

AWS

Cloud Concepts

region

Regions are separate geographic locations designed to be completely independent of other regions to achieve fault tolerance and stability.

Key things to know

- Regions are physically isolated from each other
- Each region has at least two AZs within it
- US-East is the largest and where you will find your billing information when logged into the console

availability zones

Each region has multiple, isolated data centres called Availability Zones (AZs). This protects customers applications from service disruptions.

Key things to know

- AZs are represented by a region code followed by a letter - ap-southeast-2 is Sydney
- AZs in the same region have sub 10ms latency between each other

edge location

An Edge Location is a datacentre with a direct connection to the AWS network. This allows fast downloads from AWS with CloudFront and fast uploads to AWS with API Gateway.

AWS

Identify & Access Management (IAM)

what is it?

Allows you to securely control individual and group access to your resources. Users by default have no access until you assign them a role.

users

If you are setting up an AWS account for the first time you will begin with the root user and full administrative rights. Your first step should be creating a new user rather than using the root user for day to day activities.

After that you can create user accounts for anyone else who needs access to your account.

Users can have any combination of credentials:

- AWS access key
- X.509 certificate
- SSH key
- Password for web app logins
- MFA device

Default limits:

- 5000 per account
- 50 tags per user
- 5 SSH keys per user

AWS

Identify & Access Management (IAM)

groups

Groups are a collection of users. This could be a group of marketing users who need access to campaign data, finance users who need more sensitive data or customer service users who need access to customer data.

Default limits:

- 300 per account
- An IAM user can be a member of 10 groups

roles

This is what defines the set of permissions your users and services have. This is where you decide if the 'marketing' role needs to be able to read/write to an S3 bucket, read/write to the RDS, and nothing else.

Default limits:

- 1000 per account
- 50 tags per role

policies

When they are first created users have no permissions. These are added by creating and attaching policies.

- **Managed Policies** - fixed policies provided by AWS
- **Customer Managed Policies** - policies created by the customer
- **Inline Policies** - policies which are directly attached to a user

AWS

Identify & Access Management (IAM)

creating policies

Policies can be created either using the UI or by writing JSON with what you need.

Default limits:

- 1500 customer-managed policies
- 10 policies attached to a user or role

policy example

There are three parts to the policy at its most basic level:

