1. Illustrate different service model of cloud.

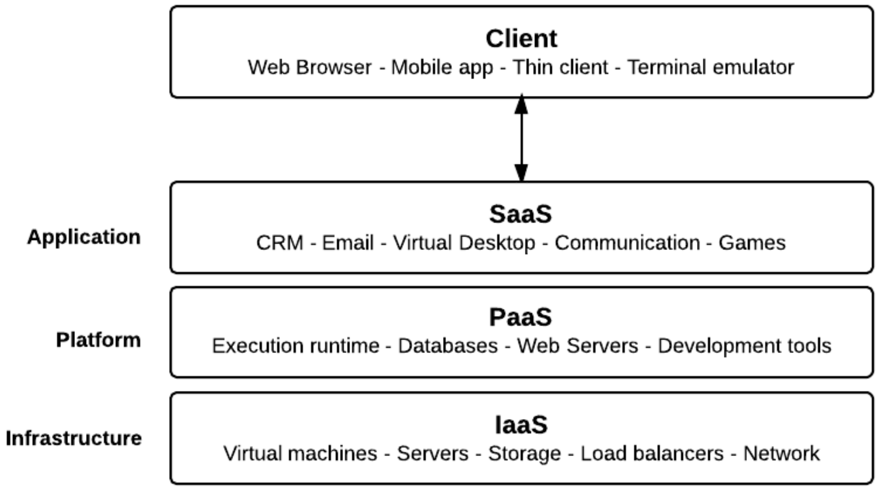
Client
Web Browser - Mobile app - Thin client - Terminal emulator
Application
SaaS
CRM - Email - Virtual Desktop - Communication - Games
Platform
PaaS
Execution runtime - Databases - Web Servers - Development tools
Infrastructure
IaaS
Virtual machines - Servers - Storage - Load balancers - Network

Fig. : Cloud Services

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Although cloud computing has evolved over the time it has been majorly divided into three broad service categories:

1. Infrastructure as a Service(IAAS),
2. Platform as a Service (PAAS) and
3. Software as a Service (SAAS)

1. Infrastructure as a Service (IAAS)

- Infrastructure as a Service (IAAS) is a form of cloud computing that provides virtualized computing resources over the internet.
- In an IAAS model, a third party provider hosts hardware, software, servers, storage and other infrastructure components on the behalf of its users.
- IAAS providers also host users' applications and handle tasks including system maintenance backup and resiliency planning.
- IAAS platforms offer highly scalable resources that can be adjusted on-demand which makes it a well-suited for workloads that are temporary, experimental or change unexpectedly.
- Other characteristics of IAAS environments include the automation of administrative tasks, dynamic scaling, desktop virtualization and policy based services.
- Technically, the IaaS market has a relatively low barrier of entry, but it may require substantial financial investment in order to build and support the cloud infrastructure.
- Mature open-source cloud management frameworks like OpenStack are available to everyone, and provide strong a software foundation for companies that want to build their private cloud or become a public cloud provider.

IAAS- Network:

- There are two major network services offered by public cloud service providers:

1. load balancing and
2. DNS (domain name systems).

- Load balancing provides a single point of access to multiple servers that run behind it. A load balancer is a network device that distributes network traffic among servers using specific load balancing algorithms.
- DNS is a hierarchical naming system for computers, or any other naming devices that use IP addressing for network identification – a DNS system associates domain names with IP addresses.

2. Platform as a Service (PAAS)

- Platform as a Service (PAAS) is a cloud computing model that delivers applications over the internet.
- In a PAAS model, a cloud provider delivers hardware and software tools, usually those needed for application development, to its users as a service.
- A PAAS provider hosts the hardware and software on its own infrastructure. As a result, PAAS frees users from having to install in-house hardware and software to develop or run a new application.
- PAAS doesn’t replace a business' entire infrastructure but instead a business relies on PAAS providers for key services, such as Java development or application hosting.
- A PAAS provider, however, supports all the underlying computing and software, users only need to login and start using the platform-usually through a Web browser interface.
- PAAS providers then charge for that access on a per-use basis or on monthly basis.
- Some of the main characteristics of PAAS are :
  1) Scalability and auto-provisioning of the underlying infrastructure.
  2) Security and redundancy.
  3) Build and deployment tools for rapid application management and deployment.
  4) Integration with other infrastructure components such as web services, databases, and LDAP.
  5) Multi-tenancy, platform service that can be used by many concurrent users.
  6) Logging, reporting, and code instrumentation.
  7) Management interfaces and/or API.

3. Software as a Service (SAAS)

- Software as a Service (SAAS) is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet.
- SAAS has become increasingly prevalent delivery model as underlying technologies that support Web services and service- oriented architecture (SOA) mature and new development approaches, such as Ajax, become popular.
- SAAS is closely related to the ASP (Application service provider) and on demand computing software delivery models.
- IDC identifies two slightly different delivery models for SAAS which are
  1) the hosted application model and
  2) the software development model.
- Some of the core benefits of using SAAS model are:
  1) Easier administration.
  2) Automatic updates and patch management.
  3) Compatibility: all users will have the same version of software.
  4) Easier collaboration, for the same reason.
  5) Global accessibility.

2. What are the characteristics of cloud computing?

The five essential characteristics of cloud computing:

1. **On-demand self-service:** A consumer can separately provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.
2. **Broad network access:** Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops and workstations).
3. **Resource pooling:** The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state or datacenter). Examples of resources include storage, processing, memory and network bandwidth.
4. **Rapid elasticity:** Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward matching with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.
5. **Measured service:** Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth and active user accounts). Resource usage can be monitored, controlled and reported, providing transparency for the provider and consumer.

3. List cloud deployment models.

Following are the four types of Cloud Deployment Models identified by NIST.

1. Private cloud
2. Community cloud
3. Public cloud
4. Hybrid cloud

4. Private cloud

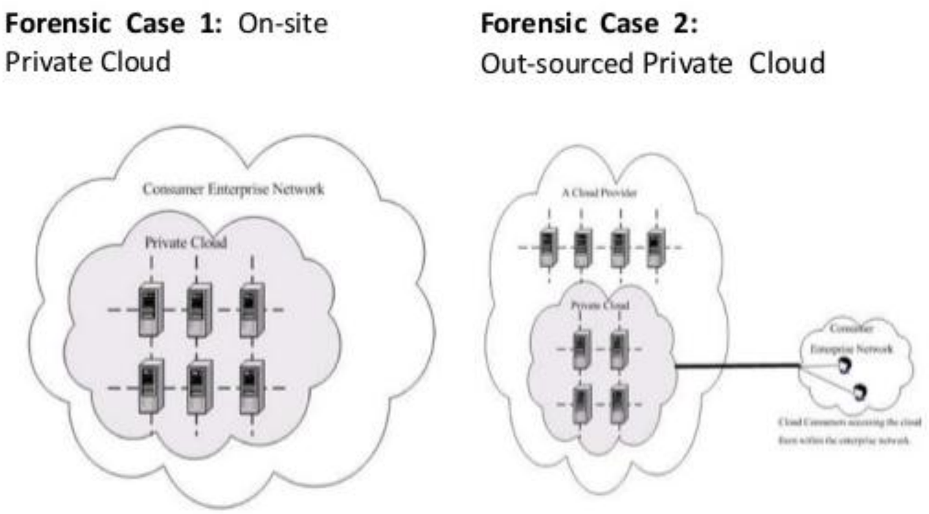
Forensic Case 1: On-site Private Cloud
Consumer Enterprise Network
Private Cloud
Forensic Case 2: Out-sourced Private Cloud
A Cloud Provider
Private Cloud
Consumer Enterprise Network
Cloud Consumers accessing the cloud from within the enterprise network

Fig.: Private Cloud

- The cloud infrastructure is operated solely for an organization.
- Contrary to popular belief, private cloud may exist off premises and can be managed by a third party. Thus, two private cloud scenarios exist, as follows:

On-site Private Cloud

- Applies to private clouds implemented at a customer's premises.

Outsourced Private Cloud

- Applies to private clouds where the server side is outsourced to a hosting company.

Examples of Private Cloud:

- Eucalyptus
- Ubuntu Enterprise Cloud - UEC (powered by Eucalyptus)
- Amazon VPC (Virtual Private Cloud)
- VMware Cloud Infrastructure Suite
- Microsoft ECI data center.

