Aws Cloud Practitioner Certification Notes

AWS Certified Cloud Practitioner

Compiled by – Ujala Siddique

Table of Contents

Cloud Computing & Amazon Web Services

Cloud Computing: The practice of using network of remote servers hosted on the internet to store, manage, and process data, rather than a local server or a personal computer.

It refers to the on-demand delivery of it resources and applications via the internet without having to invest in hardware. Automatically scale computing to meet our needs

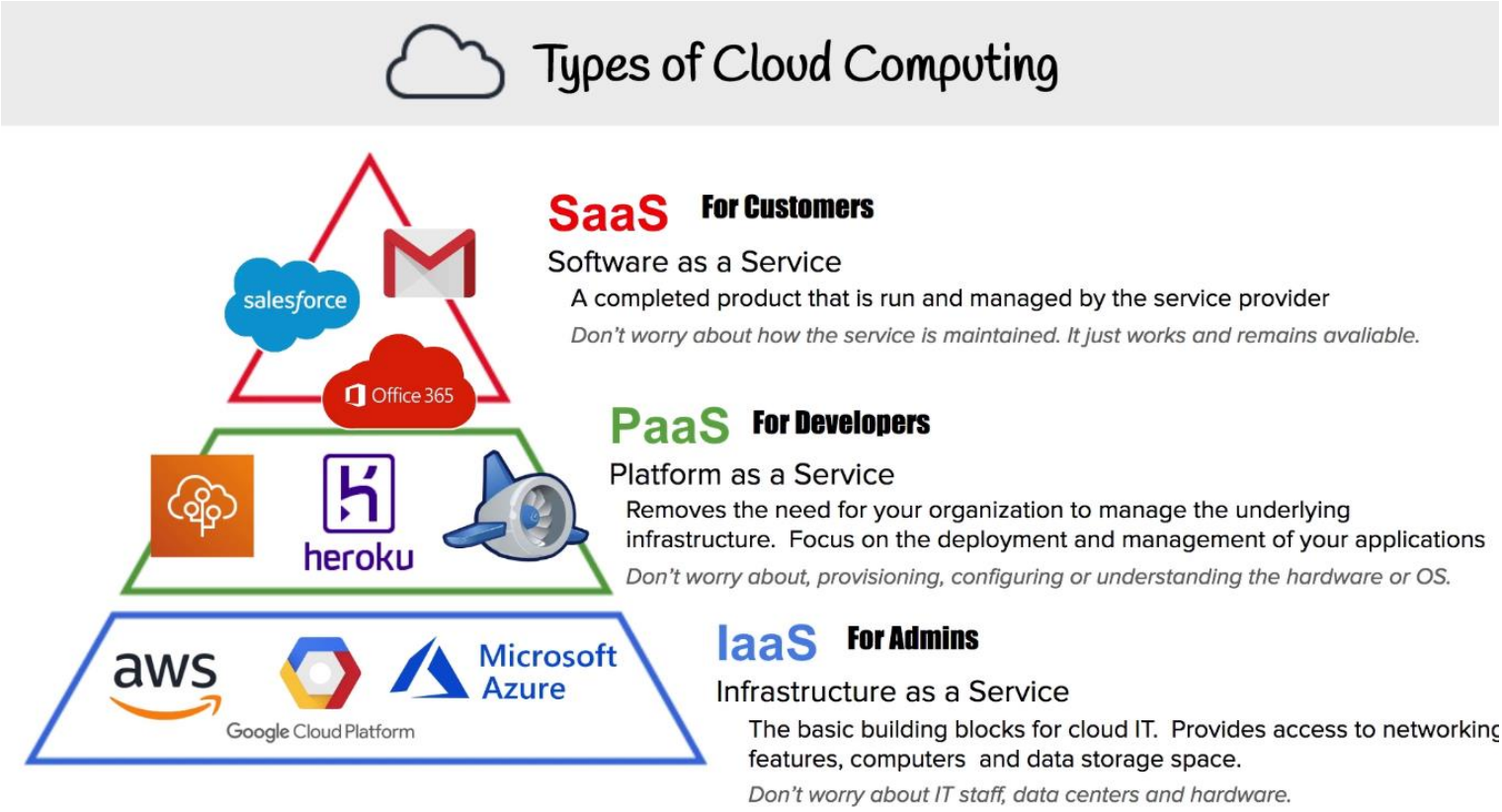
Elasticity: is the ability to scale computing up and down easily.

Agility: easy to access resources.