- **Version** - the current version of the policy language.
- **Action** - in this case to Create and Delete buckets
- **Effect** - in this case to 'Allow'. This is by default set to 'deny' in the same way that users have no permissions by default when they are created.
- **Resource** - the syntax here is used to determine the Amazon Resource Name (ARN). In this case my_new_bucket in S3.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "s3:CreateBucket",
        "s3>DeleteBucket"
      ],
      "Effect": "Allow",
      "Resource": "arn:aws:s3:::my_new_bucket"
    }
  ]
}
```

AWS

Simple Storage Service (Amazon S3)

what is it?

An S3 bucket is where objects are stored, similar to files and folders on your local machine. There is unlimited storage available, across 100 buckets, and files can be from 0 bytes to 5TB.

Each object consists of:

- **Key** - the name of the object
- **Value** - the data in the file itself made of bytes
- **VersionID**
- **Metadata**

when to use it?

- **Analytics / Data Lake** - Uncouple storage and compute to scale either up or down as needed using Amazon Athena as the query service over the top and AWS Glue as a data catalogue.
- **Archive** - When data goes from 'hot', frequently accessed, to 'cold', infrequently accessed, it can be moved to Amazon Glacier for a more cost-effective option.
- **Data Staging** - Temporary data storage before being loading into AWS Redshift.
- **Static website** - Host a website using S3 for storage and Route 53 as the DNS.

AWS

Simple Storage Service (Amazon S3)

storage options

S3

- The most expensive as it promises 11 9's of durability.
- Good for cloud apps, big data, websites, content distribution.

S3:IA

- Costs 50% less than standard S3 as availability is reduced
- Recommended for non-critical data that **CANNOT** be easily reproduced and needs to be retrieved quickly
- Good for disaster recovery, backups

S3:IA - One Zone

- Costs less than S3:IA as durability is reduced
- Recommended for non-critical data that **CAN** be easily reproduced and needs to be retrieved quickly
- Good for secondary backups as objects are only stored in one zone

Glacier

- Much cheaper as there is a 3 - 5 hour retrieval time
- Good for long term storage, archives and 'cold' data

Deep Glacier

- The cheapest option as there is a 12 hour retrieval time
- Used for documents that need to be kept for compliance reasons for 7+ years

AWS

Simple Storage Service (Amazon S3)

security

- S3 is secure by default and each new bucket and the objects in it are private.
- To keep objects even more secure you can use bucket policies, similar to IAM policies, to fine tune access.
- Presigned URLs are another option to provide security when temporary access to an object is required.
- A URL is generated via the AWS CLI and SDK which can then be used to provide temporary access to write or download object data.

encryption

Client side

This is when the client encrypts the objects and uploads to Amazon S3.

Server side

This is when the data is encrypted when written and decrypts when it is being used.

- **SSE-AES** - S3 handles the key using the AES-256 algorithm
- **SSE-KMS** - Envelope encryption via AWS KMS, you manage keys
- **SSE-C** - Customer provided key, you manage the keys

AWS

Simple Storage Service (Amazon S3)

versioning