5. Community cloud

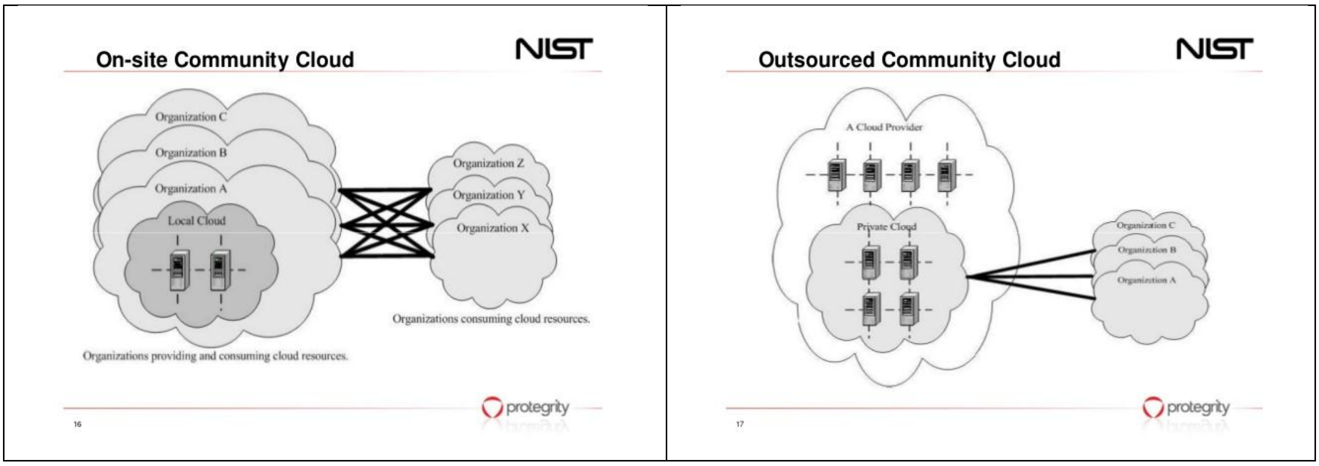

Fig. Community Cloud

- The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations).
- Government departments, universities, central banks etc. often find this type of cloud useful.
- Community cloud also has two possible scenarios:

On-site Community Cloud Scenario

- Applies to community clouds implemented on the premises of the customers composing a community cloud.

Outsourced Community Cloud

- Applies to community clouds where the server side is outsourced to a hosting company.

Examples of Community Cloud:

- Google Apps for Government
- Microsoft Government Community Cloud

6. Public cloud

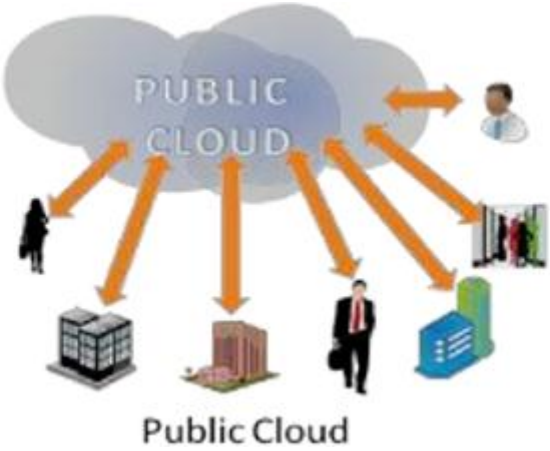

Fig. : Public Cloud

- The most ubiquitous, and almost a synonym for, cloud computing.
- The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

Examples of Public Cloud:

- Google App Engine
- Microsoft Windows Azure
- IBM Smart Cloud
- Amazon EC2

7. Hybrid Cloud

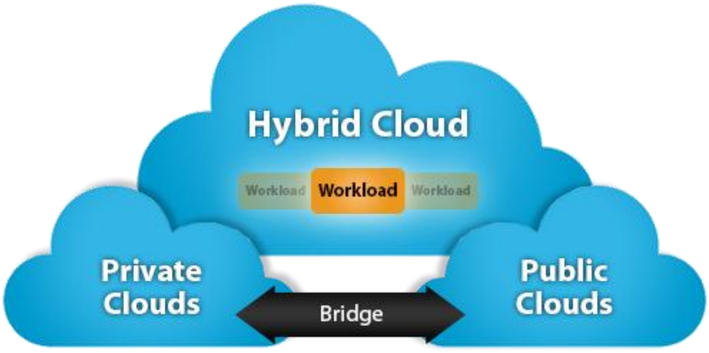

Fig. : Hybrid Cloud

- The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

Examples of Hybrid Cloud:

- Windows Azure (capable of Hybrid Cloud)
- VMware vCloud (Hybrid Cloud Services)

8. Compare different cloud service providers.

	Amazon Web Service	Azure	Rackspace
Introduction	Amazon Web Services (AWS) is a collection of remote computing services (also called web services) that together make up a cloud computing platform, offered over the Internet by Amazon.com.	Azure is a cloud computing platform and infrastructure, created by Microsoft, for building, deploying and managing applications and services through a global network of Microsoft-managed datacenters.	Rackspace is a managed cloud computing provider offering high percentage availability of applications based on RAID10.
Distinguishing Features	Rich set of services and integrated monitoring tools; competitive pricing model.	Easy-to-use administration tool, especially for Windows admins.	Easy to use control panel, especially for non-system administrators.
Virtualization	Xen hypervisor	Microsoft Hyper-V	Opensource (Xen, Kvm) and VMware
Base OS	Linux (+QEMU) and Windows	Windows and Linux	Ubuntu
Pricing model	Pay-as-you-go, then subscription	Pay-as-you-go	Pay-as-you-go
Major products	Elastic block store, IP addresses, virtual private cloud, cloud watch, Cloud Front, clusters etc.	Server Failover Clustering, Network Load Balancing, SNMP Services, Storage Manager for SANs, Windows Internet Name Service, Disaster Recovery to Azure, Azure Caching and Azure Redis Cache.	Managed cloud, block storage, monitoring
CDN Features	Origin-Pull, Purge, Gzip compression, Persistent connections, Caching headers, Custom CNAMEs, Control Panel & stats, Access Logs.	Robust security, Lower latencies, Massively scalable, Capacity on demand.	Rackspace provide CDN services through a partnership with Akamai's service.
Access interface	Web-based, API, console	Web interface	Web-based control panel
Preventive measures	Moderate	Basic	Basic
Reactive measures	Moderate	Basic	Basic
Reliability	Good	Average	Good
Scalability	Good	Good	Good
Support	Good and chargeable	Good	Excellent
Availability (%)	99.95	99.95	99.99
Server Performance (Over a period)	Good	Excellent and consistent	Average
Tools/ framework	Amazon machine image (AMI), Java, PHP, Python, Ruby	PHP, ASP.NET, Node.js, Python	-
Database RDS	MySQL, MsSQL, Oracle	Microsoft SQL Database	MySQL

9. Explain Virtualization and Hypervisor.

Hypervisor

- It is the part of the private cloud that manages the virtual machines, i.e. it is the part (program) that enables multiple operating systems to share the same hardware.
- Each operating system could use all the hardware (processor, memory) if no other operating system is on. That is the maximum hardware available to one operating system in the cloud.
- Nevertheless, the hypervisor is what controls and allocates what portion of hardware resources each operating system should get, in order every one of them to get what they need and not to disrupt each other.

Virtualization

- Virtualization is changing the mindset from physical to logical.

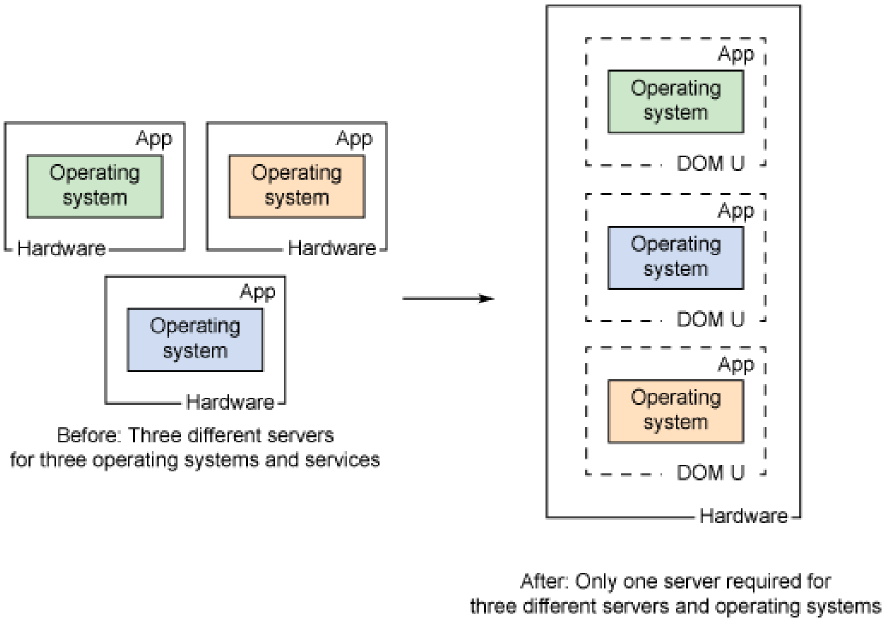

Fig. : Virtualization

- What virtualization means is creating more logical IT resources, called virtual systems, within one physical system. That's called system virtualization.
- It most commonly uses the hypervisor for managing the resources for every virtual system. The hypervisor is a software that can virtualize the hardware resources.

