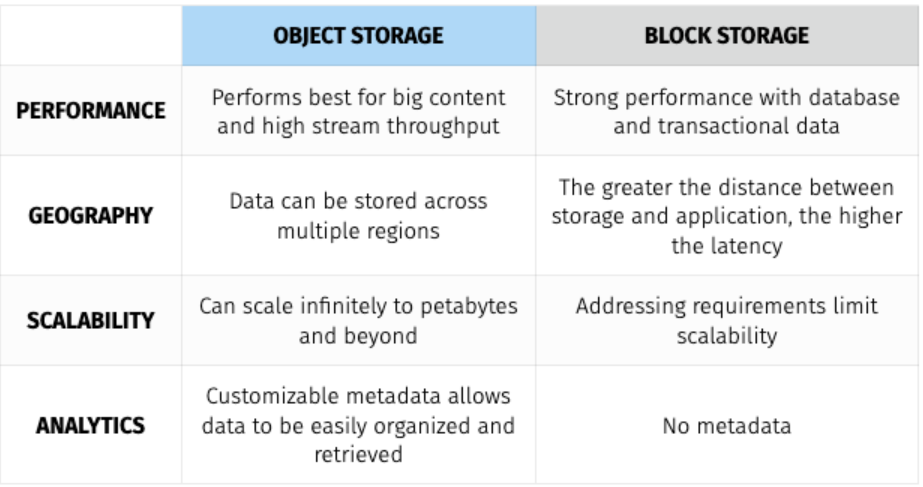
**IAM Entities:**

* **Users** - any individual end user such as an employee, system architect, CTO, etc.
* **Groups** - any collection of similar people with shared permissions such as system administrators, HR employees, finance teams, etc. Each user within their specified group will inherit the permissions set for the group.
* **Roles** - any software service that needs to be granted permissions to do its job, e.g- AWS Lambda needing write permissions to S3 or a fleet of EC2 instances needing read permissions from a RDS MySQL database.
* **Policies** - the documented rule sets that are applied to grant or limit access. In order for users, groups, or roles to properly set permissions, they use policies. Policies are written in JSON and you can either use custom policies for your specific needs or use the default policies set by AW

**Priority Levels in IAM**

* **Explicit Deny**: Denies access to a particular resource and this ruling cannot be overruled.
* **Explicit Allow**: Allows access to a particular resource so long as there is not an associated Explicit Deny.
* **Default Deny (or Implicit Deny)**: IAM identities start off with no resource access. Access instead must be granted.

**S3**

****

Data uploaded into S3 is spread across multiple files and facilities. The files uploaded into S3 have an upper-bound of 5TB per file and the number of files that can be uploaded is virtually limitless. S3 buckets, which contain all files, are named in a universal namespace so uniqueness is required. All successful uploads will return an **HTTP 200 response**.

* Objects (regular files or directories) are stored in S3 with a key, value, version ID, and metadata. They can also contain torrents and sub resources for access control lists which are basically permissions for the object itself.
* The data consistency model for S3 ensures immediate read access for new objects after the initial PUT requests. These new objects are introduced into AWS for the first time and thus do not need to be updated anywhere so they are available immediately.
* The data consistency model for S3 also ensures immediate read access for PUTS and DELETES of already existing objects.
* Amazon guarantees 99.999999999% (or 11 9s) durability for all S3 storage classes except its Reduced Redundancy Storage class.

**S3 Encryption:**

* **S3 Managed Keys / SSE - S3 (server side encryption S3 ) -** when Amazon manages the encryption and decryption keys for you automatically. In this scenario, you concede a little control to Amazon in exchange for ease of use.
* **AWS Key Management Service / SSE - KMS -** when Amazon and you both manage the encryption and decryption keys together.
* **Server Side Encryption w/ customer provided keys / SSE - C -** when I give Amazon my own keys that I manage. In this scenario, you concede ease of use in exchange for more control.

**Storage Gateway Simplified:**

Storage Gateway is a service that connects on-premise environments with cloud-based storage in order to seamlessly and securely integrate an on-prem application with a cloud storage backend. and Volume Gateway as a way of storing virtual hard disk drives in the cloud.