When versioning is turned on deleted files have a delete tag added which hides the file. To restore the file, delete the tag.

Key things to know:

- Each version takes up storage space. So a 1GB file edited three times with versioning on takes up 3GB of space
- Once turned on, versioning can only be suspended, not removed
- Versions that are deleted on the other hand are actually deleted
- Enabling multi factor authentication gives extra protection from accidental deletion

replication

Replication enables automatic, asynchronous copying of objects across Amazon S3 buckets. Buckets that are configured for object replication can be owned by the same AWS account or by different accounts. Objects can be replicated across regions or within the same region.

Key things to know:

- **Cross-Region replication (CRR)** is used to copy objects across Amazon S3 buckets in different AWS Regions
- **Same-Region replication (SRR)** is used to copy objects across Amazon S3 buckets in the same AWS Region
- Existing files won't be copied until there's been a new version, which will also replicate all previous versions and permissions

AWS

Elastic Compute Cloud (EC2)

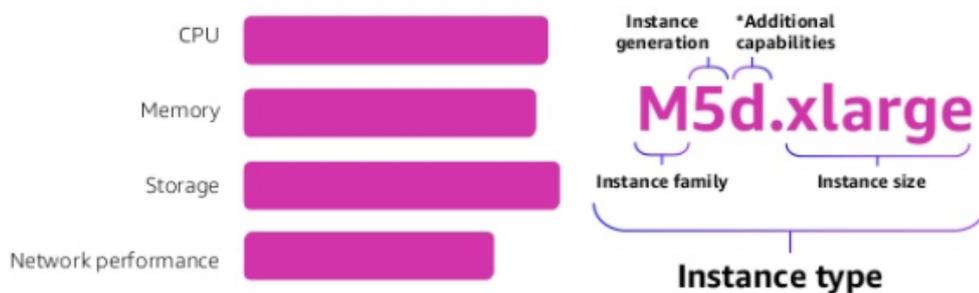
what is it?

EC2 is a service that provides virtual machines in the cloud where you only pay for the capacity you use and choose from 'families' of instance types that are good for different use cases.

what do the letters and numbers mean?

- **Family** - different instance types with resources for different use cases
- **Generation** - AWS phase out older technologies and bring in new ones with more resources using these numbers to show which is which
- **Size** - resources go up in a linear fashion, as well as the price that goes with it

Diagram below from 2018 re:Invent EC2 Fundamentals slides



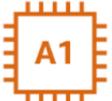
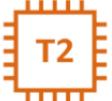
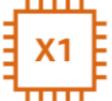
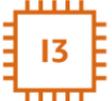
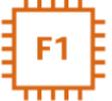
AWS

Elastic Compute Cloud (EC2)

how do i pick an instance type?

EC2 comes in variety instance types specialised for different roles:

- **General Purpose** - balanced compute, memory, and networking resources
- **Compute Optimised** - ideal for compute-bound applications that benefit from the high performance processor
- **Memory-Optimised** - fast performance for workloads that process large data sets in memory
- **Accelerated Optimised** - hardware accelerators, or co-processors
- **Storage Optimised** - high, sequential read and write access to very large data sets on local storage

General Purpose	Compute Optimised	Memory Optimised	Accelerated Computing	Storage Optimised
 ARM based core and custom silicon	 Compute - CPU intensive apps and DBs	 RAM - Memory intensive apps and DB's	 Processing optimised - Machine Learning	 High Disk Throughput - Big data clusters
 Tiny - Web servers and small DBs		 Xtreme RAM - For SAP/Spark	 Graphics Intensive - Video and streaming	 IOPS - NoSQL DBs
 Main - App servers and general purpose		 High Compute and High Memory - Gaming	 Field Programmable - Hardware acceleration	 Dense Storage - Data Warehousing

AWS

Elastic Compute Cloud (EC2)

payment options

On-Demand

- Pay for capacity by per hour or second
- No commitment
- Good for apps being developed or with unpredictable usage spikes