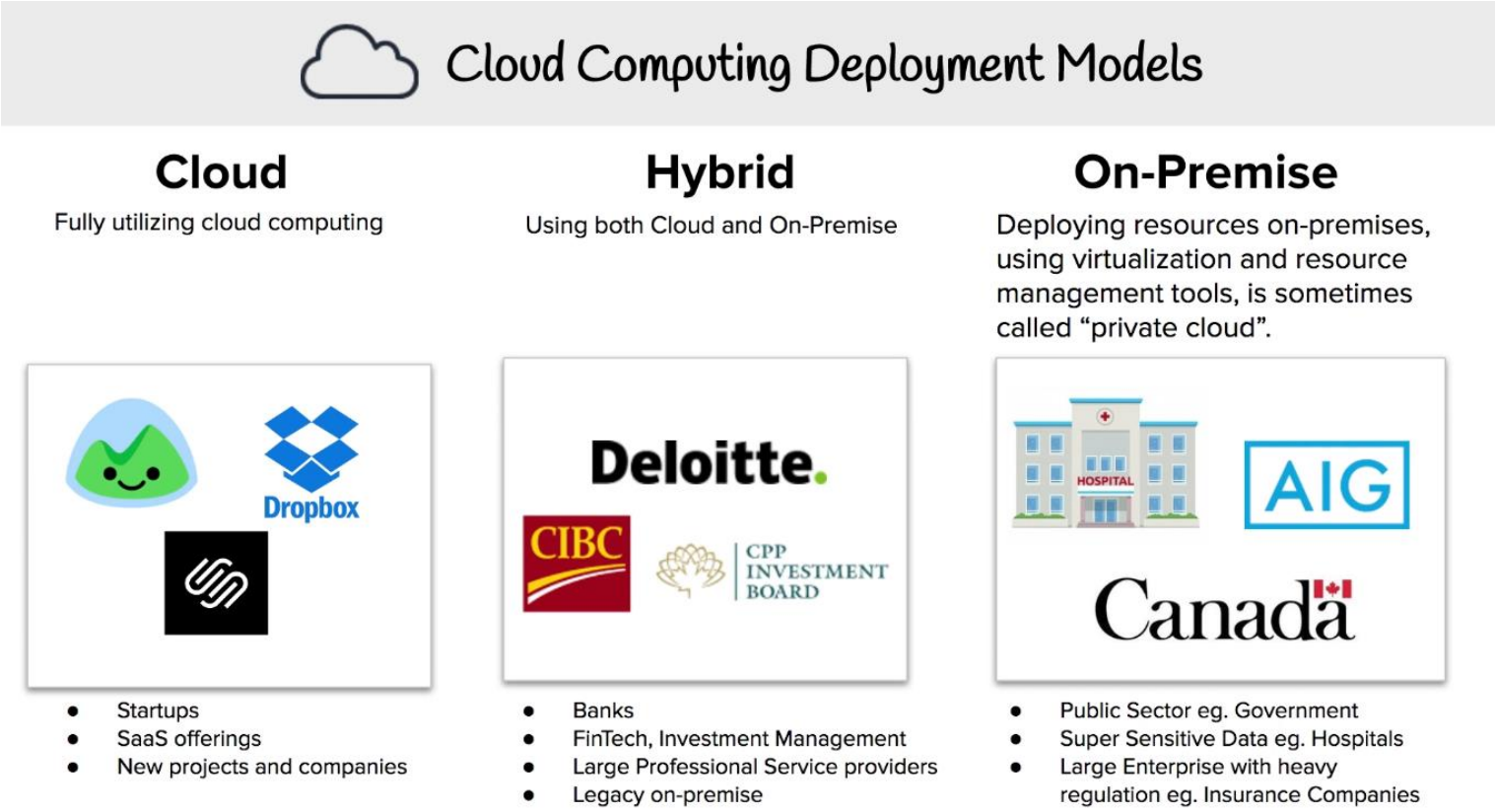
Reliability: Ability of a system to recover from failures. AWS uses regions and AZs.

3 ways to access AWS resources, that all reference the AWS API:

- Management Console – GUI
- CLI (command line interface) - Open source, language agnostic
- SDKs (software development kits)



***Pass (Aws Beanstalk, Engines for google)**



On-Premise:

- You own the servers
- You hire the IT people
- You pay or rent the real estate
- You take all the risk

Cloud Providers: (AWS, GCP, Azure, Alibaba, IBM)

- Someone else own the servers
- Someone else hires the IT People
- Someone else pay the rent of real estate

You are responsible for your configuring cloud services; someone else takes care of the rest.

Six Advantages and Benefits of Cloud Computing

Why go with a Cloud Provider over On-Premise?

1. **Trade capital expense for variable expense**
   No upfront-cost Instead of paying for data centers and servers
   Pay On-Demand Pay only when you consume computing resources

2. **Benefit from massive economies of scale**
   Usage from hundreds of thousands of customers aggregated in the cloud. You are **sharing the cost with other customers** to get unbeatable savings

3. **Stop guessing capacity**
   Eliminate guesswork about infrastructure capacity needs. I**nstead of paying for idle or underutilized servers**, you can scale up or down to meet the current need.

4. **Increase speed and agility**
   Launch resources **within a few clicks in minutes** instead of waiting days or weeks of your IT to implement the solution on-premise

5. **Stop spending money on running and maintaining data centers**
   Focus on your own customers, rather than on the heavy lifting of racking, stacking, and powering servers

6. **Go global in minutes**
   Deploy your app in **multiple regions around the world with a few clicks**. Provide lower latency and a better experience for your customers at minimal cost.

AWS Global Infrastructure

Where does all this Cloud Computing Run?