There are two types of hypervisors:

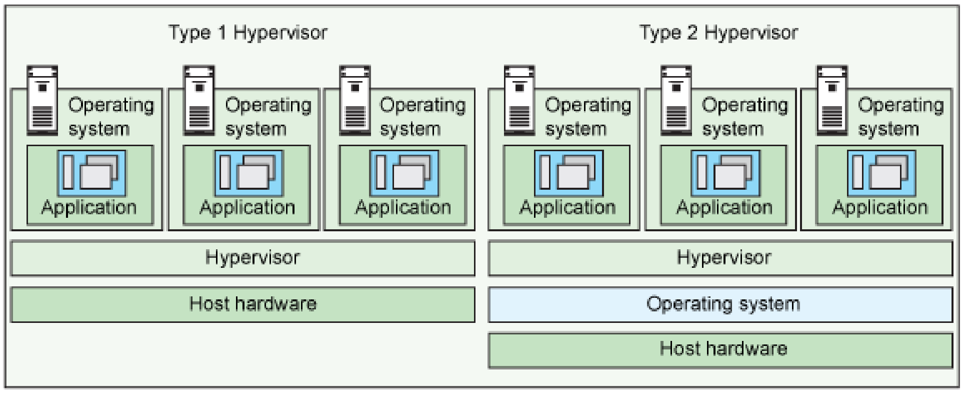

Fig. : Types of Hypervisors

- **Type 1 hypervisor**: hypervisors run directly on the system hardware – A "bare metal" embedded hypervisor. Examples are:
  1) VMware ESX and ESXi
  2) Microsoft Hyper-V
  3) Citrix XenServer
  4) Oracle VM
- **Type 2 hypervisor**: hypervisors run on a host operating system that provides virtualization services, such as I/O device support and memory management. Examples are:
  1) VMware Workstation/Fusion/Player
  2) Microsoft Virtual PC
  3) Oracle VM VirtualBox
  4) Red Hat Enterprise Virtualization

10. How machine Imaging help to achieve the goal of cloud computing?

- Machine imaging is a process that is used to achieve the goal of system portability, provision, and deploy systems in the cloud through capturing the state of systems using a system image.
- A system image makes a copy or a clone of the entire computer system inside a single file.
- The image is made by using a program called system imaging program and can be used later to restore a system image.
- For example Amazon Machine Image (AMI) is a system image that is used in the cloud computing.
- The Amazon Web Services uses AMI to store copies of a virtual machine.
- An AMI is a file system image that contains an operating system, all device drivers, and any applications and state information that the working virtual machine would have.
- The AMI files are encrypted and compressed for security purpose and stored in Amazon S3 (Simple Storage System) buckets as a set of 10MB chunks.
- Machine imaging is mostly run on virtualization platform due to this it is also called as Virtual Appliances and running virtual machines are called instances.
- Because many users share clouds, the cloud helps you track information about images, such as ownership, history, and so on.
- The IBM SmartCloud Enterprise knows what organization you belong to when you log in.
- You can choose whether to keep images private, exclusively for your own use, or to share with other users in your organization.
- If you are an independent software vendor, you can also add your images to the public catalog.

11. What is Virtual Machine? Explain Virtual Machine types.

- A virtual machine (VM) is an operating system (OS) or application environment that is installed on software, which imitates dedicated hardware. The end user has the same experience on a virtual machine as they would have on dedicated hardware.

Virtual Machine Types

1. General Purpose

- This family includes the M1 and M3 VM types. These types provide a balance of CPU, memory, and network resources, which makes them a good choice for many applications. The VM types in this family range in size from one virtual CPU with two GB of RAM to eight virtual CPUs with 30 GB of RAM. The balance of resources makes them ideal for running small and mid-size databases, more memory-hungry data processing tasks, caching fleets, and backend servers.

- M1 types offer smaller instance sizes with moderate CPU performance. M3 types offer larger number of virtual CPUs that provide higher performance. We recommend you use M3 instances if you need general-purpose instances with demanding CPU requirements.

2. Compute Optimized

- This family includes the C1 and CC2 instance types, and is geared towards applications that benefit from high compute power. Compute-optimized VM types have a higher ratio of virtual CPUs to memory than other families but share the NCs with non-optimized ones. We recommend this type if you are running any CPU-bound scale-out applications. CC2 instances provide high core count (32 virtual CPUs) and support for cluster networking. C1 instances are available in smaller sizes and are ideal for scaled-out applications at massive scale.

3. Memory Optimized

- This family includes the CR1 and M2 VM types and is designed for memory-intensive applications. We recommend these VM types for performance-sensitive database, where your application is memory-bound. CR1 VM types provide more memory and faster CPU than do M2 types. CR1 instances also support cluster networking for bandwidth intensive applications. M2 types are available in smaller sizes, and are an excellent option for many memory-bound applications.

4. Micro

- This Micro family contains the T1 VM type. The t1.micro provides a small amount of consistent CPU resources and allows you to increase CPU capacity in short bursts when additional cycles are available. We recommend this type for lower throughput applications like a proxy server or administrative applications, or for low-traffic websites that occasionally require additional compute cycles. We do not recommend this VM type for applications that require sustained CPU performance.

12. Explain AWS Infrastructure.

- Amazon Web Services (AWS) is a global public cloud provider, and as such, it has to have a global network of infrastructure to run and manage its many growing cloud services that support customers around the world.
- Now we'll take a look at the components that make up the AWS Global Infrastructure.
  1) Availability Zones (AZs)
  2) Regions
  3) Edge Locations
  4) Regional Edge Caches
- If you are deploying services on AWS, you'll want to have a clear understanding of each of these components, how they are linked, and how you can use them within your solution to YOUR maximum benefit. Let's take a closer look.

1) Availability Zones (AZ)

- AZs are essentially the physical data centers of AWS. This is where the actual compute, storage, network, and database resources are hosted that we as consumers provision within our Virtual Private Clouds (VPCs).
- A common misconception is that a single availability zone is equal to a single data center. This is not the case. In fact, it's likely that multiple data centers located close together form a single availability zone.

- Each AZ will always have at least one other AZ that is geographically located within the same area, usually a city, linked by highly resilient and very low latency private fiber optic connections. However, each AZ will be isolated from the others using separate power and network connectivity that minimizes impact to other AZs should a single AZ fail.
- These low latency links between AZs are used by many AWS services to replicate data for high availability and resilience purposes.
- Multiple AZs within a region allows you to create highly available and resilient applications and services.
- By architecting your solutions to utilize resources across more than one AZ ensures that minimal or no impact will occur to your infrastructure should an AZ experience a failure, which does happen.
- Anyone can deploy resources in the cloud, but architecting them in a way that ensures your infrastructure remains stable, available, and resilient when faced with a disaster is a different matter.
- Making use of at least two AZs in a region helps you maintain high availability of your infrastructure and it's always a recommended best practice.

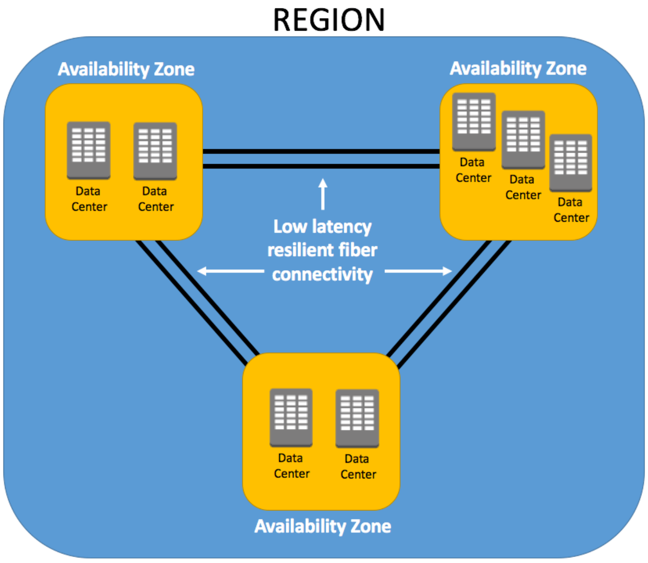

Fig. : Availability Zone and Region

2) Regions

- Region is a collection of availability zones that are geographically located close to one other.
- This is generally indicated by AZs within the same city. AWS has deployed them across the globe to allow its worldwide customer base to take advantage of low latency connections.
- Each Region will act independently of the others, and each will contain at least two Availability Zones.
- Example:if an organization based in London was serving customers throughout Europe, there would be no logical sense to deploy services in the Sydney Region simply due to the latency response times for its customers. Instead, the company would select the region most appropriate for them and their customer base, which may be the London, Frankfurt, or Ireland Region.
- Having global regions also allows for compliance with regulations, laws, and governance relating to data storage (at rest and in transit).
- Example: you may be required to keep all data within a specific location, such as Europe. Having multiple regions within this location allows an organization to meet this requirement.
- Similarly to how utilizing multiple AZs within a region creates a level of high availability, the same can be applied to utilizing multiple regions.
- You may want to use multiple regions if you are a global organization serving customers in different countries that have specific laws and governance about the use of data.

- In this case, you could even connect different VPCs together in different regions.
- The number of regions is increasing year after year as AWS works to keep up with the demand for cloud computing services.
- In July 2017, there are currently 16 Regions and 43 Availability Zones, with 4 Regions and 11 AZs planned.