* The Storage Gateway service can either be a physical device or a VM image downloaded onto a host in an on-prem data center. It acts as a bridge to send or receive data from AWS.
* Storage Gateway can sit on top of VMWare's ESXi hypervisor for Linux machines and Microsoft’s Hyper-V hypervisor for Windows machines.
* The three types of Storage Gateways are below:
  + **File Gateway -** Operates via NFS or SMB and is used to store files in S3 over a network filesystem mount point in the supplied virtual machine. Simply put, you can think of a File Gateway as a file system mount on S3.
  + **Volume Gateway -** Operates via iSCSI and is used to store copies of hard disk drives or virtual hard disk drives in S3. These can be achieved via Stored Volumes or Cached Volumes. Simply put, you can think of Volume Gateway as a way of storing virtual hard disk drives in the cloud.
  + **Tape Gateway -** Operates as a Virtual Tape Library

**Stored Volumes vs. Cached Volumes:**

* Volume Gateway's **Stored Volumes** let you store data locally on-prem and backs the data up to AWS as a secondary data source. Stored Volumes allow low-latency access to entire datasets, while providing high availability over a hybrid cloud solution. Further, you can mount Stored Volumes on application infrastructure as iSCSI drives so when data is written to these volumes, the data is both written onto the on-prem hardware and asynchronously backed up as snapshots in AWS EBS or S3.
* Volume Gateway's **Cached Volumes** differ as they do not store the entire dataset locally like Stored Volumes. Instead, AWS is used as the primary data source and the local hardware is used as a caching layer. Only the most frequently used components are retained onto the on-prem infrastructure while the remaining data is served from AWS. This minimizes the need to scale on-prem infrastructure while still maintaining low-latency access to the most referenced data

**EC2 Placement Groups:**

* Clustered Placement Groups
* Spread Placement Groups
* Partitioned Placement Groups

**Elastic Network Interfaces (ENI)**

An elastic network interface is a networking component that represents a virtual network card. When you provision a new instance, there will be an ENI attached automatically and you can create and configure additional network interfaces if desired. When you move a network interface from one instance to another, network traffic is redirected to the new instance.

**Web Application Firewall (WAF)**AWS WAF is a web application that lets you allow or block the HTTP(s) requests that are bound for CloudFront, API Gateway, Application Load Balancers, EC2, and other Layer 7 entry points into your AWS environment. AWS WAF gives you control over how traffic reaches your applications by enabling you to create security rules that block common attack patterns, such as SQL injection or cross-site scripting, and rules that filter out specific traffic patterns that you can define. WAF's default rule-set addresses issues like the OWASP Top 10 security risks and is regularly updated whenever new vulnerabilities are discovered.

**RDS Read Replicas:**

* Each Read Replica will have its own DNS endpoint.
* Automated backups must be enabled in order to use read replicas.
* Read Replication is exclusively used for performance enhancement.
* If the master DB were to fail, there is no automatic failover. You would have to manually create a new connection string to sync with one of the read replicas so that it becomes a master on its own. Then you’d have to update your EC2 instances to point at the read replica. You can have up to five copies of your master DB with read replication.

**RDS Backups:**

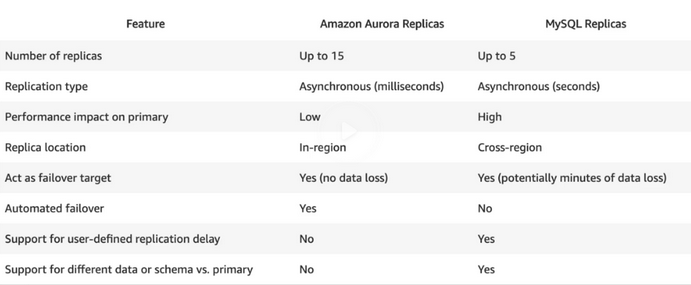
When it comes to RDS, there are two kinds of backups:

* automated backups
* database snapshots

**Aurora**

Aurora is the AWS flagship DB known to combine the performance and availability of traditional enterprise databases with the simplicity and cost-effectiveness of open source databases. It is a MySQL/PostgreSQL-compatible RDBMS that provides the security, availability, and reliability of commercial databases at 1/10th the cost of competitors. It is far more effective as an AWS database due to the 5x and 3x performance multipliers for MySQL and PostgreSQL respectively.

* In case of an infrastructure failure, Aurora performs an automatic failover to a replica of its own.
* Amazon Aurora typically involves a cluster of DB instances instead of a single instance. Each connection is handled by a specific DB instance.
* Aurora storage is self-healing and data blocks and disks are continuously scanned for errors. If any are found, those errors are repaired automatically.
* Automated failover is only possible with Aurora read replication



**DynamoDB Accelerator (DAX):**

* Amazon DynamoDB Accelerator (DAX) is a fully managed, highly available, in-memory cache that can reduce Amazon DynamoDB response times from milliseconds to microseconds, even at millions of requests per second.
* DAX does more than just increase read performance by having write through cache. This improves write performance as well.
* With DAX, your applications remain fast and responsive, even when unprecedented request volumes come your way. There is no tuning required.

**Redshift**

Amazon Redshift is a fully managed, petabyte-scale data warehouse service in the cloud. The Amazon Redshift service manages all of the work of setting up, operating, and scaling a data warehouse. These tasks include provisioning capacity, monitoring and backing up the cluster, and applying patches and upgrades to the Amazon Redshift engine.

