Cloud & AWS Workshop

Dr Simone Coughlan

October 2020

Cloud Computing Overview

Cloud computing

Cloud computing is the practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer

Cloud vs On-Premise

- **On Premise**

- You do everything (own computers, own IT staff, own buildings, electicity etc, own risk)
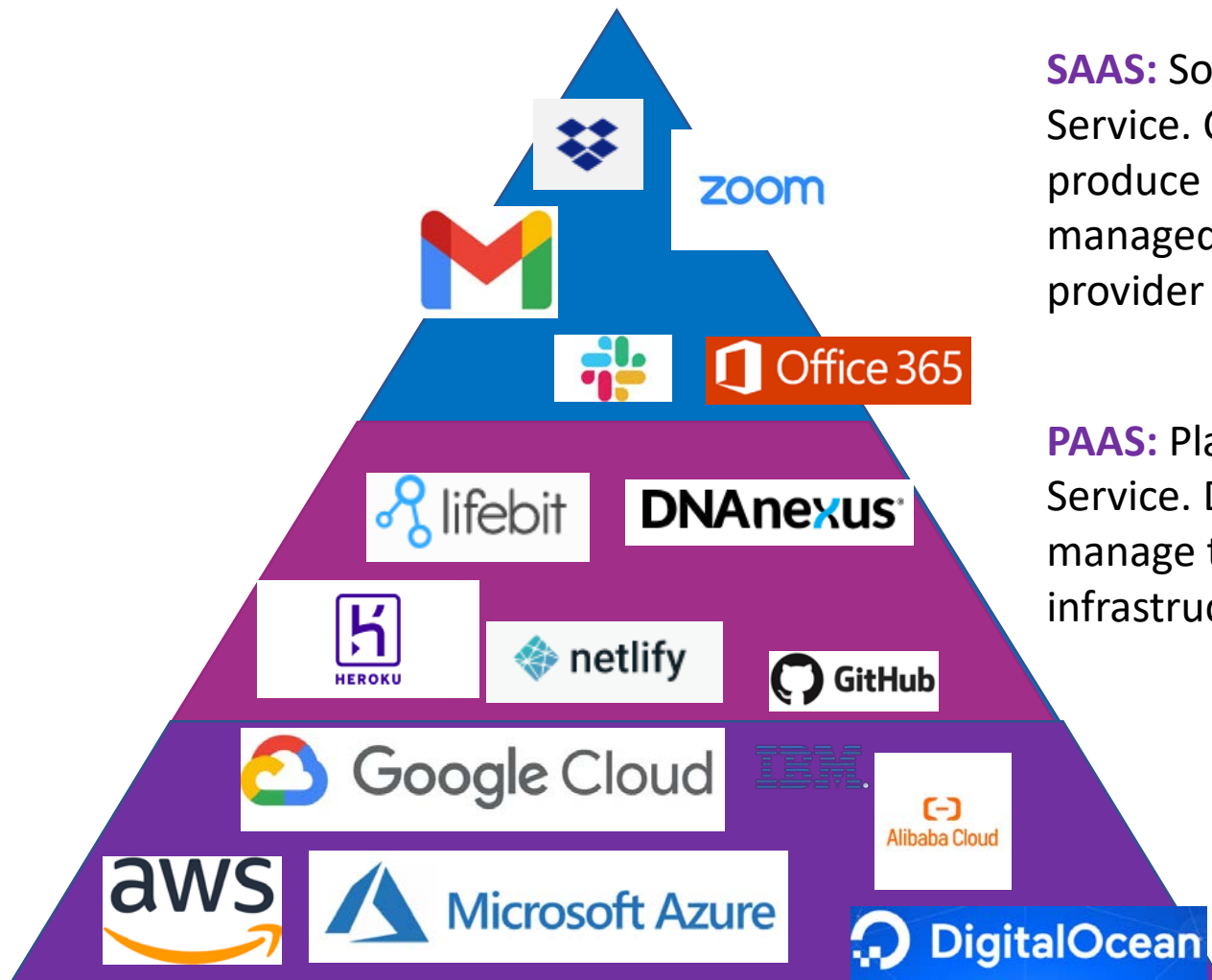
- **Cloud (provider)**

- Someone else does most stuff (owns the computers, staff, buildings but you configure the cloud services)

- Before cloud, everyone could only use On-premise

6 advantages of cloud over on-premise

1. **Trade capital expense for variable expense** –Pay only when you consume computing resources, and pay only for how much you consume.

2. **Benefit from massive economies of scale** – By using cloud computing, you can achieve a lower variable cost than you can get on your own. Because usage from hundreds of thousands of customers is aggregated in the cloud, providers can achieve higher economies of scale, which translates into lower pay as-you-go prices.

3. **Stop guessing capacity** – Eliminate guessing on your infrastructure capacity needs. You can access as much or as little capacity as you need, and scale up and down as required with only a few minutes' notice.

4. **Increase speed and agility** – In a cloud computing environment, new IT resources are only a click away, which means that you reduce the time to make those resources available from weeks to just minutes. This results in a dramatic increase in agility for the organization, since the cost and time it takes to experiment and develop is significantly lower.

5. **Stop spending money running and maintaining data centers** –don't need to worry about the heavy lifting of racking, stacking, and powering servers.

6. **Go global in minutes** – Easily deploy your application in multiple regions around the world with just a few clicks. This means you can provide lower latency and a better experience at minimal cost.

But best solution depends on your use!

Adapted from https://docs.aws.amazon.com/whitepapers/latest/aws-overview/six-advantages-of-cloud-computing.html

Cloud (service) models: Levels of abstraction

SAAS: Software as a Service. Complete produce run and managed by a service provider

PAAS: Platform as a Service. Don't need to manage the infrastructure

IAAS: Infrastructure as a Service
Building blocks of the cloud e.g. computer, storage, networking, databases

On-Premises	IaaS	PaaS	SaaS
	Infrastructure as a Service	Platform as a Service	Software as a Service
Applications	Applications	Applications	Applications
Data	Data	Data	Data
Runtime	Runtime	Runtime	Runtime
Middleware	Middleware	Middleware	Middleware
O/S	O/S	O/S	O/S
Virtualization	Virtualization	Virtualization	Virtualization
Servers	Servers	Servers	Servers
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

bmc

You Manage   Other Manages

Tech: Datacentres

- Datacentres house the computers and operations
- Cloud providers give access to datacentres all over the world