81 Availability Zones within **25 Geographic Regions** around the world
Way More **Edge Locations** than AZs!

AWS serves over **a million** active customers in **more than 190 countries**

Steadily **expanding** global infrastructure to help customers achieve lower latency and higher throughput

Regions physical location in the world with multiple Availability Zones

Availability Zones one or more discrete data centers

Edge Location datacenter owned by a trusted partner of AWS

Regions

A **geographically distinct** location which has multiple datacenters (AZs)

Every region is **physically isolated** from and independent of every other region in terms of location, power, water supply

Each region has at least ✌️ two AZs

AWS largest region is **US-EAST**

NEW services almost always become available first in **US-EAST**

Not all services are available in all regions

US-EAST-1 is the region where you see all your billing information

Availability Zones (AZs)

An AZ is a datacenter owned and operated by AWS in which AWS services run

Each region has at least ✌️ two AZs

AZs are represented by a Region Code, followed by a letter identifier eg. **us-east-1a**

Multi-AZ Distributing your instances across multiple AZs allows failover configuration for handling requests when one goes down.

< 10ms latency between AZs

Edge Locations

Get Data Fast or Upload Data Fast to AWS

An Edge Location is a datacenter owned by a trusted partner of AWS which has a **direct connection** to the AWS network.

These locations serve requests for **CloudFront** and **Route 53**. Requests going to either of these services will be routed to the nearest edge location automatically.

S3 Transfer Acceleration traffic and **API Gateway** endpoint traffic also use the AWS Edge Network.

This allows for **low latency** no matter where the end user is geographically located.

GovCloud (US)

AWS GovCloud Regions allow customers to host sensitive **Controlled Unclassified Information** and other types of regulated workloads.

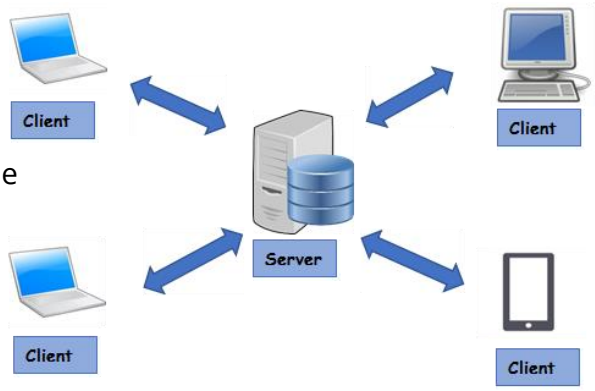
GovCloud Regions are only operated by employees who are U.S. citizens, on U.S. soil.

They are **only** accessible to U.S. entities and root account holders who pass a screening process

Customers can architect secure cloud solutions that comply with:

- 🇺🇸 FedRAMP High baseline
- 🇺🇸 DOJ's Criminal Justice Information Systems (CJIS) Security Policy
- 🇺🇸 U.S. International Traffic in Arms Regulations (ITAR)
- 🇺🇸 Export Administration Regulations (EAR)
- 🇺🇸 Department of Defense (DoD) Cloud Computing Security Requirements Guide

Core Services



Client–server model is a distributed application structure that partitions tasks or workloads between the providers of a resource or service, called servers, and service requesters,
called clients.

Amazon EC2 (Elastic Compute Cloud) is a web service interface that provides resizable compute capacity in the AWS cloud. It is designed for developers to have complete control over web-scaling and computing resources. EC2 instances can be resized and the number of instances scaled up or down as per our requirement.

EC2 – Pricing Model

On-Demand Least Commitment

- low cost and flexible
- only pay per hour
- short-term, spiky, unpredictable workloads
- cannot be interrupted
- For first time apps

Spot upto 90% Biggest Savings

- request spare computing capacity
- flexible start and end times
- Can handle interruptions (server randomly stopping and starting)
- For non-critical background jobs

Reserved upto 75% off Best Long-term

- steady state or predictable usage
- commit to EC2 over a 1 or 3 year term
- Can resell unused reserved instances

Dedicated Most Expensive

- Dedicated servers
- Can be on-demand or reserved (upto 70% off)
- When you need a guarantee of isolate hardware (enterprise requirements)

EC2 – On-Demand Instances

Least Commitment

When you launch an EC2 instance it is by default using **On-Demand** Pricing

On-demand has **no up-front payment** and **no long-term commitment**

Launch Instance

You are charged by the **hour** or by the **minute** (varies based on EC2 Instance Types)

On-Demand is for applications where the workload is for **short-term**, **spikey** or **unpredictable**.
When you have a **new app** for development or you want to run experiment.

EC2 – Reserved Instances (RI)

Best Long-term

Designed for applications that have a **steady-state**, **predictable usage**, or require **reserved capacity**.
Reduced Pricing is based on **Term** x **Class Offering** x **Payment Option**

Platform Linux/UNIX | Tenancy Default | Offering Class Standard
Instance Type t2.micro | Term 12 months - ... | Payment Option Partial Upfront | Search

Seller	Term	Effective Rate	Upfront Price	Hourly Rate	Payment Option	Offering Class	Quantity Available	Desired Quantity	Normalized units per hour	
AWS	36 months	$0.005	$66.00	$0.002	Partial Upfront	standard	Unlimited	1	0.5	Add to Cart

Standard Up to **75%** reduced pricing compared to on-demand.
Cannot change RI Attributes.

Convertible Up to **54%** reduced pricing compared to on-demand.
Allows you to change RI Attributes if greater or equal in value.

Scheduled You reserve instances for specific time periods eg. once a week for a few hours. Savings vary

Terms
You commit to a **1 Year** or **3 Year** contract.
The longer the term the greater savings.