3) Edge Locations

- Edge Locations are AWS sites deployed in major cities and highly populated areas across the globe. They far outnumber the number of availability zones available.
- While Edge Locations are not used to deploy your main infrastructures such as EC2 instances, EBS storage, VPCs, or RDS resources like AZs, they are used by AWS services such as AWS CloudFront and AWS Lambda@Edge (currently in Preview) to cache data and reduce latency for end user access by using the Edge Locations as a global Content Delivery Network (CDN).
- As a result, Edge Locations are primarily used by end users who are accessing and using your services.
- For example, you may have your website hosted on EC2 instances and S3 (your origin) within the Ohio region with a configured CloudFront distribution associated. When a user accesses your website from Europe, they would be re-directed to their closest Edge Location (in Europe) where cached data could be read on your website, significantly reducing latency.

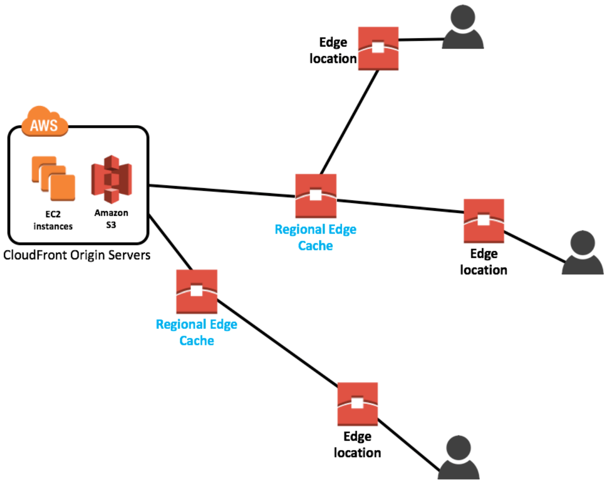

Fig. : Edge Location and Regional Edge Cache

4) Regional Edge Cache

- In November 2016, AWS announced a new type of Edge Location, called a Regional Edge Cache.
- These sit between your CloudFront Origin servers and the Edge Locations.
- A Regional Edge Cache has a larger cache-width than each of the individual Edge Locations, and because data expires from the cache at the Edge Locations, the data is retained at the Regional Edge Caches.
- Therefore, when data is requested at the Edge Location that is no longer available, the Edge Location can retrieve the cached data from the Regional Edge Cache instead of the Origin servers, which would have a higher latency.

13. Describe Glacier Storage Service.

- Amazon Glacier is an extremely low-cost storage service that provides secure, durable, and flexible storage for data backup and archival.
- With Amazon Glacier, customers can reliably store their data for as little as $0.004 per gigabyte per month.

- Amazon Glacier enables customers to offload the administrative burdens of operating and scaling storage to AWS, so that they don’t have to worry about capacity planning, hardware provisioning, data replication, hardware failure detection and repair, or time-consuming hardware migrations.
- Amazon Glacier enables any business or organization to easily and cost effectively retain data for months, years, or decades.
- With Amazon Glacier, customers can now cost effectively retain more of their data for future analysis or reference, and they can focus on their business rather than operating and maintaining their storage infrastructure.
- Customers seeking compliance storage can deploy compliance controls using Vault Lock to meet regulatory and compliance archiving requirements.

Benefits of Glacier Storage Service.

1. RETRIEVALS AS QUICK AS 1-5 MINUTES

- Amazon Glacier provides three retrieval options to fit your use case. Expedited retrievals typically return data in 1-5 minutes, and are great for Active Archive use cases. Standard retrievals typically complete between 3-5 hours work, and work well for less time-sensitive needs like backup data, media editing, or long-term analytics. Bulk retrievals are the lowest-cost retrieval option, returning large amounts of data within 5-12 hours.

2. UNMATCHED DURABILITY & SCALABILITY

- Amazon Glacier runs on the world’s largest global cloud infrastructure, and was designed for 99.999999999% of durability. Data is automatically distributed across a minimum of three physical Availability Zones that are geographically separated within an AWS Region, and Amazon Glacier can also automatically replicate data to any other AWS Region.

3. MOST COMPREHENSIVE SECURITY & COMPLIANCE CAPABILITIES

- Amazon Glacier offers sophisticated integration with AWS CloudTrail to log, monitor and retain storage API call activities for auditing, and supports three different forms of encryption. Amazon Glacier also supports security standards and compliance certifications including SEC Rule 17a-4, PCI-DSS, HIPAA/HITECH, FedRAMP, EU Data Protection Directive, and FISMA, and Amazon Glacier Vault Lock enables WORM storage capabilities, helping satisfy compliance requirements for virtually every regulatory agency around the globe.

4. LOW COST

- Amazon Glacier is designed to be the lowest cost AWS object storage class, allowing you to archive large amounts of data at a very low cost. This makes it feasible to retain all the data you want for use cases like data lakes, analytics, IoT, machine learning, compliance, and media asset archiving. You pay only for what you need, with no minimum commitments or up-front fees.

5. MOST SUPPORTED PLATFORM WITH THE LARGEST ECOSYSTEM

- In addition to integration with most AWS services, the Amazon object storage ecosystem includes tens of thousands of consulting, systems integrator and independent software vendor partners, with more joining every month. And the AWS Marketplace offers 35 categories and more than 3,500 software listings from over 1,100 ISVs that are pre-configured to deploy on the AWS Cloud. AWS Partner Network partners have adapted their services and software to work with Amazon S3 and Amazon Glacier for solutions like Backup & Recovery, Archiving, and Disaster Recovery. No other cloud provider has more partners with solutions that are pre-integrated to work with their service.

6. QUERY IN PLACE

- Amazon Glacier is the only cloud archive storage service that allows you to query data in place and retrieve only the subset of data you need from within an archive. Amazon Glacier Select helps you reduce the total cost of ownership by extending your data lake into cost-effective archive storage.

14. Explain Amazon S3? Explain Amazon S3 API? What are the operations we can execute through API

Amazon Simple Storage Service (S3)

- Amazon S3 has a simple web services interface that you can use to store and retrieve any amount of data, at any time, from anywhere on the web.
- Amazon S3 is intentionally built with a minimal feature set that focuses on simplicity and robustness.
- Following are some of advantages of the Amazon S3 service:
    - **Create Buckets** – Create and name a bucket that stores data. Buckets are the fundamental container in Amazon S3 for data storage.
    - **Store data in Buckets** – Store an infinite amount of data in a bucket. Upload as many objects as you like into an Amazon S3 bucket. Each object can contain up to 5 TB of data. Each object is stored and retrieved using a unique developer-assigned key.
    - **Download data** – Download your data any time you like or allow others to do the same.
    - **Permissions** – Grant or deny access to others who want to upload or download data into your Amazon S3 bucket.
    - **Standard interfaces** – Use standards-based REST and SOAP interfaces designed to work with any Internet-development toolkit.

Amazon S3 Application Programming Interfaces (API)

- The Amazon S3 architecture is designed to be programming language-neutral, using their supported interfaces to store and retrieve objects.
- Amazon S3 provides a REST and a SOAP interface.
- They are similar, but there are some differences. For example, in the REST interface, metadata is returned in HTTP headers. Because we only support HTTP requests of up to 4 KB (not including the body), the amount of metadata you can supply is restricted.

The REST Interface

- The REST API is an HTTP interface to Amazon S3.
- Using REST, you use standard HTTP requests to create, fetch, and delete buckets and objects.
- You can use any toolkit that supports HTTP to use the REST API.
- You can even use a browser to fetch objects, as long as they are anonymously readable.
- The REST API uses the standard HTTP headers and status codes, so that standard browsers and toolkits work as expected.
- In some areas, they have added functionality to HTTP (for example, we added headers to support access control).

The SOAP Interface

- SOAP support over HTTP is deprecated, but it is still available over HTTPS.
- New Amazon S3 features will not be supported for SOAP.
- The SOAP API provides a SOAP 1.1 interface using document literal encoding.

- The most common way to use SOAP is to download the WSDL, and use a SOAP toolkit such as Apache Axis or Microsoft .NET to create bindings, and then write code that uses the bindings to call Amazon S3.

Operations we can execute through API

- Login into Amazon S3.
- Uploading.
- Retrieving.
- Deleting etc.

15. Write S3 URL naming conventions.

- You can access your bucket using the Amazon S3 console. Using the console UI, you can perform almost all bucket operations without having to write any code.
- If you access a bucket programmatically, note that Amazon S3 supports RESTful architecture in which your buckets and objects are resources, each with a resource URI that uniquely identifies the resource.
- Amazon S3 supports both virtual-hosted–style and path-style URLs to access a bucket.
- In a **virtual-hosted–style URL**, the bucket name is part of the domain name in the URL. For example:
  - http://bucket.s3.amazonaws.com
  - http://bucket.s3-aws-region.amazonaws.com
- In a virtual-hosted–style URL, you can use either of these endpoints.
- If you make a request to the http://bucket.s3.amazonaws.com endpoint, the DNS has sufficient information to route your request directly to the Region where your bucket resides.
- In a **path-style URL**, the bucket name is not part of the domain (unless you use a Region-specific endpoint). For example:
  - US East (N. Virginia) Region endpoint, http://s3.amazonaws.com/bucket
  - Region-specific endpoint, http://s3-aws-region.amazonaws.com/bucket
- In a path-style URL, the endpoint you use must match the Region in which the bucket resides.
- For example, if your bucket is in the South America (São Paulo) Region, you must use the http://s3-sa-east-1.amazonaws.com/bucket endpoint. If your bucket is in the US East (N. Virginia) Region, you must use the http://s3.amazonaws.com/bucket endpoint.

