1. **AAA**: authorization, authentication, accounting

2. **CIA**: confidentiality, integrity, availability

3. **Network protocol**

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

自治系统内选路协议:

Rip: routing information protocol, is one of the oldest distance-vector routing protocols which employ the hop count as a routing metric. RIP prevents routing loops by implementing a limit on the number of hops allowed in a path from source to destination. The largest number of hops allowed for RIP is 15, which limits the size of networks that RIP can support.使用udp封装

Ospf: Open Shortest Path First (OSPF) is a routing protocol for Internet Protocol (IP) networks. It uses a link state routing (LSR) algorithm and falls into the group of interior gateway protocols (IGPs), operating within a single autonomous system (AS). It is defined as OSPF Version 2 for IPv4.[[1]](https://en.wikipedia.org/wiki/Open_Shortest_Path_First#cite_note-rfc2328-1) The updates for IPv6 are specified as OSPF Version 3. OSPF supports the Classless Inter-Domain Routing (CIDR) addressing model.

OSPF is a widely used IGP in large enterprise networks. IS-IS, another LSR-based protocol, is more common in large service provider networks. 使用IP封装

Isis: IS-IS is an interior gateway protocol, designed for use within an administrative domain or network. IS-IS is a link-state routing protocol, operating by reliably flooding link state information throughout a network of routers. Each IS-IS router independently builds a database of the network's topology, aggregating the flooded network information.

自治系统间选路协议**:**

Bgp: Border Gateway Protocol (BGP) is a standardized exterior gateway protocol designed to exchange routing and reachability information among autonomous systems (AS) on the Internet.[1] The protocol is classified as a path vector protocol.[2] The Border Gateway Protocol makes routing decisions based on paths, network policies, or rule-sets configured by a network administrator and is involved in making core routing decisions.

BGP may be used for routing within an autonomous system. In this application it is referred to as Interior Border Gateway Protocol, Internal BGP, or iBGP. In contrast, the Internet application of the protocol may be referred to as Exterior Border Gateway Protocol, External BGP, or eBGP. 使用tcp封装

Icmp: The Internet Control Message Protocol (ICMP) is a supporting protocol in the Internet protocol suite. It is used by network devices, including routers, to send error messages and operational information indicating success or failure when communicating with another IP address.

4**. S3 vs DynamoDB**

The basic difference between S3 and DynamoDB is that S3 is file storage whereas DynamoDB is a Database.

**S3**: S3 is a file storage and is well suited if you want to store unstructured data. S3 doesnt follow a folder structure and uses everything as an object. In place of using a hierarchical directory, S3 stores files in a flat organisation of containers called Buckets. S3 uses unique Ids called Keys to retrieve files from the bucket.

**DynamoDB**: DynamoDB is a NoSQL database which is built for high throughput and low latency. DynamoDB best fits for you when you need to deal with (semi) structured data. DynamoDb is used to store key-value. It uses items and attributes for its tables. Each item contains different number of attributes.

A screenshot of a cell phone

Description automatically generated

5. **sqs vs lambda**

Amazon SQS: Amazon Simple Queue Service (SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications.

Amazon SQS offers a reliable, highly-scalable hosted queue for storing messages as they travel between applications or microservices. It moves data between distributed application components and helps you decouple these components. Amazon SQS provides common middleware constructs such as dead-letter queues and poison-pill management. It also provides a generic web services API and can be accessed by any programming language that the AWS SDK supports. Amazon SQS supports both standard and FIFO queues.

Amazon lambda:

AWS Lambda is a serverless compute service that runs your code in response to events and automatically manages the underlying compute resources for you. You can use AWS Lambda to extend other AWS services with custom logic, or create your own back-end services that operate at AWS scale, performance, and security. AWS Lambda can automatically run code in response to multiple events, such as HTTP requests via Amazon API Gateway, modifications to objects in Amazon S3 buckets, table updates in Amazon DynamoDB, and state transitions in AWS Step Functions.

Lambda runs your code on high-availability compute infrastructure and performs all the administration of the compute resources, including server and operating system maintenance, capacity provisioning and automatic scaling, code and security patch deployment, and code monitoring and logging. All you need to do is supply the code.

6. **why nosql**

NoSQL databases are a great fit for many modern applications such as mobile, web, and gaming that require flexible, scalable, high-performance, and highly functional databases to provide great user experiences.