Reserved Instances

- Provides a reservation at 75% off the On-Demand price
- Gives you the ability to launch instances when you need them
- Reduced price as you need to commit to one or three-year terms and decide if you will pay all upfront, partial upfront, or no upfront

Spot Instances

- Bid for spare capacity for up to 90% off the On-Demand price
- Flexible start and end times
- If you're outbid the instance is terminated and you don't pay for the hour
- If you stop the instance you will pay for the hour
- Good for those background jobs which aren't critical

Dedicated Hosts

- Provides capacity on dedicated physical servers
- Good for when can't share capacity due to regulatory reasons or for licensing requirements

Savings Plan

- Provides the benefits of Reserved Instances but with more flexibility
- You will need to commit to a one or three year term but can change instance type within the same family while taking advantage of savings

AWS

Scaling

what is it?

Auto Scaling launches and terminates Amazon EC2 instances automatically according to userdefined policies.

You can use Auto Scaling to maintain a fleet of AWS EC2 instances that can adjust to any presented load. You can also use Auto Scaling to bring up multiple instances in a group at one time.

Scaling Plans

- **Maintain current level** - this plan checks all running EC2 instances and if it sees that any are unhealthy it will terminate the instance and launch a replacement
- **Manual** - this option allows you to specify the minimum and maximum capacity
- **Scheduled scaling** - if you know when demand is highest you can specify when to scale up
- **Demand based scaling** - build a policy that specifies which parameters need to be met to scale up and down based on demand

AWS

Persistent storage

Amazon Elastic Block Storage (EBS)

Persistent storage device that can be attached to a **single** EC2 instance to be used as a file system for databases and storage

Types of EBS

- **General Purpose SSD (GP2)** - general purpose, can burst up to 3000 IOPS for volumes under 1 GB
- **Magnetic HDD (Standard)** - lowest cost bootable storage, meant for infrequently accessed workloads
- **Provisioned IOPS SSD (IO1)** - use if need more than 10,000 IOPS, can provision up to 20,000 IOPS
- **Throughput optimised HDD (ST1)** - used for big data
- **Cold HDD (SC1)** - lowest cost meant for infrequently access workloads

Amazon Elastic File Storage (EFS)

Amazon Elastic File Storage (EFS) is a managed, scalable network file system that can be shared across **multiple** EC2 instances.

You can create a file system, mount the file system on an Amazon EC2 instance, and then read and write data to and from your file system.

EFS and EBS volumes can be created from the AWS Console or CLI. You can choose whether to encrypt the volume and the Availability Zone in which to create the volume. EBS volumes can only be attached to EC2 instances within the same Availability Zone.

AWS

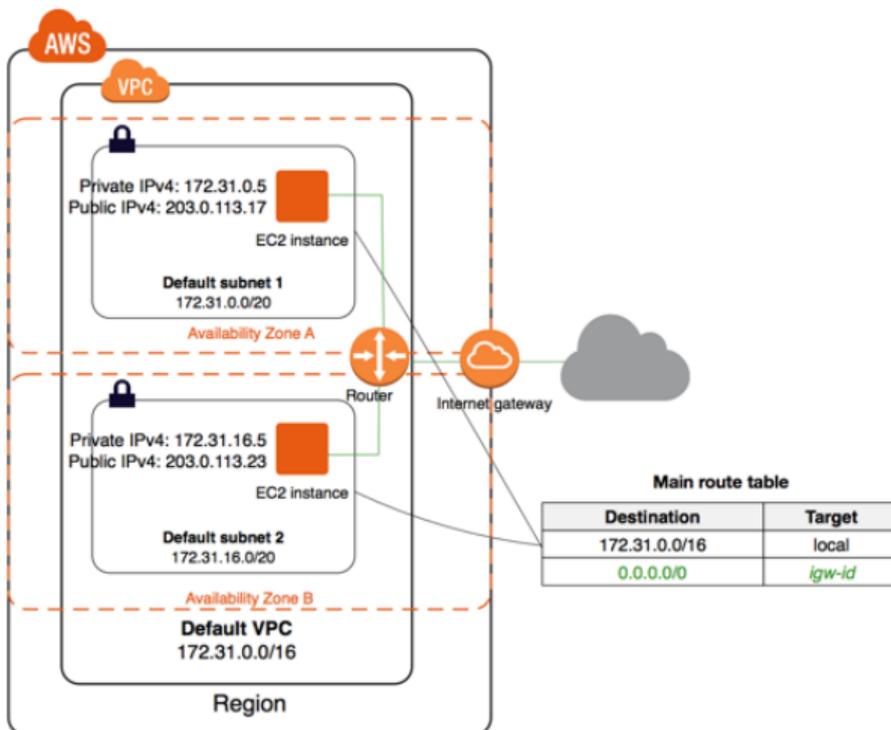
Virtual Private Cloud (VPC)

what is it?

A Virtual Private Cloud (VPC) is a virtual network dedicated to your AWS account. It is logically isolated from other virtual networks in the AWS Cloud.

You can launch your AWS resources, such as Amazon EC2 instances, into your VPC. Access is controlled using Route Tables, Internet Gateways, NAT Gateways, and Network Access Control Lists.