16. Describe Elastic Block Store (EBS)

- Amazon Elastic Block Store is an AWS block storage system that is best used for storing persistent data.
- Often incorrectly referred to as Elastic Block Storage, Amazon EBS provides highly available block level storage volumes for use with Amazon EC2 instances.
- An EC2 instance is a virtual server in Amazon's Elastic Compute Cloud (EC2) for running applications on the Amazon Web Services (AWS) infrastructure.
- To begin, create an EBS volume (General Purpose, Provisioned IOPS or Magnetic), pick a size for it (up to a terabyte of data) and attach that to any one of your EC2 instances.
- An EBS volume can only be attached to one instance at a time but if you need to have multiple copies of the volume, you can take a snapshot and create another volume from that snapshot and attach it to another drive.
- A snapshot file is equivalent to a backup of whatever the EBS volume looks like at the time. For every snapshot you create, you can make an identical EC2 instance. This will allow you to publish identical content on multiple servers.

- Amazon EBS is ideal if you’re doing any substantial work with EC2, you want to keep data persistently on a file system, and you want to keep that data around even after you shut down your EC2 instance.
- EC2 instances have local storage that you can use as long as you’re running the instance, but as soon as you shut down the instance you lose the data that was on there.
- If you want to save anything, you need to save it on Amazon EBS. Because EC2 is like having a local drive on the machine, you can access and read the EBS volumes anytime once you attach the file to an EC2 instance.

17. What do you mean by Identity Management and Access Management?

- Identity and access management (IAM) is a framework for business processes that facilitates the management of electronic or digital identities.
- The framework includes the organizational policies for managing digital identity as well as the technologies needed to support identity management.
- With IAM technologies, IT managers can control user access to critical information within their organizations.
- Identity and access management products offer role-based access control, which lets system administrators regulate access to systems or networks based on the roles of individual users within the enterprise.
- In this context, access is the ability of an individual user to perform a specific task, such as view, create or modify a file.
- Roles are defined according to job competency, authority and responsibility within the enterprise.
- Systems used for identity and access management include single sign-on systems, multifactor authentication and access management.
- These technologies also provide the ability to securely store identity and profile data as well as data governance functions to ensure that only data that is necessary and relevant is shared.
- These products can be deployed on premises, provided by a third party vendor via a cloud-based subscription model or deployed in a hybrid cloud.

18. How we can say our data is secure on cloud and what are the challenges to achieve it?

- A number of security threats are associated with cloud data services: not only traditional security threats, such as network eavesdropping, illegal invasion, and denial of service attacks, but also specific cloud computing threats, such as side channel attacks, virtualization vulnerabilities, and abuse of cloud services.
- The following security requirements limit the threats if we achieve that requirement than we can say our data is safe on cloud.
- **Identity management**
    - Every enterprise will have its own identity management system to control access to information and computing resources.
    - Cloud providers either integrate the customer’s identity management system into their own infrastructure, using federation or SSO technology, or a biometric-based identification system, or provide an identity management system of their own.
    - CloudID, for instance, provides privacy-preserving cloud-based and cross-enterprise biometric identification.

- It links the confidential information of the users to their biometrics and stores it in an encrypted fashion.
- Making use of a searchable encryption technique, biometric identification is performed in encrypted domain to make sure that the cloud provider or potential attackers do not gain access to any sensitive data or even the contents of the individual queries.

- **Physical security**
  - Cloud service providers physically secure the IT hardware (servers, routers, cables etc.) against unauthorized access, interference, theft, fires, floods etc. and ensure that essential supplies (such as electricity) are sufficiently robust to minimize the possibility of disruption.
  - This is normally achieved by serving cloud applications from 'world-class' (i.e. professionally specified, designed, constructed, managed, monitored and maintained) data centers.
- **Personnel security**
  - Various information security concerns relating to the IT and other professionals associated with cloud services are typically handled through pre-, para- and post-employment activities such as security screening potential recruits, security awareness and training programs, proactive.
- **Privacy**
  - Providers ensure that all critical data (credit card numbers, for example) are masked or encrypted and that only authorized users have access to data in its entirety. Moreover, digital identities and credentials must be protected as should any data that the provider collects or produces about customer activity in the cloud.
- **Confidentiality**
  - Data confidentiality is the property that data contents are not made available or disclosed to illegal users.
  - Outsourced data is stored in a cloud and out of the owners' direct control. Only authorized users can access the sensitive data while others, including CSPs, should not gain any information of the data.
  - Meanwhile, data owners expect to fully utilize cloud data services, e.g., data search, data computation, and data sharing, without the leakage of the data contents to CSPs or other adversaries.
- **Access controllability**
  - Access controllability means that a data owner can perform the selective restriction of access to her or his data outsourced to cloud.
  - Legal users can be authorized by the owner to access the data, while others cannot access it without permissions.
  - Further, it is desirable to enforce fine-grained access control to the outsourced data, i.e., different users should be granted different access privileges with regard to different data pieces.
  - The access authorization must be controlled only by the owner in untrusted cloud environments.
- **Integrity**
  - Data integrity demands maintaining and assuring the accuracy and completeness of data.
  - A data owner always expects that her or his data in a cloud can be stored correctly and trustworthily.
  - It means that the data should not be illegally tampered, improperly modified, deliberately deleted, or maliciously fabricated.
  - If any undesirable operations corrupt or delete the data, the owner should be able to detect the corruption or loss.
  - Further, when a portion of the outsourced data is corrupted or lost, it can still be retrieved by the data users.

19. Write a note on AWS API Security.

- API Gateway supports multiple mechanisms of access control, including metering or tracking API uses by clients using API keys.
- The standard AWS IAM roles and policies offer flexible and robust access controls that can be applied to an entire API set or individual methods.
- Custom authorizers and Amazon Cognito user pools provide customizable authorization and authentication solutions.

A. Control Access to an API with IAM Permissions

- You control access to Amazon API Gateway with IAM permissions by controlling access to the following two API Gateway component processes:
  - To create, deploy, and manage an API in API Gateway, you must grant the API developer permissions to perform the required actions supported by the API management component of API Gateway.
  - To call a deployed API or to refresh the API caching, you must grant the API caller permissions to perform required IAM actions supported by the API execution component of API Gateway.

B. Use API Gateway Custom Authorizers

- An Amazon API Gateway custom authorizer is a Lambda function that you provide to control access to your API methods.
- A custom authorizer uses bearer token authentication strategies, such as OAuth or SAML. It can also use information described by headers, paths, query strings, stage variables, or context variables request parameters.
- When a client calls your API, API Gateway verifies whether a custom authorizer is configured for the API method. If so, API Gateway calls the Lambda function.
- In this call, API Gateway supplies the authorization token that is extracted from a specified request header for the token-based authorizer, or passes in the incoming request parameters as the input (for example, the event parameter) to the request parameters-based authorizer function.
- You can implement various authorization strategies, such as JSON Web Token (JWT) verification and OAuth provider callout.
- You can also implement a custom scheme based on incoming request parameter values, to return IAM policies that authorize the request. If the returned policy is invalid or the permissions are denied, the API call does not succeed.

C. Use Amazon Cognito User Pools

- In addition to using IAM roles and policies or custom authorizers, you can use an Amazon Cognito user pool to control who can access your API in Amazon API Gateway.
- To use an Amazon Cognito user pool with your API, you must first create an authorizer of the COGNITO_USER_POOLS type and then configure an API method to use that authorizer.
- After the API is deployed, the client must first sign the user in to the user pool, obtain an identity or access token for the user, and then call the API method with one of the tokens, which are typically set to the request's Authorization header.
- The API call succeeds only if the required token is supplied and the supplied token is valid, otherwise, the client isn't authorized to make the call because the client did not have credentials that could be authorized.

D. Use Client-Side SSL Certificates for Authentication by the Backend

- You can use API Gateway to generate an SSL certificate and use its public key in the backend to verify that HTTP requests to your backend system are from API Gateway.
- This allows your HTTP backend to control and accept only requests originating from Amazon API Gateway, even if the backend is publicly accessible.
- The SSL certificates that are generated by API Gateway are self-signed and only the public key of a certificate is visible in the API Gateway console or through the APIs.

E. Create and Use API Gateway Usage Plans

- After you create, test, and deploy your APIs, you can use API Gateway usage plans to extend them as product offerings for your customers.
- You can provide usage plans to allow specified customers to access selected APIs at agreed-upon request rates and quotas that can meet their business requirements and budget constraints.