* Flexibility: NoSQL databases generally provide flexible schemas that enable faster and more iterative development. The flexible data model makes NoSQL databases ideal for semi-structured and unstructured data.
* Scalability: NoSQL databases are generally designed to scale out by using distributed clusters of hardware instead of scaling up by adding expensive and robust servers. Some cloud providers handle these operations behind-the-scenes as a fully managed service.
* High-performance: NoSQL database are optimized for specific data models and access patterns that enable higher performance than trying to accomplish similar functionality with relational databases.
* Highly functional: NoSQL databases provide highly functional APIs and data types that are purpose built for each of their respective data models.

7. **what is map and reduce and how MapReduce works**

Hadoop divides the job into tasks. There are two types of tasks:

1. **Map tasks** (Splits & Mapping)
2. **Reduce tasks** (Shuffling, Reducing)

**Input Splits:**

An input to a MapReduce job is divided into fixed-size pieces called **input splits**Input split is a chunk of the input that is consumed by a single map

**Mapping:**

This is the very first phase in the execution of map-reduce program. In this phase data in each split is passed to a mapping function to produce output values. In our example, a job of mapping phase is to count a number of occurrences of each word from input splits (more details about input-split is given below) and prepare a list in the form of <word, frequency>

**Shuffling:**

This phase consumes the output of Mapping phase. Its task is to consolidate the relevant records from Mapping phase output. In our example, the same words are clubed together along with their respective frequency.

**Reducing:**

In this phase, output values from the Shuffling phase are aggregated. This phase combines values from Shuffling phase and returns a single output value. In short, this phase summarizes the complete dataset.

The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes. Under the MapReduce model, the data processing primitives are called mappers and reducers. Decomposing a data processing application into *mappers* and *reducers* is sometimes nontrivial. But, once we write an application in the MapReduce form, scaling the application to run over hundreds, thousands, or even tens of thousands of machines in a cluster is merely a configuration change. This simple scalability is what has attracted many programmers to use the MapReduce model.

7. **broadcast(各种cast)**

Data is transported over a network by three simple methods i.e. Unicast, Broadcast, and Multicast.

Unicast: traffic, many streams of IP packets that move across networks flow from a single point, such as a website server, to a single endpoint such as a client PC. This is the most common form of information transference on networks.

Broadcast: Here, traffic streams from a single point to all possible endpoints within reach on the network, which is generally a LAN. This is the easiest technique to ensure traffic reaches to its destinations.

Multicast: multicast is a “one source to many destinations” way of traffic distribution, means that only the destinations that openly point to their requisite to accept the data from a specific source to receive the traffic stream.

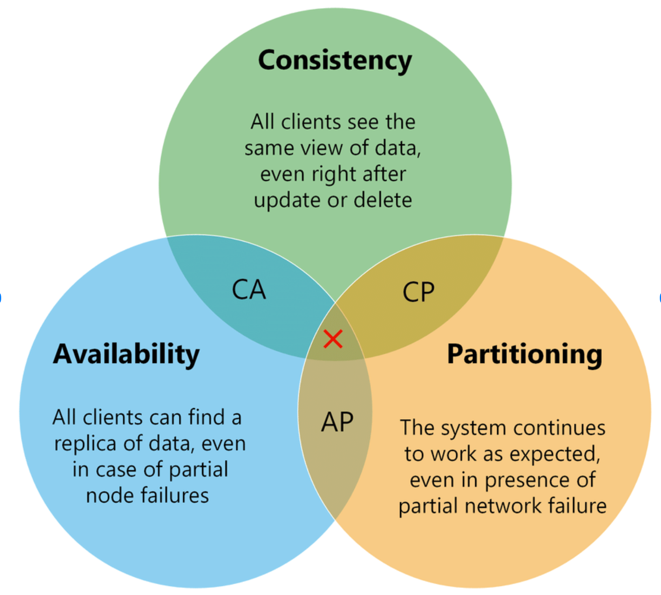
8. **cap theorem**

the CAP theorem, states that it is impossible for a distributed data store to simultaneously provide more than two out of the following three guarantees:

Consistency: Every read receives the most recent write or an error

Availability: Every request receives a (non-error) response, without the guarantee that it contains the most recent write

Partition tolerance: The system continues to operate despite an arbitrary number of messages being dropped (or delayed) by the network between nodes



9. **各种storage topics**

Object storage: Objects are discrete units of data that are stored in a structurally flat data environment. There are no folders, directories, or complex hierarchies as in a file-based system. Each object is a simple, self-contained repository that includes the data, metadata (descriptive information associated with an object), and a unique identifying ID number (instead of a file name and file path). This information enables an application to locate and access the object. Object storage removes the complexity and scalability challenges of a hierarchical file system with folders and directories. Objects can be stored locally, but most often reside on cloud servers, with accessibility from anywhere in the world. Objects (data) in an object-storage system are accessed via Application Programming Interfaces (APIs).