https://www.islandnetworks.com/solution/data-center-solutions/
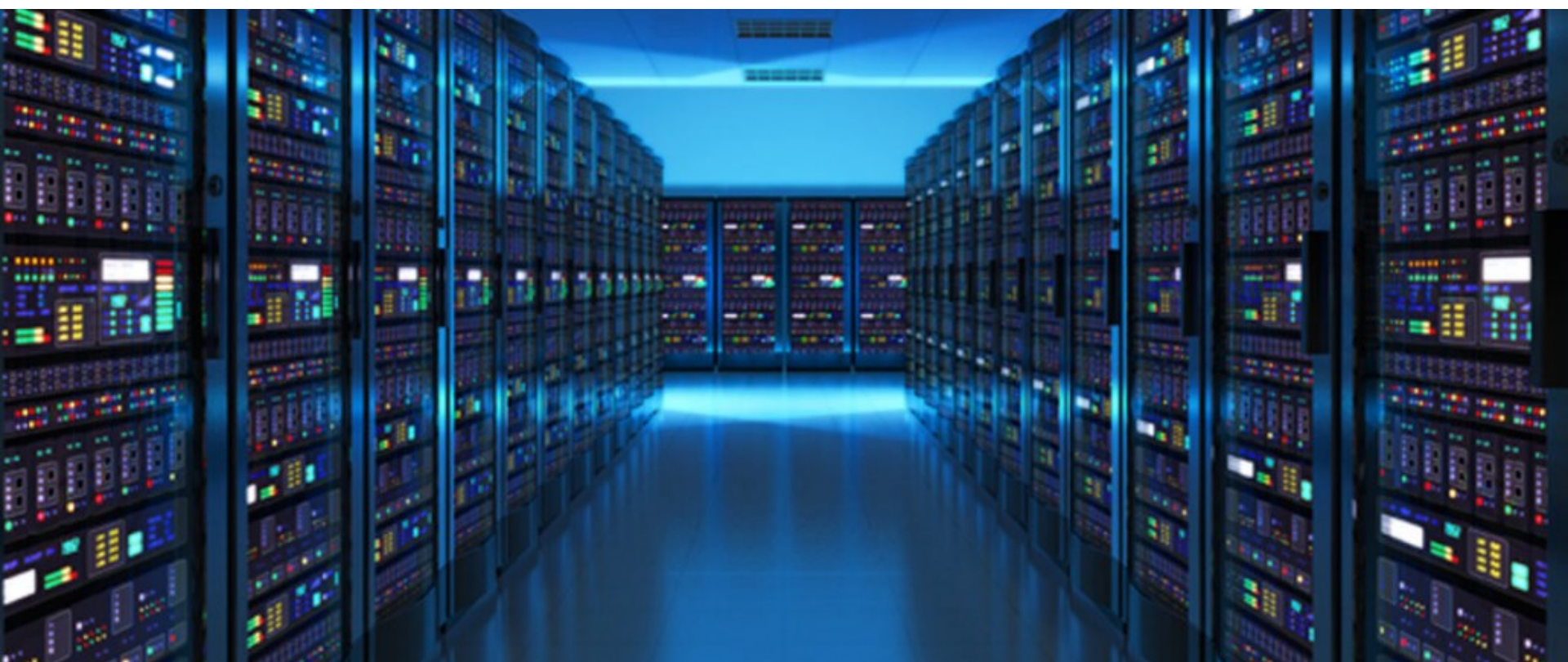
The history of the evolution of the cloud is shown in the following figure:

Figure 2.1: Cloud evolution
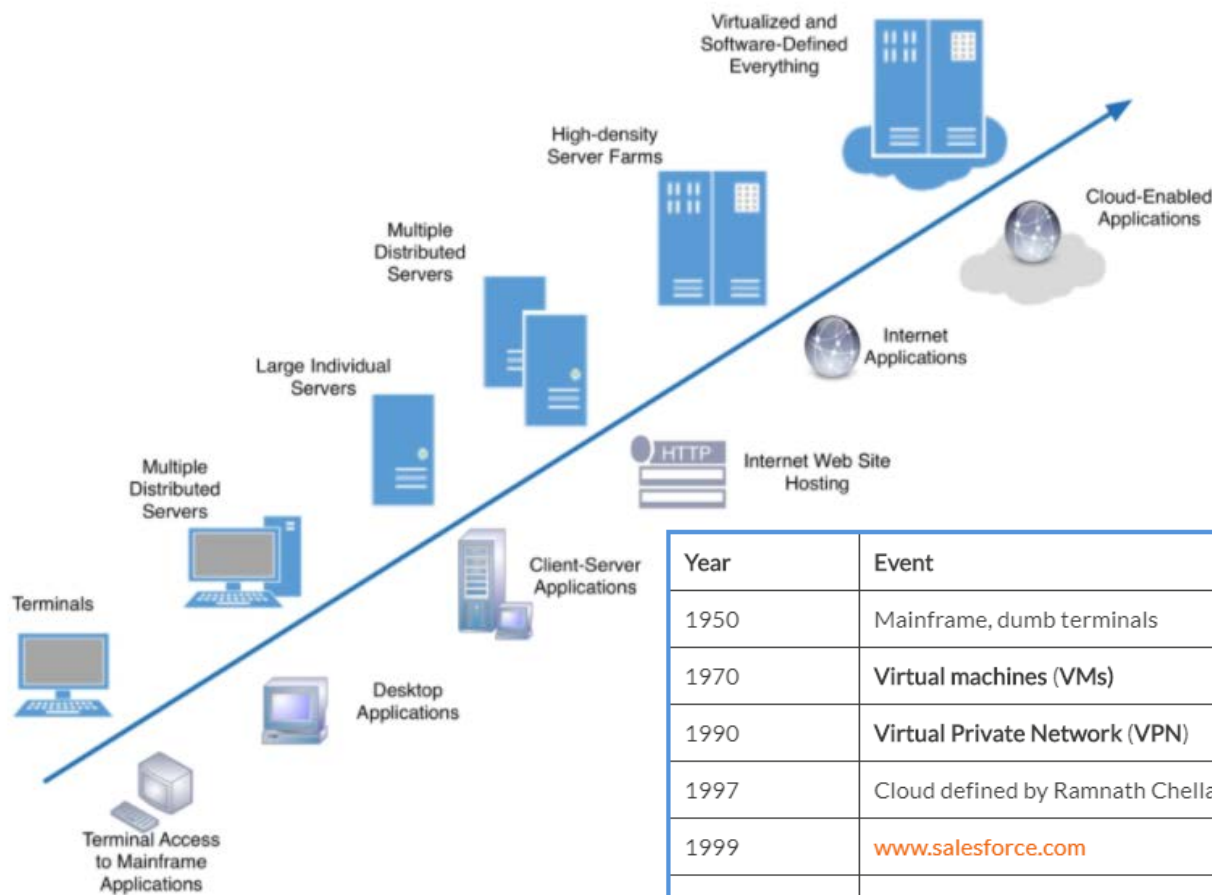
| Year	Event
1950	Mainframe, dumb terminals
1970	**Virtual machines** (VMs)
1990	**Virtual Private Network** (VPN)
1997	Cloud defined by Ramnath Chellappa
1999	www.salesforce.com
2000	Amazon's modernized data centers
2000	Google Docs service
2006	Launch of AWS Services
2008	Launch of Google app engine
2010	Launch of Microsoft Azure

Tech:Virtualisation

- Fundamental technology for cloud compute
- Produces multiple virtual servers from one physical server, maximising output of servers

