementation, but backups must be periodically verified by restoring and comparing data. It makes sense not to restore backup data into the live system. What happens if the backup isn't viable? You lose your data. The cloud allows you to create a separate, isolated environment where the application can be installed and run. The data can be restored here without affecting your live processes.

A selection of data can be examined in both systems, and once the tests have been concluded, the restored environment can be switched off. The use of IaC templates can accelerate the creation of the infrastructure needed to test backup files.

Summary

The cloud can seem a complex place. There are many acronyms and concepts, such as infrastructure as code, which can seem like an alien language, but a cloud-based system is no different than an on-premises system. "There is no cloud, it's just someone else's computer."

Cloud compute instances are the equivalent of your local servers or PCs, with processers, memory, and data storage. Instead of physical servers in an on-premises server room, in the cloud we have virtual servers, VPSs, which are provisioned inside the larger cloud data center infrastructure. The on-premises server room is replaced by the VPC, which keeps VPS, application, and data separate from the rest of the private cloud. VPSs within the VPC are networked together in a VLAN, the equivalent of the on-premise LAN. An on-premises server room is connected to other onsite departments via a LAN. In the cloud, the servers are connected to the site via a wide area network (WAN)—the internet.

The cloud can be viewed as a virtual version of your on-premises infrastructure; it's just provisioned and accessed slightly differently.

Glossary

VPC - Virtual private cloud: The division of the public cloud into secure private infrastructure and networks. A network of virtual servers and machines connected to site via the internet.

VPN - Virtual private network: Extension of a private network over a public network, such as the internet, using encryption to create a secure, virtual point-to-point connection.

CSP - Cloud service provider: A third party provider of cloud-based infrastructure, for example Amazon Web services (AWS), Microsoft Azure, or Google Cloud.

VM – Virtual machine: Virtualization of computer systems using hypervisor software, which allows the creation of isolated, virtual computers within a host computer. The host resources can be divided between VMs, depending on application requirements.

VPS – Virtual private server: A VM that acts as a secure, private server for a single organization.

IaC – Infrastructure as Code: Use of machine-readable code files to provision and manage cloud infrastructure. Allows the creation of cloud infrastructure templates.

VDS – Virtual dedicated server: As per VPS

LAN – Local area network: Computers or devices connected to form a network that is confined to a defined physical location. Typically restricted to a single organization.

WAN – Wide area network: A network that is not restricted to a single organization or physical location. Computers and devices in a LAN can communicate with other devices in separate LAN over a WAN, for example the internet.

VLAN – Virtual local area network: Essentially a custom subnetwork, where computers or devices on physically separate LANs can be grouped together in a partitioned virtual network. For example, laboratory instrumentation may be connected to an individual LAN in different laboratories on a single site or multiple sites, but can operate in a single network using VLAN.

IP address - Internet protocol address: A unique numerical label that identifies each computer or device attached to a network, allowing it to communicate over the network.

Learn more at **thermofisher.com/chromeleon**

tn-000796-na-en 03/07/22