File storage stores data as a single piece of information in a folder to help organize it among other data. This is also called hierarchical storage, imitating the way that paper files are stored. When you need access to data, your computer system needs to know the path to find it.

Block storage takes a file apart into singular blocks of data and then stores these blocks as separate pieces of data. Each piece of data has a different address, so they don't need to be stored in a file structure.

10. **Types of firewalls**

Let’s quickly discuss the three basic types of network firewalls: packet filtering (stateless), stateful, and application layer.

Packet filtering, or stateless, firewalls work by inspecting individual packets in isolation. As such, they are unaware of connection state and can only allow or deny packets based on individual packet headers.

Stateful firewalls are able to determine the connection state of packets, which makes them much more flexible than stateless firewalls. They work by collecting related packets until the connection state can be determined before any firewall rules are applied to the traffic.

Application firewalls go one step further by analyzing the data being transmitted, which allows network traffic to be matched against firewall rules that are specific to individual services or applications. These are also known as proxy-based firewalls.

11. **firewall rules**

network traffic that traverses a firewall is matched against rules to determine if it should be allowed through or not.

Firewall rules layer:

Layer 3 Firewalls (Network Firewalls)

One way is to categorize traffic according to IP addresses, port numbers and service protocols. In other words, you could tell your firewall to accept traffic from certain IP addresses while blocking all other traffic (this would constitute a whitelisting strategy). Alternatively, you could blacklist IP addresses that you know to be sources of abuse.

You could make things more granular by configuring your firewall to accept traffic from certain IP addresses only on certain ports, or when the traffic uses a certain protocol.

If you categorize traffic in these ways, you’re operating on layer 3 of your firewall. This is also sometimes known as the network layer. Layer 3 firewalls filter traffic based on the TCP/IP stack. This approach is sometimes also referred to as packet filtering, because you’re essentially allowing and blocking individual network packets depending on where they originated and which ports they want to talk to.

Layer 7 Firewalls (Application Firewalls)

The other common approach to firewall configuration involves layer 7, which is also known as the application layer.

Layer 7 lets you sort traffic according to which application or application service the traffic is trying to reach, and what the specific contents of that traffic are. Rather than simply blocking all traffic on a certain port, you could use an application firewall to accept traffic on that port in general, but block any traffic that contains a known vulnerability (like an SQL injection attack or a malicious telnet command).

12. **network security(protocols)**

TCP/IP Protocol: The TCP/IP protocol is the fundamental communication protocol of the Internet and it consists of two protocols, the TCP, and the IP. The TCP/IP was developed in the Unix operating system in 1989. The purpose was to have computers communicate in a simple way and transmit information across the network. For example, a user interacts with an application and TCP is responsible for sending the message to the destination host. After the TCP sends the message it returns an IP header which includes the location and the IP address destination.

HTTP Protocol: The HTTP (Hypertext Transfer Protocol) protocol is based on the www (World Wide Web) which transmits messages around the network. For example, when a user is entering the browser and enters in the URL a search, the URL transmits the messages by HTTP to the web server that the user requested. Then the web server responds and delivers the results of the search criteria you had requested.

UDP Protocol: The UDP (User Datagram Protocol) protocol is an additional protocol to TCP which focuses on low-latency and loss connections between applications on the internet. UDP offers port numbers to differentiate what the user requests and offers checksum to count the number of bits in a transmission and to validate that the data has been completely delivered. UDP is part of the OSI model in the transport layer. One of the advantages of this is that it does not require a connection to send data and THE UDP header is smaller than TCP header.

TCP Protocol: The TCP (Transmission Control Protocol) protocol is a standard protocol which describes how to manage a network communication using applications that can transmit data. It also communicates to other networks using Internet protocol. It requires a connection to transmit messages.  The TCP works with IP which transmits packets of data to each other. TCP is a network device used for data transmission from the network layer. In the OSI model, TCP belongs to a portion of the transport layer and session layer. The purpose is to make it easy for computers to communicate with each other and send information across the internet.