TRADITIONAL ARCHITECTURE → VIRTUAL ARCHITECTURE
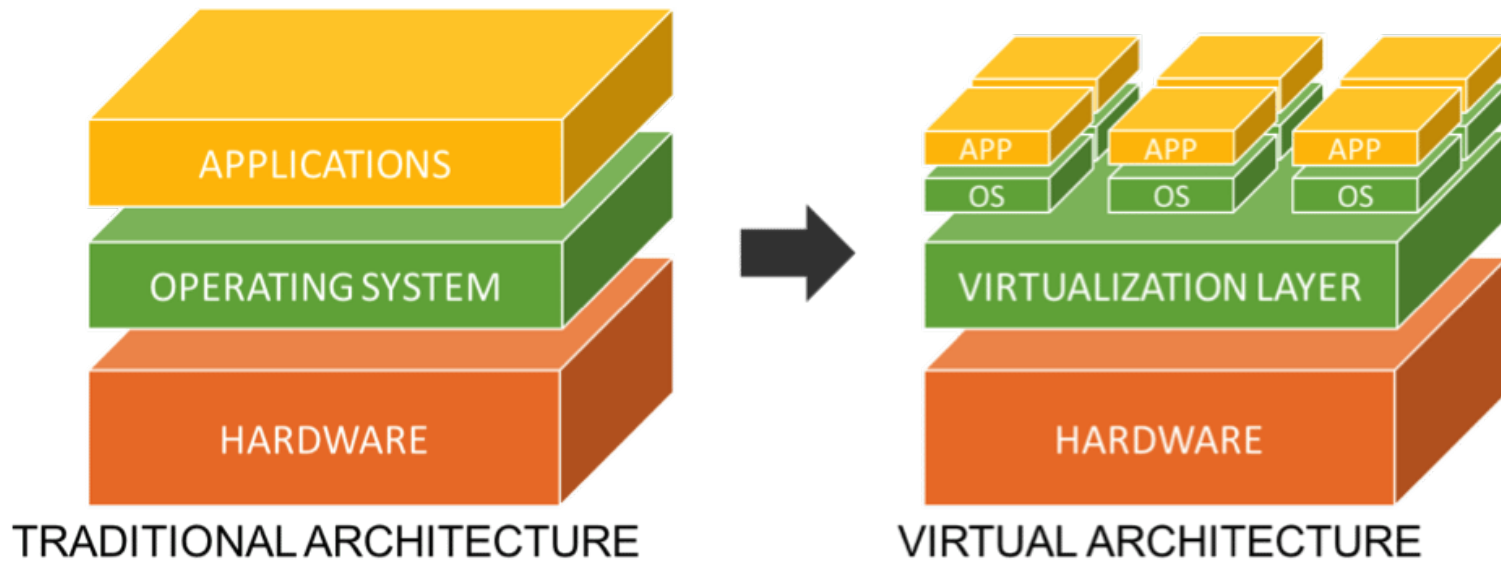
Tech: Hypervisors and VMs
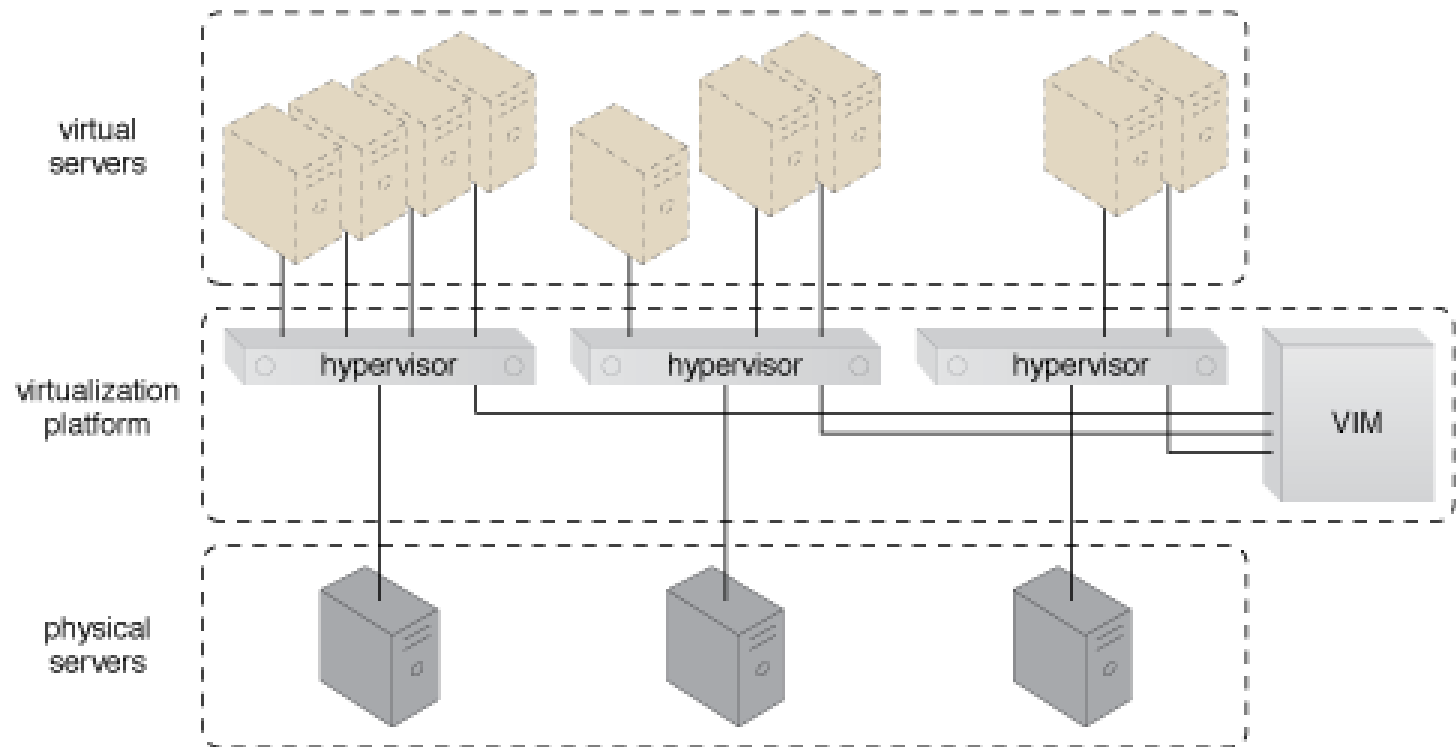
VIM= Virtualised Infrastructure Manager

Cloud Computing Deployment Models

Private Cloud

Operated solely for a single organization

Maybe on premise or off premise

Community Cloud

Shared by several entities that have a common purpose.

Maybe on premise or off premise

Public Cloud

Available to the general public and owned by a single organization selling cloud services.

Hybrid Cloud

Any combination of two or more private / community or public clouds.

Multicloud vs Hybrid Cloud

- **Multicloud** refers to the presence of more than 1 cloud deployment of the same type (public or private), sourced from different vendors e.g. using both AWS and GCP

- **Hybrid cloud :** using multiple deployment types (public or private) with some form of integration or orchestration between them. e.g. using GCP and on-premise servers

Example Users for deployments

- **On-premise/private cloud:** public sector e.g. government, really sensitive data e.g. hospitals, heavily regulated area e.g. insurance.

- In US, AWS has GovCloud

- **Hybrid:** Banks, fintech, investment management, organsiations with legacy on-premise

- **Public Cloud:** Startups, Saas offerings e.g. dropbox

Genomics?

Genomics Companies & Cloud

Our customers

Color

Google Cloud enables scientists to focus on their work by analyzing huge volumes of genomic data in seconds rather than hours.

Read case study →

Broad Institute

Replacing in-house genome sequence analysis computers and storage with Google Cloud for greater speed, scalability, and data security.

Read case study →

Clear Labs

Delivering pathogen test results to enterprise food producers in hours instead of days with Pipelines API and Google Cloud.

Read case study →

Genomics Companies & Cloud

Illumina case study

In this case study, learn how Illumina massively scales its DNA sequencing technologies using AWS. See how they support the Illumina BaseSpace Sequence Hub and store 10 PB of genomics data using products like Amazon RedShift

Learn more »

Sequence Bio case study

In this case study, learn how Sequence Bio quickly built a safe and secure platform for data-driven drug discovery on AWS.

Learn more »

Smithsonian case study

In this case study, learn how the Smithsonian Institute data science team scales AWS compute instances up and down as needed, allowing the team to annotate genomes in parallel while also managing costs.

Learn more »

UC Santa Cruz case study

In this case study, learn how the UC Santa Cruz Genomics Institute was able to reduce their genomic computational time from three months down to four days using AWS high-performance compute services, and reduce overall costs.

Learn more »

Blog: Driving momentum in genomics research

AWS Education Blog: AWS collaborates wtih Broad Institute

Learn how Cromwell, an execution engine that simplifies the orchestration of computing tasks needed for genomic analysis, is now enabled on the AWS Cloud.

Learn more »

Blog: Deploy Illumina DRAGEN with new quick start

AWS What's New: New Quick Start to deploy Illumina DRAGEN on the AWS Cloud.