Payment Options
All Upfront, **Partial Upfront**, and **No Upfront**
The greater upfront the great the savings

RIs can be **shared between multiple accounts** within an org

Unused RIs can be sold in the **Reserved Instance Marketplace**

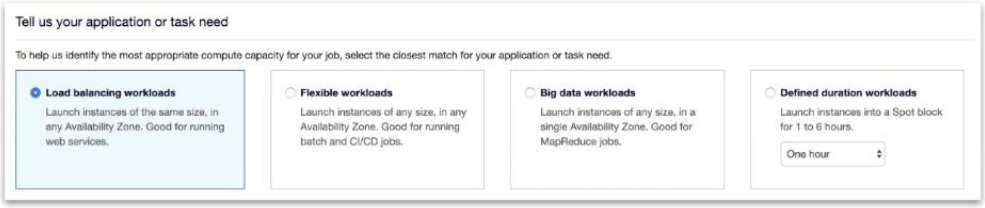
EC2 – Spot Instances

Biggest Savings

AWS has **unused compute capacity** that they want to maximize the utility of their idle servers.
It's like when a hotel offers discounts for to fill vacant suites or planes offer discount to fill vacant seats.

Spot Instances provide a discount of **90%** compared to On-Demand Pricing
Spot Instances can be terminated if the computing capacity is needed by on-demand customers.

Designed for applications that have flexible start and end times or applications that are only feasible at **very low** compute costs.



AWS Batch is an easy and convenient way to use Spot Pricing

Termination Conditions
Instances can be terminated by AWS **at anytime**
If your instance is **terminated by AWS**, **you don't get charged** for a partial hour of usage.
If **you terminate** an instance **you will still be charged** for any hour that it ran.

EC2 – Dedicated Host Instances

Most Expensive

Designed to meet regulatory requirements. When you have strict **server-bound licensing** that won't support multi-tenancy or cloud deployments.

Multi-Tenant vs Single Tenant

When multiple customers are running workloads on the same hardware. **Virtual Isolation** is what separate customers. (think apartment)

When a single customer has dedicated hardware. **Physical Isolation** is what separates customers (think house)

Offered in both **On-demand** and **Reserved** (70% off on-demand pricing)

Enterprises and **Large Organizations** may have security concerns or obligations about against sharing the same hardware with other AWS Customers.

EC2 Pricing – *CheatSheet*

- EC2 has for 4 pricing models **On-Demand**, **Spot**, **Reserved Instances (RI)** and **Dedicated**
- **On-Demand** (least commitment)
  - low cost and flexible
  - only pay per hour
  - **Use case:** short-term, spiky, unpredictable workloads, first time apps
  - Ideal when your workloads cannot be interrupted
- **Reserved Instances** upto 75% off (Best long-term value)
  - **Use case:** steady state or predictable usage
  - Can resell unused reserved instances (Reserved Instance Marketplace)
  - Reduced Pricing is based on **Term** x **Class Offering** x **Payment Option**
  - **Payment Terms**: 1 year or 3 year
  - **Payment Options**: All Upfront, Partial Upfront, and No Upfront
  - **Class Offerings**
    - **Standard** Up to 75% reduced pricing compared to on-demand. Cannot change RI Attributes.
    - **Convertible** Up to 54% reduced pricing compared to on-demand. Allows you to change RI Attributes if greater or equal in value.
    - **Scheduled** You reserve instances for specific time periods eg. once a week for a few hours. Savings vary

EC2 Pricing – *CheatSheet*

- **Spot Pricing** upto 90% off (Biggest Savings)
  - request spare computing capacity
  - flexible start and end times
  - **Use case:** Can handle interruptions (server randomly stopping and starting)
  - **Use case:** For non-critical background jobs
  - Instances can be terminated by AWS **at anytime**
  - If your instance is **terminated by AWS**, **you don't get charged** for a partial hour of usage.
  - If **you terminate** an instance **you will still be charged** for any hour that it ran.
- **Dedicated Hosting** (Most Expensive)
  - Dedicated servers
  - Can be on-demand or reserved (upto 70% off)
  - **Use case:** When you need a guarantee of isolate hardware (enterprise requirements)

EBS - Elastic Block store: Storage unit for your EC2 instances, HDD or SSD. Can create snapshots and change in size if needed

S3 - Simple Storage Service: Fully managed durable storage service. Virtually unlimited objects, securely access from anywhere.

- Files places into buckets, that have globally unique names within a given region.
- Billed for what you use.

VPC - Virtual Private Cloud: A virtual network in AWS cloud, allowing complete network control with several layers of security. Other AWS services (such as EC2) are deployed into this VPC

- Live within a Region
- **Subnets** divide a VPC and allow it to span multiple AZs.
- **Route tables** control traffic going of the subnet
- **Internet Gateways** allow access to the internet from VPCs
- **NAT gateways** allow private subnet resources access to internet
- **Network Access Control Lists (NACL**) control access to subnets, stateless

AWS Security groups: Act as built in built-in firewalls, control how accessible instances are and what traffic is allowed and denied. Default all incoming is denied and outgoing allowed.

Integrated Services

Application load balancer
Balance incoming traffic to the correct application. Additional protocols, access logs, cloud watch, health checks.

Auto Scaling
Helps ensure correct number of EC2 instances available to handle load. Automatically scale in or out depending on load based on your settings.