FTP Protocol: The FTP (File Transfer Protocol) protocol is usually used to transfer files over the Internet. FTP uses a client-server to share files by a remote computer. The way the FTP works is like HTTP for sending web pages from a server back to the user. Files can be transmitted by two computers using an FTP software. One computer would be the local host and the other computer would be the remote host also using the same FTP software. There are three steps to transmit the files: first, the local host computer connects to the remote host using their IP address; second, the user must authenticate by entering a username and a password. Third, The FTP software often has a graphic user interface (GUI) to grant the users to drag and drop the files between each other.

SSH Protocol: The SSH (Secure Socket Shell) protocol focuses on network operations and provides a secure way to access via remote computer. SSH provides authentication and encryption between two computers connecting to the Internet. SSH is well used by network administrations to manage systems by remote access. It also uses a client-server just like FTP, but instead of transmitting files IT connects into a secure application.

DNS Protocol: DNS (Domain Name System) protocol maintains a directory of domain names translated into IP addresses. The DNS tracks the user to locate the web address to the corresponding IP address. For example, if a user types in the URL google.com, the web server is not reading the name google.com is reading the NUMERICAL IP address that corresponds to google.com which in this case is 208.65.155.84.

TLS Protocol: TLS (Transport Layer Security) protocol is based on SSH but focuses on data integrity and secrecy. TLS is widely used for web browsers that ensure a secure transmission by a network, such as voice over IP, VPN connections, etc. TLS provides secure communications between servers and browsers on websites. Using TLS can provide secure connections between a client and a server using symmetric encryption which provides secrecy. Integrity is ensured by checking each message being transmitted using a message authentication code.

DHCP Protocol: DHCP (Dynamic Host Configuration Protocol) is a network protocol which provides IP host that contains IP addresses, AND A default gateway and subnet mask. DHCP is being used by network administrations to manage network configuration to make sure each user requesting information to the server using its IP address would respond back to the corresponding IP address. For example, if the DHCP server is not responding then each computer must be manually assigned to an IP address.

MAC Protocol: The MAC (Media Access Control) protocol belongs to the data link layer which is layer 2. It focuses on systems that share communications by local networks. MAC is being assigned to a network interface card to move data packets from one device to another by a shared channel. The MAC protocol is also known as MAC address that contains 12-digit hexadecimal numbers which are part of the internet protocol.

13. **vpc**

VPC主要是一个网络层面的功能，其目的是让我们可以在云平台上构建出一个隔离的、自己能够管理配置和策略的虚拟网络环境，从而进一步提升我们在AWS环境中的资源的安全性。

Ability to setup connectivity intra- or inter- datacenter to allow disaggregated group of infrastructure/services to be perceived as local for better performance or manageability and security

14. **private ip address**

Private internal addresses are not routed on the Internet and no traffic cannot be sent to them from the Internet, they only supposed to work within the local network.

15. **vm vs container**

a virtual machine is software that emulates a computer system. This enables teams to run what appear to be multiple machines on a single computer. If you need to run software on a different type of hardware or operating system, a VM provides that option without using additional hardware.

Containers are a lighter-weight, more agile way of handling virtualization. Rather than spinning up an entire virtual machine, a container packages together everything needed to run a small piece of software. The container includes all the code, its dependencies and even the operating system itself. This enables applications to run almost anywhere.

there are 2 major differences between a Virtual Machine (VM) & a Container:

1. A Container’s system requires an underlying operating system that provides the basic services to all of the containerized applications using virtual-memory support for isolation. VMs, on the other hand, have their own operating system using hardware VM support by the Hypervisor.
2. A Container provides an abstract OS, while A VM provides an abstract machine that uses device drivers targeting the abstract machine.

A screenshot of a cell phone

Description automatically generated

16. **what is big data**

A screenshot of a cell phone

Description automatically generated

17. **cgroups**

cgroups (abbreviated from control groups) is a Linux kernel feature that limits, accounts for, and isolates the resource usage (CPU, memory, disk I/O, network, etc.) of a collection of processes.

18. **MFA**

 AWS Multi-Factor Authentication (MFA) is a simple best practice that adds an extra layer of protection on top of your user name and password

**MORE:**

# **Raft**:

raft是工程上使用较为广泛的强一致性、去中心化、高可用的分布式协议。

raft是一个共识算法（consensus algorithm），所谓共识，就是多个节点对某个事情达成一致的看法，raft协议就是一种leader-based的共识算法，与之相应的是leaderless的共识算法。

Raft implements consensus by first electing a distinguished leader, then giving the leader complete responsibility for managing the replicated log. The leader accepts log entries from clients, replicates them on other servers, and tells servers when it is safe to apply log entries to their state machines. A leader can fail or become disconnected from the other servers, in which case a new leader is elected.