Learn how this Quick Start deploys Dynamic Read Analysis for GENomics Complete Suite (DRAGEN CS), a data analysis platform by Illumina, on the AWS Cloud in about 15 minutes.

Blog: precision medicine at scale

AWS Compute Blog: Accelerating Precision Medicine at Scale

Learn how Edico Genome developed a novel solution to accelerate genomics analysis using FPGA-enabled applications on AWS.

Learn more »

Blog: Human Longevity, Inc.

AWS News Blog: Changing medicine through genomics research.

Learn how Human Longevity, Inc. is using using AWS to store the massive amount of data that is being generated as part of the effort to revolutionize medicine.

Learn more »

3 basic services offered by all providers

- Compute: VMs (instances)
- Storage: store text, image files etc (data replicated between datacentres so highly reliable)
- Databases: Store structured data
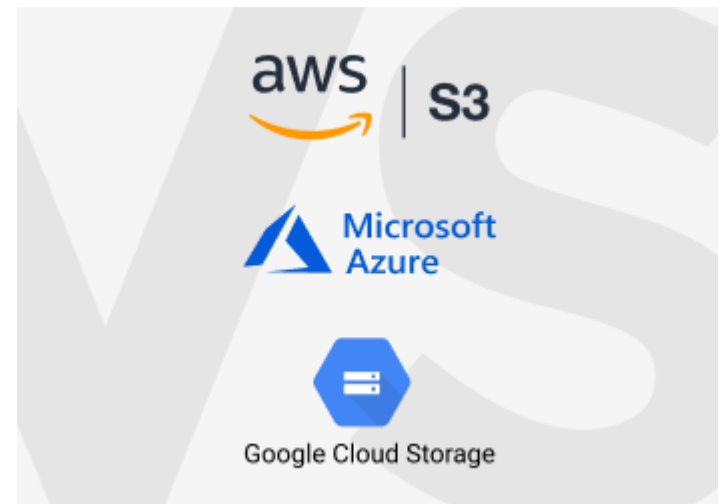
Cloud providers

Amazon Leads $100 Billion Cloud Market

Worldwide market share of leading cloud infrastructure service providers in Q2 2020*

Provider	Market share
amazon web services	33%
Azure	18%
Google Cloud	9%
Alibaba Cloud	6%
IBM Cloud	5%
salesforce	3%
Tencent Cloud	2%
ORACLE CLOUD	2%

Annual cloud infrastructure service revenue**
$111 billion
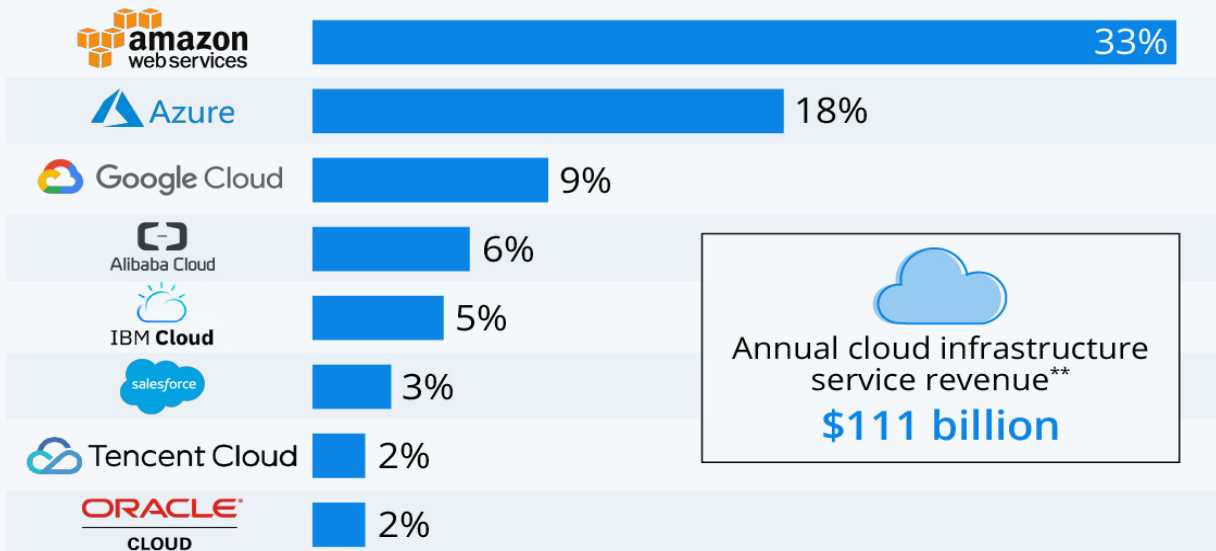
* includes platform as a service (PaaS) and infrastructure as a service (IaaS) as well as hosted private cloud services
** 12 months ended June 30, 2020
Source: Synergy Research Group

Percent of total sales generated by each company

statista

AWS: Amazon Web Services

- AWS is the IAAS arm of Amazon.com

- Launched in 2006 as a side business for Amazon

- Offering IT infrastructure services to businesses as web services i.e. cloud computing.

- Powers hundreds of thousands of businesses in 190 countries

- Can use AWS services through AWS dashboard in your account, AWS Command Line Interface(AWS CLI), python boto3 library

AWS Free Tier

AWS Free Tier

Gain free, hands-on experience with the AWS platform, products, and services

Create a Free Account

Types of offers

Explore more than 60 products and start building on AWS using the free tier. Three different types of free offers are available depending on the product used. See below for details on each product.

Always free

These free tier offers do not expire and are available to all AWS customers

12 months free

Enjoy these offers for 12-months following your initial sign-up date to AWS

Trials

Short-term free trial offers start from the date you activate a particular service

Free tier products

Filter by:
Clear all filters

▼ Tier Type
- ☐ Featured
- ☐ 12 Months Free
- ☐ Always Free
- ☐ Trials

▼ Product Categories
- ☐ Analytics
- ☐ Application Integration
- ☐ AR & VR
- ☐ Business Productivity
- ☐ Compute
- ☐ Customer Engagement
- ☐ Database
- ☐ Developer tools
- ☐ End User Computing
- ☐ Front-End Web & Mobile
- ☐ Game Tech
- ☐ Internet of Things
- ☐ Machine Learning
- ☐ Management & Governance
- ☐ Media Services

🔍 Search free tier products

COMPUTE

Free Tier   12 MONTHS FREE

Amazon EC2
750 Hours
per month

Resizable compute capacity in the Cloud.

750 hours per month of Linux, RHEL, or SLES

STORAGE

Free Tier   12 MONTHS FREE

Amazon S3
5 GB
of standard storage

Secure, durable, and scalable object storage infrastructure.

5 GB of Standard Storage

DATABASE

Free Tier   12 MONTHS FREE

Amazon RDS
750 Hours
per month of db.t2.micro database usage (applicable DB engines)

Managed Relational Database Service for MySQL, PostgreSQL, MariaDB, Oracle BYOL, or SQL Server.

DATABASE

Free Tier   ALWAYS FREE

Amazon DynamoDB
25 GB
of storage

MACHINE LEARNING

Free Tier   FREE TRIAL

Amazon SageMaker
250 Hours
per month of t2.medium notebook usage for the first two months

COMPUTE

Free Tier   ALWAYS FREE

AWS Lambda
1 Million
free requests per month

https://aws.amazon.com/free/?all-free-tier.sort-by=item.additionalFields.SortRank&all-free-tier.sort-order=asc

A note on Identity & Access Management (IAM)

- IAM enables you to manage access to AWS services and resources securely. Using IAM, you can create and manage AWS users and groups, and use permissions to allow and deny their access to AWS resource

- When you first create an Amazon Web Services (AWS) account, you begin with a single sign-in identity. That identity has *complete access* to all AWS services and resources in the account. This identity is called the AWS account *root user*.