20. What is elastic load balancing and how it reduces the workload?

- Elastic Load Balancing automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses.
- It can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones.
- Elastic Load Balancing offers three types of load balancers that all feature the high availability, automatic scaling, and robust security necessary to make your applications fault tolerant.
  - Application Load Balancers,
  - Network Load Balancers, and
  - Classic Load Balancers.

Application Load Balancer

- Application Load Balancer is best suited for load balancing of HTTP and HTTPS traffic and provides advanced request routing targeted at the delivery of modern application architectures, including micro services and containers.
- Operating at the individual request level (Layer 7), Application Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) based on the content of the request.

Network Load Balancer

- Network Load Balancer is best suited for load balancing of TCP traffic where extreme performance is required.
- Operating at the connection level (Layer 4), Network Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) and is capable of handling millions of requests per second while maintaining ultra-low latencies.
- Network Load Balancer is also optimized to handle sudden and volatile traffic patterns.

Classic Load Balancer

- Classic Load Balancer provides basic load balancing across multiple Amazon EC2 instances and operates at both the request level and connection level.
- Classic Load Balancer is intended for applications that were built within the EC2-Classic network.

Benefits of Elastic Load Balancing for reducing workload

Highly Available

- Elastic Load Balancing automatically distributes incoming traffic across multiple targets – Amazon EC2 instances, containers, and IP addresses – in multiple Availability Zones and ensures only healthy targets receive traffic. Elastic Load Balancing can also load balance across a Region, routing traffic to healthy targets in different Availability Zones.

Secure

- Elastic Load Balancing works with Amazon Virtual Private Cloud (VPC) to provide robust security features, including integrated certificate management and SSL decryption. Together, they give you the flexibility to centrally manage SSL settings and offload CPU intensive workloads from your applications.

Elastic

- Elastic Load Balancing is capable of handling rapid changes in network traffic patterns. Additionally, deep integration with Auto Scaling ensures sufficient application capacity to meet varying levels of application load without requiring manual intervention.

Flexible

- Elastic Load Balancing also allows you to use IP addresses to route requests to application targets. This offers you flexibility in how you virtualize your application targets, allowing you to host more applications on the same instance. This also enables these applications to have individual security groups and use the same network port to further simplify inter-application communication in microservices based architecture.

Robust Monitoring and Auditing

- Elastic Load Balancing allows you to monitor your applications and their performance in real time with Amazon CloudWatch metrics, logging, and request tracing. This improves visibility into the behavior of your applications, uncovering issues and identifying performance bottlenecks in your application stack at the granularity of an individual request.

Hybrid Load Balancing

- Elastic Load Balancing offers ability to load balance across AWS and on-premises resources using the same load balancer. This makes it easy for you to migrate, burst, or failover on-premises applications to the cloud.

21. Define load balancing. What is need of load balancing in cloud computing? List network resources that can be load balanced.

Load balancing

In computing, load balancing improves the distribution of workloads across multiple computing resources, such as computers, a computer cluster, network links, central processing units, or disk drives.

Need of load balancing in cloud computing

(i) High Performing applications

- Cloud load balancing techniques, unlike their traditional on premise counterparts, are less expensive and simple to implement. Enterprises can make their client applications work faster and deliver better performances, that too at potentially lower costs.

(ii) Increased scalability

- Cloud balancing takes help of cloud's scalability and agility to maintain website traffic. By using efficient load balancers, you can easily match up the increased user traffic and distribute it among various servers or network devices. It is especially important for ecommerce websites, who deals with thousands of website visitors every second. During sale or other promotional offers they need such effective load balancers to distribute workloads.

(iii) Ability to handle sudden traffic spikes

- A normally running University site can completely go down during any result declaration. This is because too many requests can arrive at the same time. If they are using cloud load balancers, they do not need to worry about such traffic surges. No matter how large the request is, it can be wisely distributed among different servers for generating maximum results in less response time.

(iv) Business continuity with complete flexibility

- The basic objective of using a load balancer is to save or protect a website from sudden outages. When the workload is distributed among various servers or network units, even if one node fails the burden can be shifted to another active node.
- Thus, with increased redundancy, scalability and other features load balancing easily handles website or application traffic.

Network resources that can be load balanced

- Servers
- Routing mechanism

22. Explain Virtual Private cloud

- A **virtual private cloud (VPC)** is an on-demand configurable pool of shared computing resources allocated within a public cloud environment, providing a certain level of isolation between the different organizations.
- You already know that there are three major types of clouds: Public, Private and Hybrid. Now, there's a newer player in the game: Virtual Private Clouds.
- What makes these different from public and private clouds, and what is the benefit? Is it just a fancy name for public cloud, or is it a private one?
- VPC are related to the public cloud, but they are not the same. Instead of sharing resources and space in a public infrastructure, you get a changeable allotment of resources to configure.
- There is a certain level of isolation between you and other users, via a private IP subnet and virtual communication construct (such as a VLAN) on a per user basis.
- This ensures a secure method of remotely accessing your cloud resources. This isolation within a public cloud lends the name "virtual private" because you are essentially operating a private cloud within a public cloud.
- That also doesn't mean Virtual Private Clouds and private clouds are the same.
- Private clouds are entirely dedicated to your organization, and that includes the hardware.
- Virtual Private clouds do not have the same hardware dedication; it just creates a more secure environment on public infrastructure.
- Think of it as operating like a VPN: You use them to send messages over the public internet in a secure way as if you had your own personal network, but it's not the same as actually having your own.

- What's the benefit to this? Wouldn't it just be easier to have a private cloud? Not necessarily. Private clouds are expensive to operate, and because the hardware as well as the resources required to run it belong to you alone, there is no one to share that cost with.
- Virtual Private Clouds give you the best of both worlds: A private cloud for security and compliance purposes, reduced infrastructure costs that come with public clouds. The allotment of resources is yours to use, so there is no worry about running out or having to share with others. You simply are sharing the infrastructure.
- Virtual Private Clouds are commonly used with Infrastructure as a Service (IaaS) providers.
- Because the shared resources (CPU, RAM, etc.) are not always the responsibility of the hardware provider, it is possible to have different infrastructure and VPC providers.
- However, having the same VPC and infrastructure provider can help cut down on the confusion and communication process between you and your vendor.

23. Explain DynamoDB

- Amazon DynamoDB -- also known as Dynamo Database or DDB -- is a fully managed NoSQL database service provided by Amazon Web Services. DynamoDB is known for low latencies and scalability.
- According to AWS, DynamoDB makes it simple and cost-effective to store and retrieve any amount of data, as well as serve any level of request traffic.
- All data items are stored on solid-state drives, which provide high I/O performance and can more efficiently handle high-scale requests.
- An AWS user interacts with the service by using the AWS Management Console or a DynamoDB API.
- DynamoDB uses a NoSQL database model, which is nonrelational, allowing documents, graphs and columnar among its data models.
- A user stores data in DynamoDB tables, then interacts with it via GET and PUT queries, which are read and write operations, respectively.
- DynamoDB supports basic CRUD operations and conditional operations. Each DynamoDB query is executed by a primary key identified by the user, which uniquely identifies each item.

Scalability, availability and durability

- DynamoDB enforces replication across three availability zones for high availability, durability and read consistency.
- A user can also opt for cross-region replication, which creates a backup copy of a DynamoDB table in one or more global geographic locations.
- The DynamoDB scan API provides two consistency options when reading DynamoDB data:
  - Eventually consistent reads
  - Strongly consistent reads
- The former, which is the AWS default setting, maximizes throughput at the potential expense of not having a read reflect the latest write or update. The latter reflects all writes and updates.
- There are no DynamoDB limits on data storage per user, nor a maximum throughput per table.

Security

- Amazon DynamoDB offers Fine-Grained Access Control (FGAC) for an administrator to protect data in a table.
- The admin or table owner can specify who can access which items or attributes in a table and what actions that person can perform.

- FGAC is based on the AWS Identity and Access Management service, which manages credentials and permissions.
- As with other AWS products, the cloud provider recommends a policy of least privilege when granting access to items and attributes.
- An admin can view usage metrics for DynamoDB with Amazon CloudWatch.

Additional DynamoDB features

- The DynamoDB Triggers feature integrates with AWS Lambda to allow a developer to code actions based on updates to items in a DynamoDB table, such as sending a notification or connecting a table to another data source.
- The developer associates a Lambda function, which stores the logic code, with the stream on a DynamoDB table.
- AWS Lambda then reads updates to a table from a stream and executes the function.
- The DynamoDB Streams feature provides a 24-hour chronological sequence of updates to items in a table.
- An admin can access the stream via an API call to take action based on updates, such as synchronizing information with another data store. An admin enables DynamoDB Streams on a per-table basis.

24. Write short note on Relational Database Service.

- Amazon Relational Database Service (Amazon RDS) makes it easy to set up, operate, and scale a relational database in the cloud.
- It provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching and backups.
- It frees you to focus on your applications so you can give them the fast performance, high availability, security and compatibility they need.
- Amazon RDS is available on several database instance types optimized for memory, performance or I/O and provides you with six familiar database engines to choose from, including Amazon Aurora, PostgreSQL, MySQL, MariaDB, Oracle, and Microsoft SQL Server.
- You can use the AWS Database Migration Service to easily migrate or replicate your existing databases to Amazon RDS.