Diagram from AWS documentation



AWS

Virtual Private Cloud (VPC)

subnet

A subnet can be public or private. There can be up to 200 subnets per VPC. If you would like to increase this you will need to request this through AWS Support.

Key things to know

- You define which subnets you want to be exposed to the internet by attaching public IP addresses

route table

Each Subnet has a route table attached. This creates a set of rules to allow traffic to flow within a set of guidelines. This means that traffic stays inside the subnet until a route is created to allow it to travel to the next stop on the network.

Key things to know

- Route Tables create a set of rules to allow traffic to flow within a set of guidelines
- The Internet Gateway allows devices on a Public Subnet to connect to the internet
- In contrast, a Network Address Translation Gateway (NAT Gateway) facilitates the connection between Private Subnets and the internet

AWS

Virtual Private Cloud (VPC)

network access control lists (nacls)

Network Access Control Lists (NACL) allow us to limit traffic to safeguard against mistakes and accidents. Using NACLs lets us control traffic flow using a set of rules.

Key things to know:

- NACLs allow us to create a set of rules to allow and deny traffic to safeguard against mistakes and accidents
- The default NACL that comes with your VPC will allow all outbound and inbound traffic. When you create a NACL it will deny by default
- They are stateless which means the incoming rule will not be applied to the outgoing

web application firewall (waf)

WAFs protect web applications from attacks by filtering traffic based on rules. A WAF can be deployed on Amazon CloudFront, protecting resources and content at edge locations.

Use cases

- **Block IP addresses that exceed request limits** - this lets you control access to your content whether that's by IP address, country, blocking SQL injections, malicious scripts and the length of requests.
- **Block IP addresses that submit bad requests** - this lets you block IP addresses using Lambda, CloudWatch and AWS WAF to block requests after a threshold has been reached.

AWS

Relational Databases

what is it?

Amazon RDS makes it easy to provision a managed database instance in the cloud. At the time of writing the following database engines were available.

- Amazon Aurora - MySQL and PostgreSQL
- MySQL
- PostgreSQL
- MariaDB
- Oracle
- MS SQL Server

disaster recovery

Disaster recovery relates to the backups, logs and replication instances that are maintained while everything is working fine.

These are switched on, switched over, and analysed when something does go wrong, like a hardware failure, natural disaster or even human error.

- **Failover** - multiple clusters are set up so if one fails the other can take over
- **Mirroring** - maintaining two copies of the same database at different locations. One in offline mode so we know where things are at when we need to use it
- **Replication** - the secondary database is online and can be queried. This is not only good for disaster recovery but can be useful if you utilise one instance for reporting and one for live queries

AWS

NoSQL Databases

what is it?

You would use a non-relational database if:

- You have a lot of data with little structure