- It is strongly recommended that you *do not use the root user for your everyday tasks, even the administrative ones*.

- Instead, adhere to the best practice of using the root user only to create your first IAM user. Then securely lock away the root user credentials and use them to perform only a few account and service management tasks.

- To view the tasks that require you to sign in as the root user, see AWS Tasks That Require Root User in Resources slide.

- Create access policies for users which grant least amount of privileges needed

AWS Global Infrastructure

Regions
Coming Soon

245 Countries and Territories Served

24 Launched Regions
Each with multiple Availability Zones (AZ's)

77 Availability Zones

217 Points of Presence
205 Edge Locations and 12 Regional Edge Caches

https://aws.amazon.com/about-aws/global-infrastructure/

AWS: Regions, AZ's & Edge Locations

- Over 1 million active customers
- **Region:** Physical location which has multiple availability zones
- **Availability Zone (AZ):** One or more discrete data centres
- **Edge Location:** Datacentre owned by an AWS partner

Region

- A geographically distinct location which contains multiple datacentres

- Physically isolated from every other region (diff location, electric, water supply etc)

- Has at least 2 AZ's

- Largest region is US-EAST-> new services always available here first

- Billing info always shown in US-East-1 (N.Virginia)

- Not all services available in all regions

Region Maps and Edge Networks

| North America | South America | Europe/Middle East/Africa | Asia Pacific |

Map Key

○ Regions

● Edge locations

Europe (Frankfurt) Region
Availability Zones: 3
Launched 2014

Europe (London) Region
Availability Zones: 3
Launched 2016

Europe (Paris) Region
Availability Zones: 3
Launched 2017

Middle East (Bahrain) Region
Availability Zones: 3
Launched 2019

Europe (Ireland) Region
Availability Zones: 3
Launched 2007

Europe (Milan) Region
Availability Zones: 3
Launched 2020

Europe (Stockholm) Region
Availability Zones: 3
Launched 2018

AWS Africa (Cape Town) Region
Availability Zones: 3
Launched 2020

Europe (Ireland) Region
Availability Zones: 3
Launched 2007

https://aws.amazon.com/about-aws/global-infrastructure/regions_az/

Availability Zones (AZ's)

- An AZ is a datacentre *owned and operated* by AWS in which AWS services run

- At least 2 AZs per region (nearly 3 in all regions)

- Represented by a region code followed by a letter which identifies the AZ e.g. us-east-1a (a is the datacentre, 6 Azs in US-east-1 so a to f)

- Can distribute instances between AZs for failover

- < 10 ms of latency between Azs (position to achieve this)

Edge Locations

- Aim to move data fast!

- **Edge location:** A datacentre *owned by a trusted AWS partner*, which has a direct connection to the AWS network

- Achieve low latency regardless of where the user is geographically located

- Serve requests for CloudFront and Route53. These will route requests to the nearest edge location automatically

- CloudFront = fast content delivery network (CDN) service that securely delivers data, videos, applications

- Route53 = highly available and scalable cloud Domain Name System (DNS) web service.

- S3 transfer acceleration (S3TA) traffic and API gateway traffic also used the AWS Edge Network

Amazon API Gateway

S3TA = speed up content transfers to and from Amazon S3 by as much as 50-500% for long-distance transfer of larger objects

API gateway = fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale

Edge Locations

Europe/Middle East/Africa Edge Locations

AWS Edge Network Locations:

Edge locations - Amsterdam, The Netherlands; Berlin, Germany; Brussels, Belgium; Cape Town, South Africa; Copenhagen, Denmark; Dubai, United Arab Emirates; Dublin, Ireland; Frankfurt, Germany; Fujairah, United Arab Emirates; Helsinki, Finland; Johannesburg, South Africa; London, England; Madrid, Spain; Manchester, England; Marseille, France; Milan, Italy; Munich, Germany; Oslo, Norway; Palermo, Italy; Paris, France; Prague, Czech Republic; Rome, Italy; Stockholm, Sweden; Vienna, Austria; Warsaw, Poland; Zurich, Switzerland; Tel Aviv, Israel

GDPR: personal data

General Data Protection Regulation (GDPR)

- Regulates how personal data is collected, processed, and stored from users in the EU

- Examples:
    - Users must explicitly consent to data collection

    - Notify users of any data breaches

    - Personal data information must be encrypted, anonymized, and/or pseudonymized

    - **Personal data can't leave EU borders, unless you can guarantee the same level of protection**

- Fine: 20 million Euros or up to 4% of the worldwide annual revenue

Source: Datacamp

GDPR: personal data

Art. 4 GDPR
Definitions

For the purposes of this Regulation:

(1) 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

(2) 'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

Excerpt from https://gdpr-info.eu/art-4-gdpr/

Art. 32 GDPR
Security of processing

1. Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:

   (a) the pseudonymisation and encryption of personal data;

   (b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;

   (c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;

   (d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

2. In assessing the appropriate level of security account shall be taken in particular of the risks that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

AWS Compute

- Virtual machines (VMs) called instances

- Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. (EC2 instances)

Millions of customers ranging from enterprises to startups

350+ instances for virtually every business need

24 regions and **76** availability zones globally

Choice of Intel, AMD, and Arm-based processors

AWS Compute

High performance storage

Amazon Elastic Block Store (EBS) provides easy to use, high performance block storage for use with Amazon EC2. Amazon EBS is available in a range of volume types that allow you to optimize storage performance and cost for your workloads. Many EC2 instance types also come with options for local NVMe SSD storage for applications that require low latency.

Choice of purchasing model

We offer a choice of multiple purchasing models with On-Demand, Spot Instances, and Savings Plan. With Spot Instances, you can save up to 90% for fault-tolerant workloads. With Savings Plan, you can save up to 72% savings with committed usage and flexibility across EC2, Fargate, and Lambda. You can also optimize your costs with recommendations on instances, built into EC2 with AWS Compute Optimizer or through tools such as Cost Explorer.

EC2 Instances
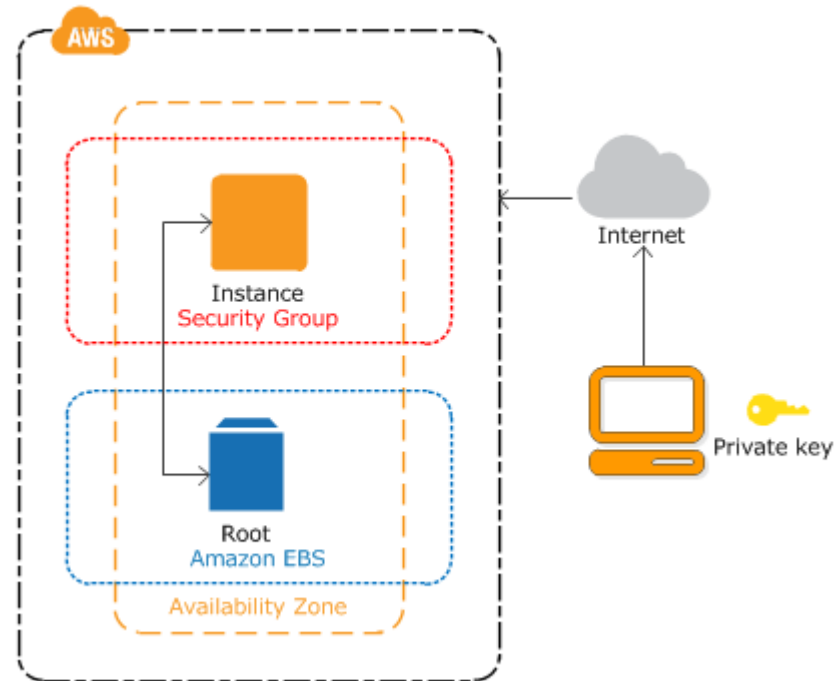
EC2 Instance Types

- General purpose

- Compute optimised

- Storage optimised

- Memory optimised

- Accelerated Computing

- https://docs.aws.amazon.com/AWSEC2/latest/User Guide/instance-types.html

- Cost depends on type and region the instance is used in along with pricing model chosen. Also need to consider adding storage to machine and moving data to machine