上面的引文对raft协议的工作原理进行了高度的概括：raft会先选举出leader，leader完全负责replicated log的管理。leader负责接受所有客户端更新请求，然后复制到follower节点，并在“安全”的时候执行这些请求。如果leader故障，followes会重新选举出新的leader。

raft协议中，一个节点任一时刻处于以下三个状态之一：

* leader
* follower
* candidate

# **Paxos**:

A screenshot of a social media post

Description automatically generated

Paxos is an algorithm that is used to achieve consensus among a distributed set of computers that communicate via an asynchronous network. One or more clients proposes a value to Paxos and we have consensus when a majority of systems running Paxos agrees on one of the proposed values.

# **Eventual consistency**

Eventual consistency is a consistency model used in distributed computing to achieve high availability that informally guarantees that, if no new updates are made to a given data item, eventually all accesses to that item will return the last updated value.

# **CQRS**

Command Query Responsibility Segregation (CQRS) is an architectural pattern that separates reading and writing into two different models. This means that every method should either be a Command that performs an action or a Query that returns data. A Command cannot return data and a Query cannot change the data.

# **Actor Model**

The actor model in computer science is a mathematical model of concurrent computation that treats actor as the universal primitive of concurrent computation. In response to a message it receives, an actor can: make local decisions, create more actors, send more messages, and determine how to respond to the next message received. Actors may modify their own private state, but can only affect each other indirectly through messaging (obviating lock-based synchronization).

# **Non-repudiation**

refers to a situation where a statement's author cannot successfully dispute its authorship or the validity of an associated contract.

# **Mandantory access controll (mac)**

In computer security, mandatory access control (MAC) refers to a type of access control by which the operating system constrains the ability of a subject or initiator to access or generally perform some sort of operation on an object or target.

# **FROM SLIDES 1**

1. Cloud architecture overview -- different aspects of cloud and its computing architecture

Could architecture: computing, software, network, data, service

***data***:

1. Storage: fast/local storage, scale out storage as a service, traditional storage access(smb, nfs)
2. Database: relational database, managed software service(mostly RDBMS), software as a service

compute: vm, container, platform services, functions/serverless

***network***:

VPC: VPC主要是一个网络层面的功能，其目的是让我们可以在云平台上构建出一个隔离的、自己能够管理配置和策略的虚拟网络环境，从而进一步提升我们在AWS环境中的资源的安全性。

Ability to setup connectivity intra- or inter- datacenter to allow disaggregated group of infrastructure/services to be perceived as local for better performance or manageability and security

load balancing: Provide balancing mechanism to application or web service in a specific region or across multiple regions

在分布式系统中，负载均衡（Load Balancing）是一种将任务分派到多个服务端进程的方法

DNS: Public and private name resolution for public or automation in software services (e.g. my-service.neu-coe.com -> 52.95.245.77)

DNS（Domain Name Server，域名服务器）是进行域名(domain name)和与之相对应的IP地址 (IP address)转换的服务器。

***security***:

Identity and access management: Cloud Native Authentication and Authorization Service For Managing User, Objects and Services

Web service authorization: Ability to leverage latest web application security standards for repeatable, scalable user authentication (OAuth, SAML, et al)

Secrets: Ability for OS, App, or Cloud Services to Leverage Cryptographically Private Metadata

Key management: Cloud or Customer Managed Certificates for Cloud Service, Application and API-driven functionality

More: Intrusion Detection Systems, SSO Providers, Data Loss Prevention, Web Application Firewalls et al

OSI MODEL:

A screenshot of a cell phone

Description automatically generated

Layer 2 vs layer 3 communication:

Layer 2 - frames, broadcast domain (use physical addressing) -- Leverages MAC (physical) address

Layer 3 - packets, routing domains (use logical addressing) -- ARP resolves broadcasts for layer 3, Broadcast domains are limited by layer 3

Routed vs routing protocols:

Layer 3 routing: interior gateway protocols, designed for single datacenter, OSPF, RIP, EIGRP, exterior gateway protocols, designed to work across datacenters, EGP, BGP

Layer 2 routing: IS-IS, routed protocols: IP, IPv6, et al

Cloud DNS

IP Firewalls

Key infrastructure services

1. TCP/IP model -- the model for which software communicates and a way to review the world
2. Networking -- key network protocols used in cloud computing environment