- **launch configuration**: EC2 types
- **auto scaling group**: Where & how - VPC, min, max desired
- **auto scaling policy**: When to scale up/down - dynamic w/ cloudwatch or scheduled

Route53
DNS service (domain name system), translating "example.com" into "54.85.178.219".

RDS - Relations Database Services
Managed service to setup databases. You manage the data, AWS the rest. Ability to configure Multi-AZ for availability++ & durability++.

Challenges with traditional DBs: Maintenance, patching, backups, availability, scalability, security

Database instance: Type of database (MySQL, Aurora, SQL Server, PostgreSQL, mariaDB, oracle), underlying CPU/memory

- **Read replica**: Updates to the database are automatically replicated in the secondary instance read replica. Can be created in a different region for disaster resilience and better global availability

Other:

AWS Lambda: Event driven, server less compute service. No servers to manage, continues scaling, pay for each second used. Connective tissue between AWS services.

Elastic Beanstalk: PaaS. Easily provision resources for your application.

SNS - Simple Notification Service: Send messages/emails/notifications to individuals/groups based on events.

CloudWatch: Monitor AWS resources and applications in real time. CPU, data transfer, Disk IO, log files, set alarms, react to changes.

CloudFormation: Simplifies task of repeatedly creating groups of related resources by using JSON/YAML template files. Infrastructure through code.

Architecture

Well architected framework

There to help customers. Guide to help you with the design of your architecture. 5 pillars are:

Security:
Ability to protect systems while delivering value through risk assessment and mitigation. Secure. IAM (Only authorized users can access), detective controls, infrastructure protection, data protections

Principles: Implement at all layers, traceability, least privilege, secure your system (shared responsibility), automate

Reliability:
Ability to recover from failure & to meet demand. Foundations, change management (know how change impacts systems), failure management. Test recovery procedures, automatically recover, scale horizontally, stop guessing capacity, manage change in automation.

Performance efficiency:

Select the best solution, review when new things come out, monitor performance and know the tradeoffs for your solution. Democratize advanced technologies, go server less, experiment

Cost optimization: Use cost effective resources, match supply with demand, increase cost awareness, optimize over time. Adopt a consumption model, measure efficiency, reduce spending, use managed services and analyze and attribute cost

Operational excellence: Manage and automate changes, respond to events, define the standards to manage daily operations.

Fault tolerance: Ability of a system to remain operational**.** SQS, S3**,** RDS - Auto backup, multi-AZ

Highly available: Ensure systems are always accessible**.** Elastic load balancers, elastic IP, route53, auto scaling, Cloudwatch

Web hosting Can host many types of web applications. AWS allows you to scale as your business grows and to meet sudden spikes of demand**.** Traditional architecture has no way of meet these demands on the fly, but need time and up-front money to setup.

Security

Shared responsibility model: AWS secures the infrastructure. You secure what you provision and build.

- **AWS**: Physical, network, hypervisor, OS
- **Middle**: EC2
- **You**: OS (choose it), Application, User Data

IAM - Identity and Access management

- **Users**: Permanent named operator (human or machine). E.g. John Doe
- **Groups**: Collection of users
- **Roles**: Operator with temporary authentications. E.g. Developer, admin, etc.
- **Policy document**: Permissions (allow/explicit deny) that attach to users/groups/roles in JSON. Defines what can and cannot be done.

Amazon Inspector: Automated security assessment service. Vulnerability and deviations in best practices.

AWS Shield: Free (and premium) managed Dos/DDos protection service safeguarding applications on AWS.

Security Compliance: Openly publish certifications, get legal/regulatory support, regularly undergoes audits. 3 components:

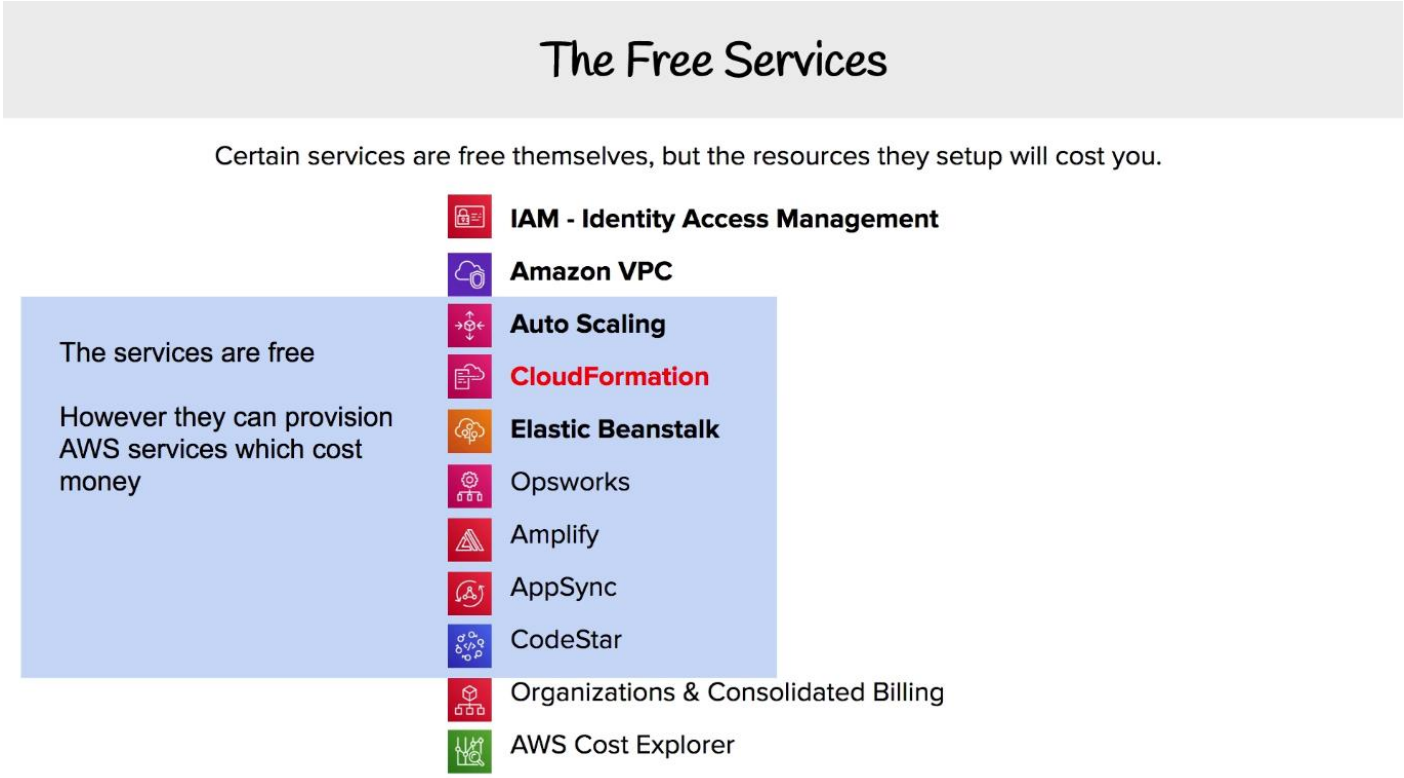
- Risk management: Establish frameworks, policies, maintenance, training and reviews.
- Control environment
- InfoSec: Confidentiality, availability

Pricing

Pricing fundamentals: Only pay for services you consume. Pay as you go, pay less if you reserve, pay less if you order more, and pay less as AWS grows. Reserved instances pay all or partially upfront and can save up to 75% of on-demand.

Cost fundamentals: Pay for compute, storage and outbound data. No charge for inbound data or between services in regions.

- **EC2**: Hourly cost & Data Load balancer processing.
    - Auto Scaling, Elastic IP and Cloudwatch is free (unless w/ detailed monitoring)
- **S3**: Type of storage, number & size of object, number & types of requests
- **EBS**: Type of storage, per snapshot, outbound data transfers tiered
- **RDS**: Clock hours of servers, DB characteristics (engine, size, memory), type (on-demand, reserved), number of Availability zones. Free to backup for active DB, pay per GB/month for terminated DB.
- **CloudFront**: Requests and data transfer out



AWS Marketplace

AWS Marketplace is a curated digital catalogue with **thousands** of software listings from independent software vendors.

Easily find, buy, test, and deploy software that already runs on AWS.

The product can be **free** to use or can have an **associated charge**. The charge becomes part of your AWS bill, and once you pay, AWS Marketplace pays the provider.

The sales channel for ISVs and Consulting Partners allows you to **sell your solutions** to other AWS customers.

Products can be offered as

- Amazon Machine Images (AMIs)
- AWS CloudFormation templates
- Software as a service (SaaS) offerings
- Web ACL
- AWS WAF rules

AWS Support Plans

Basic	Developer	Business	Enterprise
Email Support only For **Billing and Account**	Tech Support via **Email** ~24 hours until reply		
	No third party support	Tech Support via **Chat, Phone** Anytime 24/7	
	General Guidance		< 24 hrs
	System Impaired		< 12 hrs
		Production System Impaired	< 4 hrs
		Production System **DOWN!**	< 1 hrs
			Business-Critical System **DOWN!** < 15m
			😎 Personal Concierge
			🤓 TAM
7 Trusted Advisor Checks		**All** Trusted Advisor Checks	
$0 USD /month	**$20 USD** /month	**$100 USD** / month	**$15,000 USD** / month

AWS Trusted Advisor

FREE - 7 Trusted Advisor Checks
Business, Enterprise - All Trusted Advisor Checks

Advises you on **security**, **saving money**, **performance**, **service limits** and **fault tolerance**

Think of it like an automated checklist of best practices on AWS

AWS Trusted Advisor

Cost Optimization

Amazon EC2 Reserved Instances Optimization
Low Utilization Amazon EC2 Instances
Underutilized Amazon EBS Volumes
Amazon EC2 Reserved Instance Lease Expiration
Amazon RDS Idle DB Instances
Amazon Route 53 Latency Resource Record Sets
Idle Load Balancers
Unassociated Elastic IP Addresses
Underutilized Amazon Redshift Clusters

Performance

CloudFront Alternate Domain Names
Amazon EBS Provisioned IOPS (SSD) Volume Attachment Configuration
Amazon EC2 to EBS Throughput Optimization
Amazon Route 53 Alias Resource Record Sets
CloudFront Content Delivery Optimization
CloudFront Header Forwarding and Cache Hit Ratio
High Utilization Amazon EC2 Instances
Large Number of EC2 Security Group Rules Applied to an Instance
Large Number of Rules in an EC2 Security Group
Overutilized Amazon EBS Magnetic Volumes