- Your data structure may change over time
- You want to make frequent changes to how the data is structured

Dynamodb

A key-value database which stores data with

- **a key** - as a unique identifier
- **a value** - anything from an integer to a JSON structure.

Amazon Neptune

A graph database set up to show relationships between people, places, objects and entities. If you've seen LinkedIn's 'Recommended Connections' you've encountered a Graph database.

AWS Timestream

A time-series database which aims to collect data points over time.

AWS Quantum Ledger

A ledger database designed to record a history of economic and financial activity.

AWS

CloudWatch

what is it?

CloudWatch measures **'what'** is happening in an AWS Account.

- **CloudWatch Logs** - logs data from AWS services - CPU utilisation
- **CloudWatch Metrics** - captures variables - Utilisation over time
- **CloudWatch Events** - triggers an event based on a condition - every hour take a snapshot of a server
- **CloudWatch Alarms** - triggers notifications when a threshold is breached
- **CloudWatch Dashboards** - create visualisations based on metrics

what can you watch?

- Load balancers
- Auto scaling groups
- EC2 instances
- SQS queues
- SNS topics
- Databases
- Objects in S3

AWS

CloudTrail

what is it?

CloudTrail is concerned with the **'who?'**

- Who made the API call?
- Which IP address has done something?
- How did a user access a bucket?

CloudTrail is turned on by default and sends logs to an S3 bucket for further analysis.

when to use it?

- CloudTrail can help with auditing and allows us to start with the problem, and trackback to where the problem began.
- It's timestamps a record of 'who' and lets us follow the trail to find the cause of any problems.
- CloudTrail is free of charge BUT the storing of the logs on S3 is not. Check out the [Monthly Cost Calculator](#)⁵ to find out how much you will be charged given your use case.

AWS

CloudFront

what is it?

Amazon CloudFront is the AWS Cloud Delivery Network (CDN). It caches information closest to the user so the next user can download a copy faster.

CloudFront can distribute content including dynamic, static, and streaming content from services like S3 or your own server

Key things to know:

- Edge locations can be used to read and write to
- Time to Live (TTL) defines how long until the cache expires and refreshes
- Distribution comes in two types - web distribution for static content and RTMP for streaming
- Signed URLs and cookies can be used to protect content

AWS

Kinesis

what is it?

Amazon Kinesis Data Firehose is a reliable way to stream data in near real-time. Data can be streamed to S3, Redshift or Elasticsearch.

Kinesis allows data to be streamed in real-time from a Producer to a Processor or Storage option.

This is a huge change from Batch Processing that has been the traditional way to land data from one location to another.

- **Batch Processing** - data is landed in chunks and analysed when the transfer is complete
- **Stream Processing** - streams of data pour in and don't have an end... unless you create one. This allows us to act on the data and make decisions faster

AWS

Kinesis Data Streams

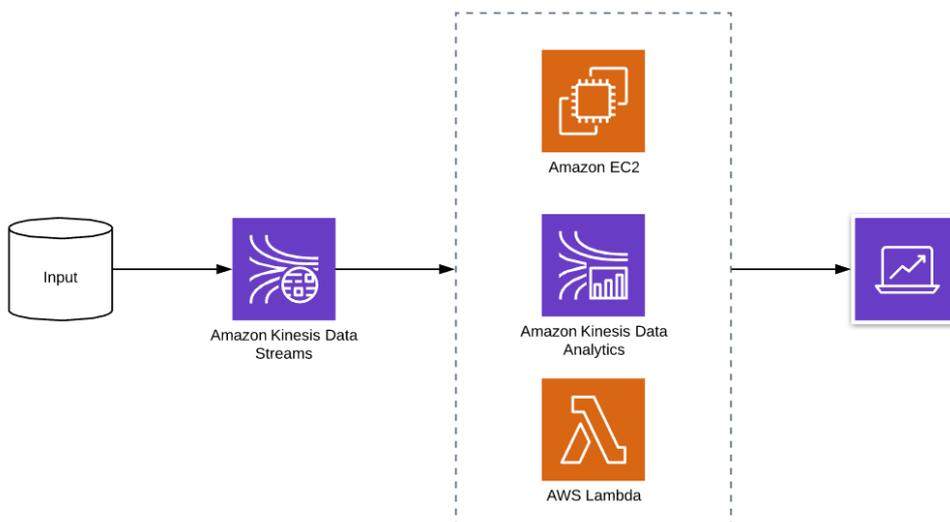
what is it?

This is an example of how a streaming process can work using AWS Kinesis.