* Redshift is used for business intelligence and pulls in very large and complex datasets to perform complex queries in order to gather insights from the data.
* Snapshots are point-in-time backups of a cluster. These backups are enabled by default with a 1 day retention period. The maximum retention period is 35 days.
* Redshift can also asynchronously replicate your snapshots to a different region if desired.
* A Highly Available Redshift cluster would require 3 copies of your data. One copy would be live in Redshift and the others would be standby in S3.
* Redshift is encrypted in transit using SSL and is encrypted at rest using AES-256. By default, Redshift will manage all keys, but you can do so too via AWS CloudHSM or AWS KMS.
* Redshift is not multi-AZ, if you want multi-AZ you will need to spin up a separate cluster ingesting the same input. You can also manually restore snapshots to a new AZ in the event of an outage.

**Route53 Simplified:**

Amazon Route 53 is a highly available and scalable Domain Name System (DNS) service. You can use Route 53 to perform three main functions in any combination: domain registration, DNS routing, and health checking.

* When you create a record, you choose a routing policy, which determines how Amazon Route 53 responds to DNS queries. The routing policies available are:
  + Simple Routing
  + Weighted Routing
  + Latency-based Routing
  + Failover Routing
  + Geolocation Routing
  + Geo-proximity Routing
  + Multivalue Answer Routing

**Bastion Hosts:**

* Bastion Hosts are special purpose computers designed and configured to withstand attacks. This server generally runs a single program and is stripped beyond this purpose in order to reduce attack vectors.
* The purpose of Bastion Hosts are to remotely access the instances behind the private subnet for system administration purposes without exposing the host via an internet gateway.
* The best way to implement a Bastion Host is to create a small EC2 instance that only has a security group rule for a single IP address. This ensures maximum security.
* It is perfectly fine to use a small instance rather than a large one because the instance will only be used as a jump server that connects different servers to each other.
* If you are going to RDP or SSH into the instances of your private subnet, use a Bastion Host. If you are going to be providing internet traffic into the instances of your private subnet, use a NAT.
* Similar to NAT Gateways and NAT Instances, Bastion Hosts live within a public-facing subnet.
* There are pre-baked Bastion Host AMIs.

**Virtual Private Networks (VPNs):**

* VPCs can also serve as a bridge between your corporate data center and the AWS cloud. With a VPC Virtual Private Network (VPN), your VPC becomes an extension of your on-prem environment.
* Naturally, your instances that you launch in your VPC can't communicate with your own on-premise servers. You can allow the access by first:
  + attaching a virtual private gateway to the VPC
  + creating a custom route table for the connection
  + updating your security group rules to allow traffic from the connection
  + creating the managed VPN connection itself.

**AWS DirectConnect:**

* Direct Connect is an AWS service that establishes a dedicated network connection between your premises and AWS. You can create this private connectivity to reduce network costs, increase bandwidth, and provide more consistent network experience compared to regular internet-based connections.
* The use case for Direct Connect is high throughput workloads or if you need a stable or reliable connection
* VPN connects to your on-prem over the internet and DirectConnect connects to your on-prem off through a private tunnel.

**VPC Flow Logs:**

* VPC Flow Logs is a feature that captures the IP information for all traffic flowing into and out of your VPC. Flow log data is sent to an S3 bucket or CloudWatch where you can view, retrieve, and manipulate this data.
* You can capture the traffic flow at various stages through its travel:
  + Traffic flowing into and out of the VPC (like at the IGW)
  + Traffic flowing into and out of the subnet
  + Traffic flowing into and out of the network interface of the EC2 instance (eth0, eth1, etc.)
* VPS Flow Logs capture packet metadata and not packet contents. Things like:
  + The source IP
  + The destination IP
  + The packet size
  + Anything which could be observed from outside of the packet.
* Your flow logs can be configured to log valid traffic, invalid traffic, or both

**Simple Workflow Service (SWF)**

SWF is a web service that makes it easy to coordinate work across distributed application components. SWF has a range of use cases including media processing, web app backend, business process workflows, and analytical pipelines.

* SWF is a way of coordinating tasks between application and people. It is a service that combines digital and human-oriented workflows.
* An example of a human-oriented workflow is the process in which Amazon warehouse workers find and ship your item as part of your Amazon order.
* SWF provides a task-oriented API and ensures a task is assigned only once and is never duplicated
* The SWF pipeline is composed of three different worker applications that help to bring a job to completion:
  + SWF Actors are workers that trigger the beginning of a workflow.
  + SWF Deciders are workers that control the flow of the workflow once it's been started.
  + SWF Activity Workers are the workers that actually carry out the task to completion.
* With SWF, workflow executions can last up to one year compared to the 14 days maximum retention period for SQS.