EC2 instance pricing

Region: Europe (Ireland) ⇕

	vCPU	ECU	Memory (GiB)	Instance Storage (GB)	Linux/UNIX Usage
General Purpose - Current Generation					
a1.medium	1	N/A	2 GiB	EBS Only	$0.0288 per Hour
a1.large	2	N/A	4 GiB	EBS Only	$0.0576 per Hour
a1.xlarge	4	N/A	8 GiB	EBS Only	$0.1152 per Hour
a1.2xlarge	8	N/A	16 GiB	EBS Only	$0.2304 per Hour
a1.4xlarge	16	N/A	32 GiB	EBS Only	$0.4608 per Hour

https://aws.amazon.com/ec2/pricing/on-demand/

EC2 Pricing Model

- On demand (least commitment)

- Spot (lowest cost, biggest saving)

- Reserved (best for long term use)

- Dedicated (most expensive)

How do you pay for AWS?

Pay-as-you-go

Pay-as-you-go allows you to easily adapt to changing business needs without overcommitting budgets and improving your responsiveness to changes. With a pay-as-you-go model, you can adapt your business depending on need and not on forecasts, reducing the risk of overpositioning or missing capacity.
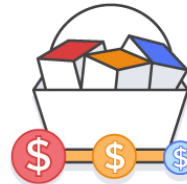
Save when you reserve

For certain services like Amazon EC2 and Amazon RDS, you can invest in reserved capacity. With Reserved Instances, you can save up to 75% over equivalent on-demand capacity. When you buy Reserved Instances, the larger the upfront payment, the greater the discount.

Pay less by using more

With AWS, you can get volume based discounts and realize important savings as your usage increases. For services such as S3, pricing is tiered, meaning the more you use, the less you pay per GB. AWS also gives you options to acquire services that help you address your business needs.

Image: https://aws.amazon.com/pricing/

On-demand

- Low cost and flexible

- No upfront payment

- Pay by the second for linux instances (min 60 secs), by the hour for some others (all prices shown per hour)

- Suited to short-term, unpredictable or spiky workloads

- Won't be interrupted

Spot

- Save up to 90 % of the cost of an on-demand instance
- Request spare computing capacity by bidding for it (if a machines or machines that you request are available then you will get them at the current spot price)
- If the cost rises above your bid or AWS needs the compute back, instance is terminated
- Suitable for workloads that can handle interruptions e.g. those with flexible start and end times, or applications that only feasible at very low cost
- Also defined duration (1 hour and 6 hour duration spot instances available)
- Can use with AWS Batch (plans, schedules, and executes your batch computing workloads)

Reserved Instances (RI)

- Save up to 75 % off the cost of on-demand
- Suitable for steady state or predictable use or you require reserved capacity
- Commit to EC2 over a 1 or 3 year term
- Can be shared between multiple accounts in an organisation
- Pricing based on

Term x Class Offering x Payment Option

- Term: 1 or 3 year term (longer is cheaper)
- Class: Standard , Convertible ,Scheduled
- Payment: All upfront, partial upfront, no upfront (the more upfront, the cheaper it is)

R1 Classes

- Standard: Up to 75 % off on-demand pricing. Cannot change attributes of instance. Can change the size (within same instance type) and AZ though. Can resell unused reserved instances on the Reserved Instance Marketplace

- Convertible: Up to 54 % off on-demand pricing. Allows you to change RI attributes if greater or equal value. Cannot resell on the Reserved Instance Marketplace

- Scheduled: Reserve for specific time periods e.g. once a month for a few hours. Savings vary

Dedicated servers

- Most expensive option
- Dedicated server(s)
- Use when need a guarantee of isolated hardware
- Designed to meet regulatory requirements
- For example you have strict server bound licensing that don't support multi-tenancy (i.e. you can't share the same underlying hardware as other customers)
- Multi-tenant: Using virtualisation as described before
- Single-tenant: Single customer has dedicated hardware (own physical server)
- Can be on demand or reserved pricing

Spot pricing

https://aws.amazon.com/ec2/spot/pricing/

Spot Instance Prices

Spot Instances	Defined Duration for Linux	Defined Duration for Windows

Note: T4g and T3 instances launch as unlimited by default. If you launch T4g or T3 Spot Instances as unlimited and plan to use them immediately and for a short duration, with no idle time for accruing CPU credits, you will incur charges for surplus credits. If the average CPU usage over a 24-hour period exceeds the baseline, you will also incur charges for surplus credits. We recommend that you launch your T4g or T3 Spot Instances in standard mode to avoid paying higher costs. For more information, see Surplus Credits Can Incur Charges and Spot Instance limits.

Region: Europe (Ireland) ⬍

	Linux/UNIX Usage	**Windows Usage**
General Purpose - Current Generation		
a1.medium	$0.0089 per Hour	N/A*
a1.large	$0.0178 per Hour	N/A*
a1.xlarge	$0.0356 per Hour	N/A*
a1.2xlarge	$0.0712 per Hour	N/A*

Compute Services

Services Pricing

Compute	Storage	Database	Migration & Transfer	Networking & Content Delivery

Amazon EC2
Virtual Servers in the Cloud

Amazon Elastic Container Service
Run and Manage Docker Containers

Amazon Elastic Container Registry
Store and Retrieve Docker Images

Amazon Lightsail
Launch and Manage Virtual Private Servers

AWS Batch
Run Batch Jobs at any Scale

AWS Elastic Beanstalk
Run and Manage Web Apps

AWS Fargate
Run Containers without Managing Servers or Clusters

AWS Lambda
Run your Code in Response to Events

Amazon Elastic Container Service for Kubernetes
Run managed Kubernetes on AWS

serverless

AWS vs GCP Terminology

Terminology

Feature	Amazon Web Services	Google Cloud Platform	Comments
Compute Service	Elastic Compute Cloud (EC2)	Google Compute Engine (GCE)	
Basic Virtual Server	On Demand Instance	Non-preemptible Instance	Both called "Instances"
Lowest Cost Option	Spot Instance	Preemptible Instance	Preemptible can be stopped
Scaling/Load Balancing	Autoscaling Groups	Instance Groups	
Stacks	CloudFormation Template	Instance Group Template	

AWS Storage

Service	Description
Amazon Simple Storage Service (Amazon S3)	A service that provides scalable and highly durable object storage in the cloud.
Amazon S3 Glacier	A service that provides low-cost highly durable archive storage in the cloud.
Amazon Elastic File System (Amazon EFS)	A service that provides scalable network file storage for Amazon EC2 instances.
Amazon Elastic Block Store (Amazon EBS)	A service that provides block storage volumes for Amazon EC2 instances.
Amazon EC2 Instance Storage	Temporary block storage volumes for Amazon EC2 instances.
AWS Storage Gateway	An on-premises storage appliance that integrates with cloud storage.
AWS Snowball	A service that transports large amounts of data to and from the cloud.
Amazon CloudFront	A service that provides a global content delivery network (CDN).