The data comes into the stream, can then be processed by AWS Kinesis Analytics or another AWS Service and then lands at its destination.

use cases

- Streaming data like website clicks and transactional data
- Migrating data from databases
- Applications with specialised data pipelines



AWS

Kinesis Firehose

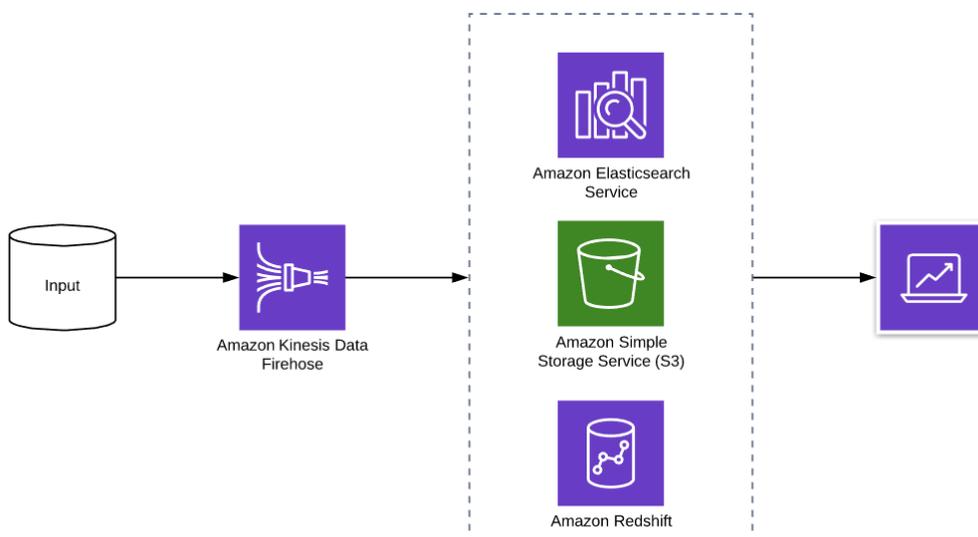
what is it?

Kinesis Firehose differs from Kinesis Data Streams as it takes the data, batches, encrypts and compresses it.

Then persists it somewhere such as Amazon S3, Amazon Redshift, or Amazon Elasticsearch.

use cases

- IoT events
- Security monitoring as Splunk can be configured as a destination
- Auto archiving



AWS

Kinesis Analytics

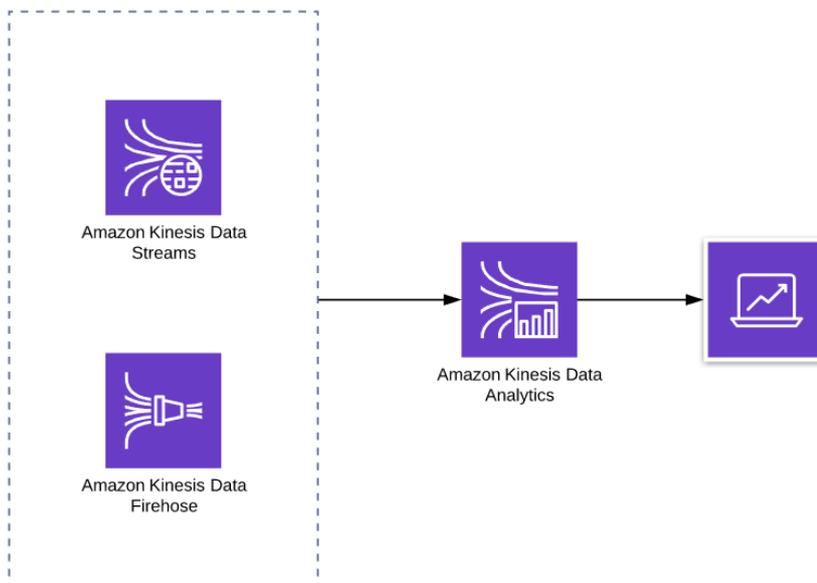
what is it?

Kinesis Data Analytics allows us to both process events and analyse them using SQL queries on-the-fly.

The service recognises formats like JSON and CSV, then sends the output to an analytics tool for visualisation or action.

use cases

- Processing of events data from applications
- Exploratory analysis
- Analysing clickstream anomalies



AWS

Simple Notification Service (SNS)

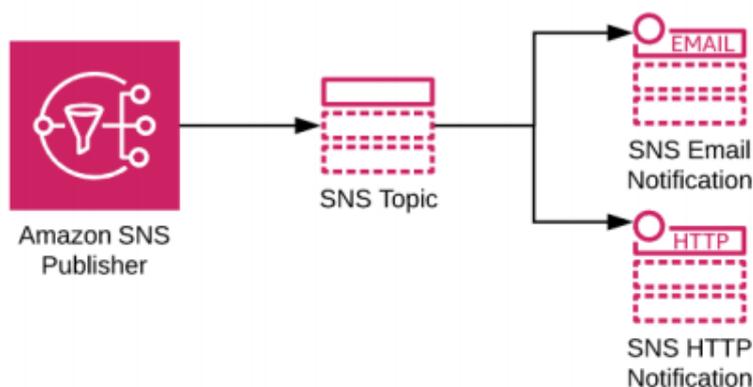
what is it?

Amazon Simple Notification Service (SNS) is a fully managed pub/sub messaging service

It is used in application integration and allows apps and services to communicate with pushed messages when these are decoupled.

Key things to know

- SNS can be used to send large numbers of time-sensitive messages to end-users in the form of a push notification, SMS and email
- Messages are in JSON format and are pushed to subscribers who can subscribe to topics and specify the endpoint
- 'Topics' are logical access points and allow recipients to subscribe to identical copies of the same message
- Notifications are then formatted for the protocol receiving the message and can be delivered as text messages, emails, and to SQS or HTTP endpoints



AWS

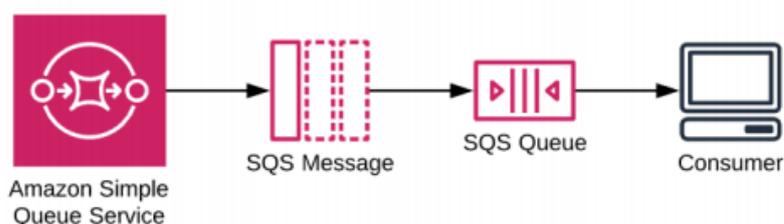
Simple Queue Service (SQS)

what is it?

Amazon Simple Queue Service (SQS) is a fully managed message queue service. It is used in application integration and allows apps and services to communicate with pulled messages when these are decoupled.

Key things to know

- SQS can be used to throttle workload and process work in batches with non-time sensitive messages being retained for up to 14 days
- Messages can be up to 256KB in any format
- If the job isn't processed before the time out expires the message will be placed back in the queue, which could cause the message to be processed twice
- The standard queue uses 'best-effort' ordering, FIFO queues guarantee messages are pulled in the order that they arrive
- By default short polling returns messages immediately. Long polling waits until there is a message in the queue or until the timeout expires
- Auto-scaling groups can monitor the SQS group and scale up and down depending on the number of messages in the queue



AWS

Security

Key Management Service (KMS)

AWS KMS makes it easy to create and control encryption keys. The service leverages Hardware Security Modules (HSM) which guarantees security and integrity of the generated keys.

Key things to know

- Keys have a unique alias and description
- Import your own key material
- Define which IAM users and roles can manage keys and decrypt data
- Automatic key rotation