Advantages/Benefits

(i) Easy to Administer

Amazon RDS makes it easy to go from project conception to deployment. Use the AWS Management Console, the AWS RDS Command-Line Interface, or simple API calls to access the capabilities of a production-ready relational database in minutes. No need for infrastructure provisioning, and no need for installing and maintaining database software.

(ii) Highly Scalable

We can scale our database's compute and storage resources with only a few mouse clicks or an API call, often with no downtime. Many Amazon RDS engine types allow you to launch one or more Read Replicas to offload read traffic from your primary database instance.

(iii) Available and Durable

Amazon RDS runs on the same highly reliable infrastructure used by other Amazon Web Services. When you provision a Multi-AZ DB Instance, Amazon RDS synchronously replicates the data to a standby instance in a different Availability Zone (AZ). Amazon RDS has many other features that enhance reliability for critical production databases, including automated backups, database snapshots, and automatic host replacement.

(iv) Fast

Amazon RDS supports the most demanding database applications. You can choose between two SSD-backed storage options: one optimized for high-performance OLTP applications, and the other for cost-effective general-purpose use. In addition, Amazon Aurora provides performance on par with commercial databases at 1/10th the cost.

(v) Secure

Amazon RDS makes it easy to control network access to your database. Amazon RDS also lets you run your database instances in Amazon Virtual Private Cloud (Amazon VPC), which enables you to isolate your database instances and to connect to your existing IT infrastructure through an industry-standard encrypted IPsec VPN. Many Amazon RDS engine types offer encryption at rest and encryption in transit.

(vi) Inexpensive

You pay very low rates and only for the resources you actually consume. In addition, you benefit from the option of On-Demand pricing with no up-front or long-term commitments, or even lower hourly rates via Reserved Instance pricing.

25. Explain Redshift

- Perhaps one of the most exciting outcomes of the public cloud was addressing the shortcomings of traditional enterprise data warehouse (EDW) storage and processing. The fast provisioning, commodity costs, infinite scale, and pay-as-you-grow pricing of public cloud are a natural fit for EDW needs, providing even the smallest of users the ability to now get valuable answers to BI questions.
- **Amazon Redshift** is one such system built to address EDW needs, and it boasts low costs, an easy SQL-based access model, easy integration to other Amazon Web Services (AWS) solutions, and most importantly, high query performance.
- Amazon Redshift gets its name from the astronomical phenomenon noticed by Hubble, which explained the expansion of the universe. By adopting the Amazon Redshift moniker, AWS wanted to relay to customers that the service was built to handle the perpetual expansion of their data.

Amazon Redshift Architecture:

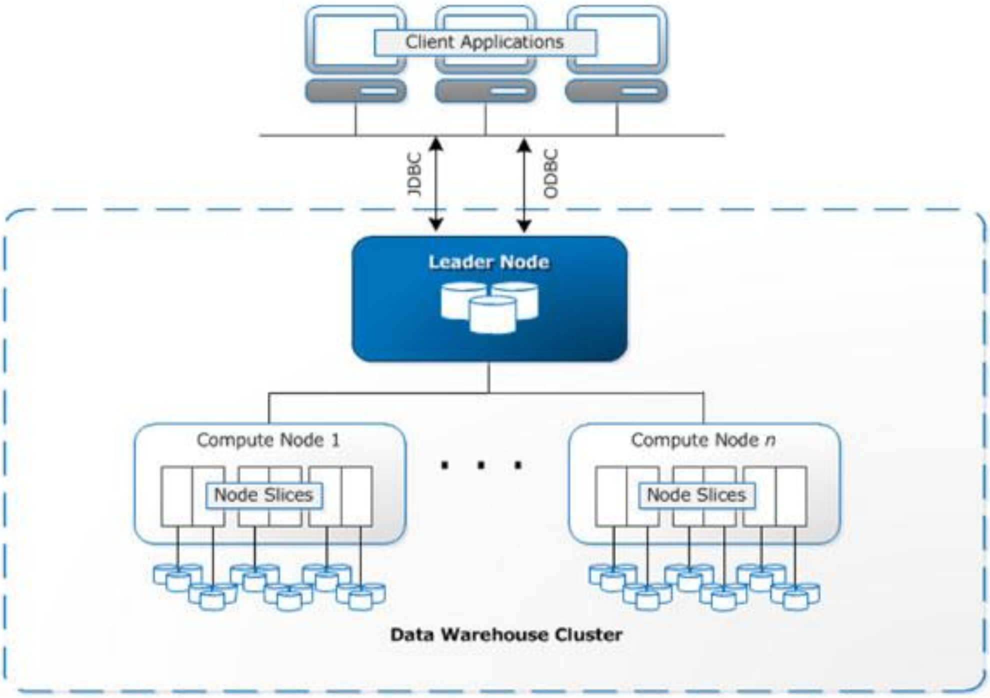

Fig. : Amazon Redshift Architecture

- An Amazon Redshift cluster consists of one leader node (which clients submit queries to) and one or more follower (or “compute”) nodes, which actually perform the queries on locally stored data.
- By allowing for unlimited expansion of follower nodes, Amazon Redshift ensures that customers can continue to grow their cluster as their data needs grow.
- Customers can start with a “cluster” as small as a single node (acting as both leader and follower), and for the smallest supported instance type (a DW2), that could be as low cost as $0.25/hour or about $180/month. By using “Reservations” (paying an up-front fee in exchange for a lower hourly running cost) for the underlying instances, Amazon Redshift can cost as little as $1,000/TB/year — upwards of one-fifth to one-tenth of the cost of a traditional EDW.
- Because Amazon Redshift provides native Open Database Connectivity (ODBC) and Database Connectivity (JDBC) connectivity (in addition to PostgresSQL driver support), most third-party BI tools (like Tableu, Qlikview, and MicroStrategy) work right out of the box. Amazon Redshift also uses the ubiquitous Structured Query Language (SQL) language for queries, ensuring that your current resources can quickly and easily become productive with the technology.
- Amazon Redshift was custom designed from the ParAccel engine — an analytic database which used columnar storage and parallel processing to achieve very fast I/O.
- Columns of data in Amazon Redshift are stored physically adjacent on disk, meaning that queries and scans on those columns (common in online analytical processing [OLAP] queries) run very fast.
- Additionally, Amazon Redshift uses 10GB Ethernet interconnects, and specialized EC2 instances (with between three and 24 spindles per node) to achieve high throughput and low latency.
- For even faster queries, Amazon Redshift allows customers to use column-level compression to both greatly reduce the amount of data that needs stored, and reduce the amount of disk I/O.
- Amazon Redshift, like many of AWS’s most popular services, is also fully managed, meaning that low-level, time-consuming administrative tasks like OS patching, backups, replacing failed hardware, and software upgrades are handled automatically and transparently.

- With Amazon Redshift, users simply provision a cluster, load it with their data, and begin executing queries. All data is continuously, incrementally, automatically backed up in the highly durable S3, and enabling disaster recovery across regions can be accomplished with just a few clicks.
- Spinning a cluster up can be as simple as a few mouse clicks, and as fast as a few minutes.
- A very exciting aspect of Amazon Redshift, and something that is not possible in traditional EDWs, is the ability to easily scale a provisioned cluster up and down.
- In Amazon Redshift, this scaling is transparent to the customer—when a resize is requested, data is copied in parallel from the source cluster (which continues to function in read-only mode) to a new cluster, and once all data is live migrated, DNS is flipped to the new cluster and the old cluster is de-provisioned.
- This allows customers to easily scale up and down, and each scaling event nicely re-stripes the data across the new cluster for a balanced workload.
- Amazon Redshift offers mature, native, and tunable security. Clusters can be deployed into a Virtual Private Cloud (VPC), and encryption of data is supported via hardware accelerated AES-256 (for data at rest) and SSL (for data on the wire).
- Compliance teams will be pleased to learn that users can manage their own encryption keys via AWS's Hardware Security Module (HSM) service, and that Amazon Redshift provides a full audit trail of all SQL connection attempts, queries, and modifications of the cluster.

26. Explain ElastiCache

- ElastiCache is a web service that makes it easy to set up, manage, and scale a distributed in-memory data store or cache environment in the cloud.
- It provides a high-performance, scalable, and cost-effective caching solution, while removing the complexity associated with deploying and managing a distributed cache environment.
- With ElastiCache, you can quickly deploy your cache environment, without having to provision hardware or install software.
- You can choose from Memcached or Redis protocol-compliant cache engine software, and let ElastiCache perform software upgrades and patch management for you.
- For enhanced security, ElastiCache can be run in the Amazon Virtual Private Cloud (Amazon VPC) environment, giving you complete control over network access to your clusters.
- With just a few clicks in the AWS Management Console, you can add or remove resources such as nodes, clusters, or read replicas to your ElastiCache environment to meet your business needs and application requirements.
- Existing applications that use Memcached or Redis can use ElastiCache with almost no modification.
- Your applications simply need to know the host names and port numbers of the ElastiCache nodes that you have deployed.
- The ElastiCache Auto Discovery feature for Memcached lets your applications identify all of the nodes in a cache cluster and connect to them, rather than having to maintain a list of available host names and port numbers.
- In this way, your applications are effectively insulated from changes to node membership in a cluster.
- ElastiCache has multiple features to enhance reliability for critical production deployments:
  - Automatic detection and recovery from cache node failures.
  - Multi-AZ with Automatic Failover of a failed primary cluster to a read replica in Redis clusters that support replication (called replication groups in the ElastiCache API and AWS CLI.
  - Flexible Availability Zone placement of nodes and clusters.