Separate cloud storage

EC2 storage

EBS persists independently from EC2 instance so can attach the EBS to new instances etc

Contrast with ephemeral storage which is attached to instance by default. Once instance terminated, data type disappears

https://docs.aws.amazon.com/whitepapers/latest/aws-storage-services-overview/introduction.html

AWS storage: S3

- Amazon Simple Storage Service (Amazon S3)
- Store objects in buckets
- Object: A file and any optional metadata that describes the file.
- Buckets: Containers for objects.
- You can control access for each bucket, deciding who can create, delete, and list objects in it. You can also choose the geographical Region where Amazon S3 will store the bucket and its contents and view access logs for the bucket and its objects.
- Global service with buckets in specific regions
- Bucket names *must be globally unique*
- Designed for 99.999999999% (11 9's) of durability

S3 types

- S3 Standard: General purpose storage of frequently accessed data
- S3 Standard-Infrequent Access (Standard-IA): For long-lived but less frequently accessed data
- S3 Glacier: For low cost archival data, can be slow to get data back out (faster = more expensive)

Case studies

Sony DADC NEW MEDIA SOLUTIONS

Sony DADC New Media Solutions moves its complete 20-petabyte video archive from LTO tape to Amazon S3

Celgene

Celgene uses Amazon S3 to store hundreds of terabytes of genomic data.

SOUNDCLOUD

Soundcloud uses Amazon S3 to store and process massive data sets every day.

King County

King County saved $1M in the first year after replacing tapes with Amazon S3.

illumina®

Illumina uploads and stores massive genomics datasets in Amazon S3 cloud archive storage.

https://aws.amazon.com/glacier/

S3 Security

- Manage access to Amazon S3 by granting other AWS accounts and users permission to perform the resource operations (using access policies).

- Protect data while *in-transit* (as it travels to and from Amazon S3) and *at rest* (while it is stored on disks in Amazon S3 data centers) using either:

- Server-Side Encryption: Request Amazon S3 to encrypt your object before saving it on disks in its data centers and then decrypt it when you download the objects.

or

- Client-Side Encryption: Encrypt data client-side and upload the encrypted data to Amazon S3. In this case, you manage the encryption process, the encryption keys, and related tools.

https://docs.aws.amazon.com/AmazonS3/latest/dev/UsingEncryption.html

Server-side encryption

3 mutually exclusive options

1. Server-Side Encryption with Amazon S3-Managed Keys (SSE-S3)

Each object is encrypted with a unique key. As an additional safeguard, it encrypts the key itself with a master key that it regularly rotates. Amazon S3 server-side encryption uses one of the strongest block ciphers available, 256-bit Advanced Encryption Standard (AES-256), to encrypt data

2. Server-Side Encryption with Customer Master Keys (CMKs) Stored in AWS Key Management Service (SSE-KMS)

Similar to SSE-S3. There are separate permissions for the use of a CMK that provides added protection against unauthorized access of your objects in Amazon S3. SSE-KMS also provides you with an audit trail that shows when your CMK was used and by whom.

Server-Side Encryption with Customer-Provided Keys (SSE-C)

You manage the encryption keys and Amazon S3 manages the encryption, as it writes to disks, and decryption, when you access your objects.

S3 pricing

any storage class. Estimate your costs using the AWS Pricing Calculator.

Region: [Europe (Ireland) ÷]

	Storage pricing
S3 Standard - General purpose storage for any type of data, typically used for frequently accessed data	
First 50 TB / Month	$0.023 per GB
Next 450 TB / Month	$0.022 per GB
Over 500 TB / Month	$0.021 per GB
S3 Intelligent - Tiering * - Automatic cost savings for data with unknown or changing access patterns	
Frequent Access Tier, First 50 TB / Month	$0.023 per GB
Frequent Access Tier, Next 450 TB / Month	$0.022 per GB
Frequent Access Tier, Over 500 TB / Month	$0.021 per GB
Infrequent Access Tier, All Storage / Month	$0.0125 per GB
Monitoring and Automation, All Storage / Month	$0.0025 per 1,000 objects
S3 Standard - Infrequent Access * - For long lived but infrequently accessed data that needs millisecond access	
All Storage / Month	$0.0125 per GB
S3 One Zone - Infrequent Access * - For re-createable infrequently accessed data that needs millisecond access	
All Storage / Month	$0.01 per GB
S3 Glacier ** - For long-term backups and archives with retrieval option from 1 minute to 12 hours	
All Storage / Month	$0.004 per GB
S3 Glacier Deep Archive ** - For long-term data archiving that is accessed once or twice in a year and can be restored within 12 hours	
All Storage / Month	$0.00099 per GB

https://aws.amazon.com/s3/pricing/

S3 pricing

You pay for requests made against your S3 buckets and objects. S3 request costs are based on the request type, and are charged on the quantity of requests as listed in the table below. When you use the Amazon S3 console to browse your storage, you incur charges for GET, LIST, and other requests that are made to facilitate browsing. Charges are accrued at the same rate as requests that are made using the API/SDK. Reference the S3 developer guide for technical details on the following request types: PUT, COPY, POST, LIST, GET, SELECT, Lifecycle Transition, and Data Retrievals. DELETE and CANCEL requests are free.

LIST requests for any storage class are charged at the same rate as S3 Standard PUT, COPY, and POST requests.

You pay for retrieving objects that are stored in S3 Standard – Infrequent Access, S3 One Zone – Infrequent Access, S3 Glacier, and S3 Glacier Deep Archive storage. Reference the S3 developer guide for technical details on Data Retrievals.

Region: Europe (Ireland) ⇕

	PUT, COPY, POST, LIST requests (per 1,000 requests)	GET, SELECT, and all other requests (per 1,000 requests)	Lifecycle Transition requests (per 1,000 requests)	Data Retrieval requests (per 1,000 requests)	Data retrievals (per GB)
S3 Standard	$0.005	$0.0004	n/a	n/a	n/a
S3 Intelligent - Tiering	$0.005	$0.0004	$0.01	n/a	n/a
S3 Standard - Infrequent Access*	$0.01	$0.001	$0.01	n/a	$0.01
S3 One Zone - Infrequent Access*	$0.01	$0.001	$0.01	n/a	$0.01
S3 Glacier **	$0.055	$0.0004	$0.055	See below	See below
Expedited	n/a	n/a	n/a	$11.00	$0.03
Standard	n/a	n/a	n/a	$0.055	$0.01
Bulk	n/a	n/a	n/a	$0.0275	$0.0025
Provisioned Capacity Unit ***	n/a	n/a	n/a	n/a	$100.00 per unit

Storage Services

Services Pricing

Compute	Storage	Database	Migration & Transfer	Networking & Content Delivery

Amazon S3
Scalable Storage in the Cloud

Amazon EBS
Block Storage for EC2

Amazon Elastic File System
Managed File Storage for EC2

Amazon Glacier
Low-cost Archive Storage in the Cloud

Amazon FSx for Lustre
High-performance file system for processing Amazon S3 or on-premises data

Amazon FSx for Windows File Server
Fully managed native Microsoft Windows file system

AWS Storage Gateway
Hybrid Storage Integration

AWS Snowball
Petabyte-scale Data Transport

AWS Snowball Edge
Petabyte-scale Data Transport with On-board Compute

AWS Snowmobile
Exabyte-scale Data Transport

AWS Backup
Centrally manage and automate backups

Migrating data to cloud: AWS Snowball

- <u>Expedited bulk transfer of data to or from Amazon S3.</u>

- Snowball helps transfer data in situations where you have connectivity limitations, bandwidth constraints, high network connection costs, legacy environment challenges, and when data is collected in remote locations.

- <u>Edge computing applications</u>, to collect data, process the data to gain immediate insight, and then transfer the data to AWS.
- Transfer data that is continuously generated by sensors or machines online to AWS from hospitals, factory floors, or at other edge locations.

- Snowball Edge Storage Optimized devices provide 40 vCPUs of compute capacity coupled with 80 terabytes of usable block or Amazon S3-compatible object storage.