IAM Security

Identity and Access Management (IAM) helps you to securely control who has access to your resources and how they access them

Key things to know

- Follow the best practice to enable MFA, delete root account credentials and create new roles with administrator permissions
- Create users, roles and groups to grant 'least access' and assign permissions to your users
- Use roles to allow access to services like EC2 rather than individual users
- Put conditions in place so users create strong passwords which get rotated regularly
- IAM Access Analyzer helps you identify resources in your organisation and accounts, such as Amazon S3 buckets or IAM roles, that are shared with an external entity.

AWS

Security

VPC and GuardDuty

Key things to know

- **Network Access Control Lists** - control inbound and outbound traffic at the subnet level
- **Security Groups** - act as a firewall at the EC2 level to control inbound and outbound traffic
- **VPC Flow logs** - capture information about how traffic is flowing
- **GuardDuty** - analyses data from VPC Flow Logs, and profiles them for anomaly detection. This service can detect a brute force attack on an EC2, suspicious API calls, or unauthorised behaviour

S3 and Macie

Key things to know

- **IAM policies** - control access to S3, and bucket policies to make sure buckets are kept private
- **MFA Delete and Versioning** - stops accidental deletion of objects and allow objects to be recovered using Cross-region replication
- **Amazon S3 Object Lock** - locks objects to prevent them being deleted during a fixed term or indefinitely
- **KMS or S3-Managed Keys** - for Server Side Encryption
- **Macie** - identifies personally identifiable information, API keys, and credentials

AWS

Security

EC2 and Inspector

Key things to know

- **Security Groups** - control inbound and outbound traffic to instances
- **Elastic Block Store (EBS) Encryption** - adds an extra layer of security
- **Inspector** - checks for access to your instances from the internet, remote root login being enabled, or vulnerable software versions installed

RDS and Redshift

Key things to know

- Encrypt data using AES-256⁵² level encryption
- Encrypt data in transit using SSL⁵³. This creates and installs the certificate when the instance is provisioned
- When using Redshift, enable cluster encryption⁵⁴ to encrypt user-created tables

CloudTrail

Key things to know

- Enable CloudTrail to provide a history of API calls made across your account
- Integrate with CloudWatch and SNS to support compliance and monitoring by setting up logs, metrics and alarms

AWS

Athena

what is it?

A service used to query files in S3 buckets

advantages

- Query data directly on a pay-for-what-you-use basis.
- No need to import data into a database.
- Structured and unstructured data formats are supported like CSV, JSON, and Apache Parquet.
- The service can be used on its own, integrated with AWS Glue as a Data Catalogue or with AWS Lambda as part of a bigger architecture.
- Use Amazon QuickSight to visualise data or for further analysis

limitations

- Consider Amazon Redshift if you need to store results or combine data from many different sources.
- AWS Athena is priced by query and the amount of data scanned. Consider compressing data or converting to columnar formats to reduce cost.