- Integration with other AWS services such as Amazon EC2, Amazon CloudWatch, AWS CloudTrail, and Amazon SNS to provide a secure, high-performance, managed in-memory caching solution.

27. Write a short note on High performance AWS Networking.

- High performance AWS Networking is nothing but use of various network services provided by AWS for better performance.
- AWS Networking include following services:
  1. Private DNS Servers
     - The Private DNS are name servers that reflect your domain name rather than our default ones.
     - Having private nameservers could be useful if you intend to resell hosting services or want to brand your business.
     - Also, when using Private DNS, if a domain name is migrated to another server, there is no need to change any nameservers and the domain names will automatically point to the new location.
  2. Virtual Private Clouds (Explain Earlier),
  3. Cloud Models (Explain Earlier), etc.

28. What is cloud watch and cloud formation?

Cloud watch

- Amazon CloudWatch is a monitoring service for AWS cloud resources and the applications you run on AWS.
- You can use Amazon CloudWatch to collect and track metrics, collect and monitor log files, set alarms, and automatically react to changes in your AWS resources.
- Amazon CloudWatch can monitor AWS resources such as Amazon EC2 instances, Amazon DynamoDB tables, and Amazon RDS DB instances, as well as custom metrics generated by your applications and services, and any log files your applications generate.
- You can use Amazon CloudWatch to gain system-wide visibility into resource utilization, application performance, and operational health.
- You can use these insights to react and keep your application running smoothly.

Cloud formation

- AWS CloudFormation provides a common language for you to describe and provision all the infrastructure resources in your cloud environment.
- CloudFormation allows you to use a simple text file to model and provision, in an automated and secure manner, all the resources needed for your applications across all regions and accounts.
- This file serves as the single source of truth for your cloud environment.
- AWS CloudFormation is available at no additional charge, and you pay only for the AWS resources needed to run your applications.

Advantage of Cloud Formation

MODEL IT ALL

- AWS CloudFormation allows you to model your entire infrastructure in a text file. This template becomes the single source of truth for your infrastructure. This helps you to standardize infrastructure components used across your organization, enabling configuration compliance and faster troubleshooting.

AUTOMATE AND DEPLOY

- AWS CloudFormation provisions your resources in a safe, repeatable manner, allowing you to build and rebuild your infrastructure and applications, without having to perform manual actions or write custom scripts. CloudFormation takes care of determining the right operations to perform when managing your stack, and rolls back changes automatically if errors are detected.

IT'S JUST CODE

- Codifying your infrastructure allows you to treat your infrastructure as just code. You can author it with any code editor, check it into a version control system, and review the files with team members before deploying into production.

29. How AWS deals with Disaster recovery?

Disaster Recovery (DR)

- Our data is the most precious asset that we have and protecting it is our top priority.
- Creating backups of our data to an off shore data center, so that in the event of an on premise failure we can switch over to our backup, is a prime focus for business continuity.
- As AWS says, 'Disaster recovery is a continual process of analysis and improvement, as business and systems evolve. For each business service, customers need to establish an acceptable recovery point and time, and then build an appropriate DR solution.'
- Backup and DR on Cloud reduces costs by half as compared to maintaining your own redundant data centers. And if you think about it, it's really not that surprising.
- Imagine the kind of cost you would entail in buying and maintaining servers and data centers, providing secure and stable connectivity and not to mention keeping them secure.
- You would also be underutilizing severs; and in times of unpredictable traffic rise it would be strenuous to set up new ones. To all these cloud provides a seamless transition reducing cost dramatically.

4 Standard Approaches of Backup and Disaster Recovery Using Amazon Cloud

1. Backup and Recovery

- To recover your data in the event of any disaster, you must first have your data periodically backed up from your system to AWS.
- Backing up of data can be done through various mechanisms and your choice will be based on the RPO (Recovery Point Objective- So if your disaster struck at 2 pm and your RPO is 1 hr, your Backup & DR will restore all data till 1 pm.) that will suit your business needs.
- AWS offers AWS Direct connect and Import Export services that allow for faster backup.
- For example, if you have a frequently changing database like say a stock market, then you will need a very high RPO. However if your data is mostly static with a low frequency of changes, you can opt for periodic incremental backup.
- Once your backup mechanisms are activated you can pre-configure AMIs (operating systems & application software).
- Now when a disaster strikes, EC2 (Elastic Compute Capacity) instances in the Cloud using EBS (Elastic Block Store) coupled with AMIs can access your data from the S3 (Simple Storage Service) buckets to revive your system and keep it going.

2. Pilot Light Approach

- The name pilot light comes from the gas heater analogy. Just as in a heater you have a small flame that is always on, and can quickly ignite the entire furnace; a similar approach can be thought of about your data system.
- In the preparatory phase your on premise database server mirrors data to data volumes on AWS. The database server on cloud is always activated for frequent or continuous incremental backup.
- This core area is the pilot from our gas heater analogy. The application and caching server replica environments are created on cloud and kept in standby mode as very few changes take place over time.
- These AMIs can be updated periodically. This is the entire furnace from our example. If the on premise system fails, then the application and caching servers get activated; further users are rerouted using elastic IP addresses to the ad hoc environment on cloud. Your Recovery takes just a few minutes.

3. Warm Standby Approach

- This Technique is the next level of the pilot light, reducing recovery time to almost zero.
- Your application and caching servers are set up and always activated based on your business critical activities but only a minimum sized fleet of EC2 instances are dedicated.
- The backup system is not capable of handling production load, but can be used for testing, quality assurance and other internal uses.
- In the event of a disaster, when your on premise data center fails, two things happen.
- Firstly multiple EC2 instances are dedicated (vertical and horizontal scaling) to bring your application and caching environment up to production load. ELB and Auto Scaling (for distributing traffic) are used to ease scaling up.
- Secondly using Amazon Route 53 user traffic is rerouted instantly using elastic IP addresses and there is instant recovery of your system with almost zero down time.

4. Multi-Site Approach

- Well this is the optimum technique in backup and DR and is the next step after warm standby.
- All activities in the preparatory stage are similar to a warm standby; except that AWS backup on Cloud is also used to handle some portions of the user traffic using Route 53.
- When a disaster strikes, the rest of the traffic that was pointing to the on premise servers are rerouted to AWS and using auto scaling techniques multiple EC2 instances are deployed to handle full production capacity.
- You can further increase the availability of your multi-site solution by designing Multi-AZ architectures.

30. Definitions

1. Cloud

- The cloud is a term referring to accessing computer, information technology (IT), and software applications through a network connection, often by accessing data centers using wide area networking (WAN) or Internet connectivity.

2. On Demand self service

- On-demand self-service refers to the service provided by cloud computing vendors that enables the provision of cloud resources on demand whenever they are required. In on-demand self-service, the user accesses cloud services through an online control panel.

3. Resource pooling

- Resource pooling is an IT term used in cloud computing environments to describe a situation in which providers serve multiple clients, customers or "tenants" with provisional and scalable services.

4. Broad Network access

- Broad network access refers to resources hosted in a private cloud network (operated within a company's firewall) that are available for access from a wide range of devices, such as tablets, PCs, Macs and smartphones.

5. Rapid elasticity

- Rapid elasticity is a cloud computing term for scalable provisioning, or the ability to provide scalable services. Experts point to this kind of scalable model as one of five fundamental aspects of cloud computing.

6. Measured services

- This is a reference to services where the cloud provider measures or monitors the provision of services for various reasons, including billing, effective use of resources, or overall predictive planning.

References:

1. *www.aws.amazon.com*
2. *www.docs.aws.amazon.com*
3. *www.bluepiit.com*
4. *www.inforisktoday.com*
5. *www.techno-pulse.com*
6. *www.exelanz.com*
7. *www.ibm.com*
8. *www.iarjset.com/upload/2017/july-17/IARJSET%2018.pdf*
9. *www.searchservervirtualization.techtarget.com*
10. *www.docs.eucalyptus.com*
11. *www.cloudacademy.com*
12. *www.searchaws.techtarget.com*
13. *www.searchsecurity.techtarget.com*
14. *www.en.wikipedia.org/wiki/Cloud_computing_security*
15. *www.znetlive.com*
16. *www.en.wikipedia.org/wiki/Virtual_private_cloud*
17. *www.resource.onlinetech.com*
18. *www.globalknowledge.com*
19. *www.blog.blazeclan.com/4-approaches-backup-disaster-recovery-explained-amazon-cloud/*