Security

AWS CloudTrail Logging
IAM Password Policy
MFA on Root Account
Security Groups - Specific Ports Unrestricted
Security Groups - Unrestricted Access
Amazon S3 Bucket Permissions
IAM Access Key Rotation
Amazon EBS Public Snapshots
Amazon RDS Public Snapshots
Amazon RDS Security Group Access Risk
Amazon Route 53 MX Resource Record Sets and Sender Policy Framework
CloudFront Custom SSL Certificates in the IAM Certificate Store
CloudFront SSL Certificate on the Origin Server
ELB Listener Security
ELB Security Groups
Exposed Access Keys
IAM Use

AWS Trusted Advisor

Fault Tolerance

Amazon EBS Snapshots
Amazon RDS Multi-AZ
Amazon S3 Bucket Logging
Amazon S3 Bucket Versioning
Amazon Aurora DB Instance Accessibility
Amazon EC2 Availability Zone Balance
Amazon RDS Backups
Amazon Route 53 Deleted Health Checks
Amazon Route 53 Failover Resource Record Sets
Amazon Route 53 High TTL Resource Record Sets
Amazon Route 53 Name Server Delegations
Auto Scaling Group Health Check
Auto Scaling Group Resources
ELB Connection Draining
ELB Cross-Zone Load Balancing
Load Balancer Optimization
VPN Tunnel Redundancy
AWS Direct Connect Connection Redundancy
AWS Direct Connect Location Redundancy
AWS Direct Connect Virtual Interface Redundancy
EC2Config Service for EC2 Windows Instances
ENA Driver Version for EC2 Windows Instances
NVMe Driver Version for EC2 Windows Instances
PV Driver Version for EC2 Windows Instances

Service Limits

Auto Scaling Groups
Auto Scaling Launch Configurations
CloudFormation Stacks
DynamoDB Read Capacity
DynamoDB Write Capacity
EBS Active Snapshots
EBS Active Volumes
EBS Cold HDD (sc1) Volume Storage
EBS General Purpose SSD (gp2) Volume Storage
EBS Magnetic (standard) Volume Storage
EBS Provisioned IOPS (SSD) Volume Aggregate IOPS
EBS Provisioned IOPS SSD (io1) Volume Storage
EBS Throughput Optimized HDD (st1) Volume Storage
EC2 Elastic IP Addresses
EC2 On-Demand Instances
EC2 Reserved Instance Leases
ELB Active Load Balancers
IAM Group
IAM Instance Profiles
IAM Policies
IAM Roles
IAM Server Certificates
IAM Users
Kinesis Shards per Region
RDS Cluster Parameter Groups
RDS Cluster Roles
RDS Clusters
RDS DB Instances
RDS DB Parameter Groups
RDS DB Security Groups
RDS DB Snapshots Per User
RDS Event Subscriptions
RDS Max Auths per Security Group
RDS Option Groups
RDS Read Replicas per Master
RDS Reserved Instances
RDS Subnet Groups
RDS Subnets per Subnet Group
RDS Total Storage Quota
Route 53 Hosted Zones
Route 53 Max Health Checks
Route 53 Reusable Delegation Sets
Route 53 Traffic Policies
Route 53 Traffic Policy Instances
SES Daily Sending Quota
VPC
VPC Elastic IP Address
VPC Internet Gateways

Consolidated Billing

One bill for all of your accounts

Consolidate your billing and payment methods **across** multiple AWS accounts into **one bill**

For billing AWS treats all the accounts in an organization as if they were one account.

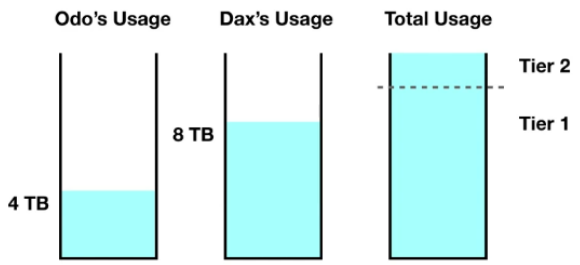
You can designate one **master account that pays the charges** of all the other **member accounts.**

Consolidated billing is offered at no additional cost!

Use **Cost Explorer** to visualize usage for consolidated billing

Consolidated Billing – Volume Discounts

AWS has **Volume Discounts** for many services
The more you use, the more you save.
Consolidated Billing lets you take advantage of Volume Discounts



Data Transfer	
First 10 TB	$0.17 per GB
Next 40 TB	$0.13 per GB

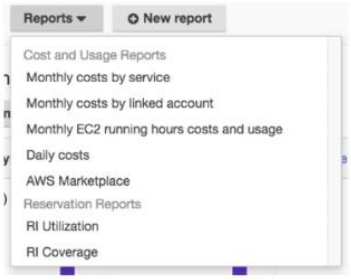
Odo	(4*1024)*0.17	= $696.32
Dax	(8*1024)*0.17	= $1392.64
Unconsolidated	696.32+1392.64	= **$2088.96**
Consolidated	((10*1024)*0.17)+((2*1024)*0.13)	= **$2007.04**

1 TB = 1024 GB

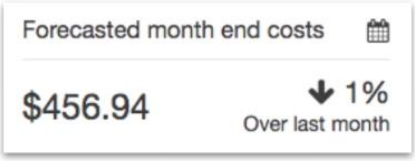
AWS Cost Explorer

AWS Cost Explorer lets you **visualize**, **understand**, and **manage** your AWS costs and usage over time.
If you are have multiple AWS accounts within an AWS Organization costs will be consolidated in the **master account**.