- Can put approx 80 TB in one snowball

What is Snowball? Petabyte scale data transport

Ruggedized case "8.5G Impact"

E-ink shipping label

50TB & 80TB 10G network

Rain & dust resistant

Tamper-resistant case & electronics

All data encrypted end-to-end

amazon web services

If you really have a lot of data to migrate...AWS Snowmobile

- An Exabyte-scale (1 EB = 1000 PB), data transfer service used to move extremely large amounts of data to AWS.

- You can transfer up to 100PB per Snowmobile, a 45-foot long ruggedized shipping container, pulled by a semi-trailer

- Can move 100 petabytes of data in as little as a few weeks, plus transport time. Same transfer could take more than 20 years to accomplish over a direct connect line with a 1Gbps connection.

- Uses multiple layers of security to help protect data including dedicated security personnel, GPS tracking, alarm monitoring, 24/7 video surveillance, and an optional escort security vehicle while in transit

- Data is encrypted (256 bit encryption) with keys that you provide before it is written to Snowmobile.

- Only available in some regions

https://aws.amazon.com/snowmobile/

AWS Snowcone

AWS Marketplace

https://aws.amazon.com/marketplace

Find AWS Marketplace products that meet your needs.

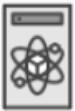| Categories	Vendors	Pricing Plans	Delivery Methods
All categories ▼	All vendors ▼	All pricing plans ▼	All delivery methods ▼

Total results: 9404

Clear selection   **View results**
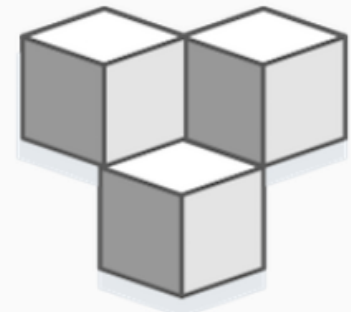
Popular Categories

Operating Systems

Security

Networking

Storage

Data Analytics

Dev Ops

Machine Learning

Data Products

View all categories

Vendors

illumina

◯ **Illumina Inc. (1)**

DRAGEN Complete Suite

By: **Illumina Inc.** ⬈   Latest Version: 3.6.3

The DRAGEN Complete Suite enables ultra-rapid analysis of Next Generation Sequencing (NGS) data for large data sets, such as whole genomes, exomes, and genes/panels.

Linux/Unix   ★★★★★ **1 AWS review**

Free Trial ◀

Continue to Subscribe

Save to List

Typical Total Price

$18.40/hr

Total pricing per instance for services hosted on f1.4xlarge in US East (N. Virginia). **View Details**

Overview   Pricing   Usage   Support   Reviews

Product Overview

The DRAGEN Complete Suite* enables ultra-rapid analysis of Next Generation Sequencing (NGS) data for large data sets, such as whole genomes, exomes, and genes/panels. This application uses the DRAGEN Platform and includes highly-optimized algorithms for mapping, aligning, sorting, duplicate marking, and haplotype variant calling. The DRAGEN CS includes a host of pipelines including our DRAGEN Germline Pipeline, DRAGEN Somatic Pipeline (T and T/N), DRAGEN Copy Number Variant (CNV) Pipeline, DRAGEN RNA Gene Fusion, DRAGEN Joint Genotyping Pipeline, and GATK Best Practices. The DRAGEN Germline and Somatic pipelines have greatly improved accuracy in calling SNPs and Indels compared to industry standard. This app also supports Illumina NovaSeq BCL conversion, download/upload of data streaming, and compressed reference hash tables for a more seamless and efficient workflow.

Highlights

- Ultra-Rapid
- Highest Accuracy
- Lowest Cost

All Products > Healthcare & Life Sciences > Sentieon Genomics Suite

Sentieon Genomics Suite

By: **Sentieon Inc** ⧉   Latest Version: 201911

The Sentieon tools are computationally efficient award winning software tools for germline and somatic variant calling.

Linux/Unix   ☆☆☆☆☆   **0 AWS reviews**

BYOL   **Free Tier**

Continue to Subscribe

Save to List

Typical Total Price
$1.728/hr
Total pricing per instance for services hosted on c5d.9xlarge in US East (N. Virginia). **View Details**

Overview   Pricing   Usage   Support   Reviews

Product Overview

Sentieon (http://www.sentieon.com ⧉) supplies award-winning software tools for secondary analysis of NGS data.

Sentieon DNAseq and TNseq produce results identical to the Broad Institute's BWA-GATK HaplotypeCaller/MuTect/MuTect2 Best Practice Workflow by implementing the same mathematics but with more efficient computing algorithms and enterprise-strength software implementation. Furthermore, the Sentieon tools do not downsample in high-coverage regions, are able to handle arbitrary depth of coverage, and have no thread dependency, resulting in 100% consistency.

Sentieon DNAscope and TNscope build upon and improve over the mathematical models from Haplotyper/Mutect2, providing additional accuracy and supporting the use of Machine Learning models for filtering.

The Sentieon tools are enterprise-strength software tools that are inherently

Highlights

- Identical mathematics as Broad Institute's BWA-GATK Best Practice Workflow, more efficient compute algorithms saving compute cost and enabling fast turn-around time and high throughput

- Award-winning enterprise strength software, no down-sampling, no run-to-run difference

- Pure software solution running on any generic-CPU-based system

https://aws.amazon.com/marketplace/pp/Sentieon-Inc-Sentieon-Genomics-Suite/B07WJ7CZBX

OpenEMR Cloud - Express Edition

Version 5.0.2-3 | Sold by OEMR

★★★★★ 3 AWS reviews | 1 external review ⓘ

OpenEMR Express is a single-instance OpenEMR cloud service designed for long-term, low-maintenance installations. OpenEMR Express features automated daily backups of patient data, portable cross-instance recovery process, SSL encryption and more. OpenEMR is an open source electronic health records...

Linux/Unix, Ubuntu 16.04 - 64-bit Amazon Machine Image (AMI)

DRAGEN Complete Suite

Free Trial

Version 3.6.3 | Sold by Illumina Inc.

★★★★★ 1 AWS review

Starting from $10.35 to $22.10/hr for software + AWS usage fees

The DRAGEN Complete Suite* enables ultra-rapid analysis of Next Generation Sequencing (NGS) data for large data sets, such as whole genomes, exomes, and genes/panels. This application uses the DRAGEN Platform and includes highly-optimized algorithms for mapping, aligning, sorting, duplicate...

Linux/Unix, CentOS Version 7.5 - 64-bit Amazon Machine Image (AMI)

AWS Latest vTiger Instance AMI Ready To Go

Version Version 7.2.0 | Sold by Chrisranjana aws developer team

Latest vTiger instance installed in AWS latest Ubuntu. Once this AMI is installed the admin password is auto generated and saved in /root/temp.txt No 2 instances admin passwords are same. You can straightaway access this vTiger using http://IP address/vtigercrm/ and can start using it. Contact...

Linux/Unix, Ubuntu 18.04 - 64-bit Amazon Machine Image (AMI)

https://aws.amazon.com/marketplace/b/2649342011?searchTerms=&category=9068362c-23da-4a02-b090-51ae4b4ca8b2&page=1

Resources

- Virtualisation basics: https://www.slideshare.net/sagaroceanic11/virtualisation-basics

- AWS Tasks That Require Root User https://docs.aws.amazon.com/general/latest/gr/root-vs-iam.html#aws_tasks-that-require-root

- IAM: https://docs.aws.amazon.com/AmazonS3/latest/gsg/SigningUpforS3.html

- Create IAM users: Create IAM users: https://docs.aws.amazon.com/IAM/latest/UserGuide/best-practices.html#create-iam-users