Default reports help you gain insight into your cost drivers and usage trends.

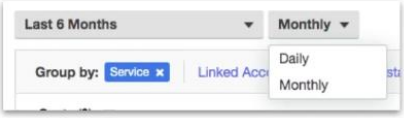


Use **forecasting** to get an idea of future costs

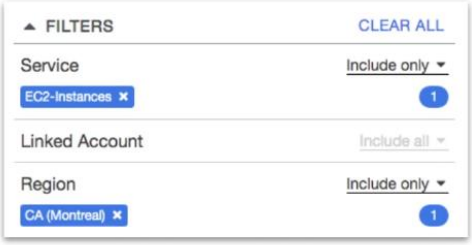


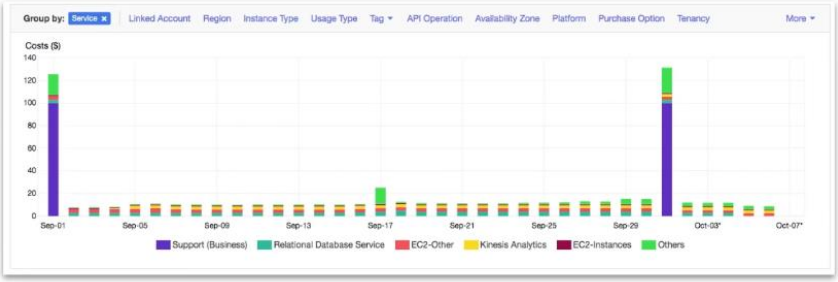
AWS Cost Explorer

Choose if you want to view your data at a **monthly** or **daily** level of granularity



Use **filter** and **grouping** functionalities to dig even deeper into your data!





AWS Budgets

first two budgets are **free** of charge
Each budget is **$0.02** per day **~0.60 USD / mo**
20,000 budgets limit

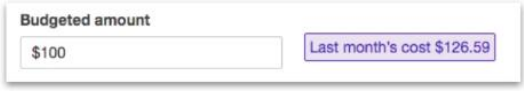
Plan your **service usage, service costs and Instance reservations**

Think of it like an billing alarms on steroids

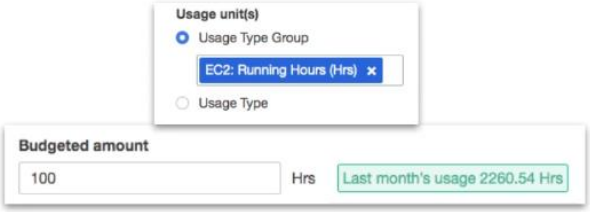
AWS Budgets

AWS Budgets give you the ability to setup alerts if you **exceed** or are **approaching** your defined budget

Create **Cost**, **Usage** or **Reservation** Budgets



Can be tracked at the **monthly**, **quarterly**, or **yearly levels**, with customizable start and end dates



Alerts support **EC2**, **RDS**, **Redshift**, and **ElastiCache** reservations.

Budget based on a fixed cost or plan your upfront based on your chosen level

Can be easily manage from the **AWS Budgets** dashboard or via the **Budgets API**.

Get Notified by providing an email or **Chatbot** and threshold how close to the current or forecasted budget

Things to Remember

O - Multitenancy: sharing of underlying hardware.

O - Different use of instances in EC2

General Purpose: different workloads, web services and code repositories, good balance of memory compute, memory and networking services

- Application servers
- Gaming servers
- Backend servers for enterprise applications
- Small and medium databases

Compute optimized: intensive high permanence, gaming servers and scientific modeling (batch workload)

Accelerated Compute Optimized: float numbers calculations, graphic processing, data pattern matching as they use hardware resources. game, and application streaming.

Memory optimized: (large datasets in memory) memory intensive tasks (pre load data)

Storage optimized: high permanence of locally stored data (input output operations in one sec)

O - Elastic load balancing: Automatically distribute application traffic through EC2 instances and help of auto scaling provides high performance and availability.

O - Loosely coupled architecture (doesn't affect the other component of system)

O - SQS: simple queue service, send, store and deliver.

O - SNS: simple notification service, end users get notifications.

O - Monolithic Applications fail, because tightly coupled architecture instead micro service architecture is preferred

O - Containers: set your application dependencies and code into a single object to avoid environment changes in deployment.

O - Container orchestration: helps to manage, deploy and scale container applications

O - Amazon elastic container service (ECS): highly scalable and highly performance container management system, supports Dockers community

- furgate launch type (app in container)
- EC2 lunch type (customizing to servers)

O - Amazon Elastic Kubernetes Service(EKS): a fully managed service and deploy that run containers applications on scale.

O - ECR: data stored repositories, Docker container images

O - AWS furgate: server less compute engine, works for both, manage infrastructure for your cluster management, remove your need of management of clusters and servers

O - AWS lambda: only code and configurations, not recommended for deep learning, less than 15 minutes' workload.