AWS Cloud Practitioner Quick Notes

Base concepts

AWS regions

- Isolated between each other
- ends with a number (e.g. eu-west-1)
- Have at least 2 AZ

Availability Zones

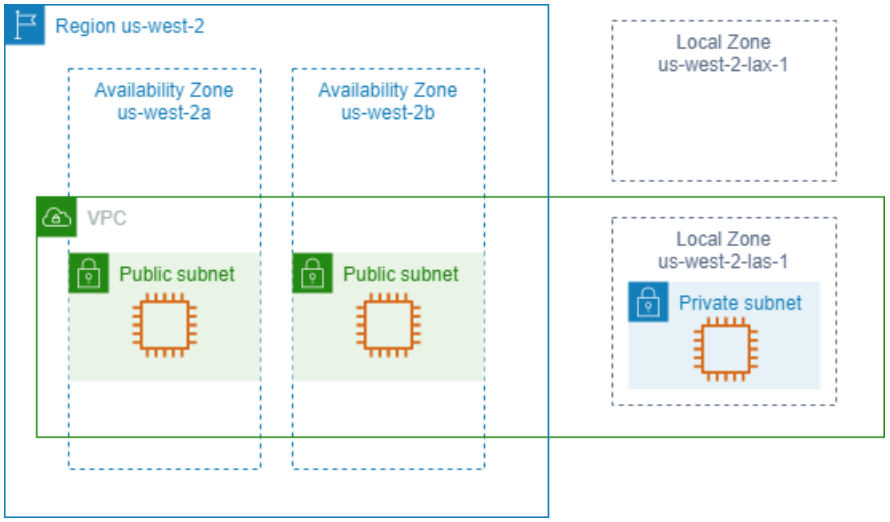
- children of regions
- include ONE or more data centers with redundant networking, power and connectivity
- suffix with a b c (eu-west-1a)
- connected with high bandwidth, ultra-low latency connections

Edge locations

- the places where data are cached to reduce latency
- used by CloudFront to cache copies

Local zones

- An extension of AWS
- Multi-tenants (used by multiple parties, not just one company like Outposts)



Global services

Services that are not bound to any region

- IAM
- CloudFront
- Route 53
- S3

AWS Outposts

- Provide local access to AWS-managed infra
- Build and run apps on-premises
- Deploy on customer site
- Managed by AWS

Wavelength zones

- Let developers build apps with ultra-low latencies to 5G and users

VPC

- Virtual Private Cloud
- Span across multiple availability zones but stay within on region

Internet gateway

- Connect a public subnet to the internet

Subnet

- Public and private
- Reside within an AZ

NAT gateway & NAT instances

- NAT gateway is Managed by AWS
- NAT instances are managed by the user
- Both allow a private subnet to connect to the internet

Security Group

- Can only have ALLOW rule
- Control access to EC2 or Elastic Network Interface (ENI)
- A kind of firewall
- Return traffic is automatically allowed, no matter the rules (Stateful)

Network Access Control List (NACL)

- Can contain rules for IP only
- can have ALLOW/DENY rules
- Filter traffic in and out Subnet
- Return traffic must be explicitly allowed

VPC Flow logs

- Provides info about IP traffic in and out of interfaces
- Can store in S3/CloudWatch logs
- Must be enabled manually

VPC Peering

- Connect two VPC
- CIDR must not overlap
- Only work with two VPC
- VPC can be in different Regions

VPC Endpoints

- Connect to AWS services using a private network

VPN CloudHub

- operates on a simple hub-and-spoke model that you can use with or without a VPC
- Use this approach if you have multiple branch offices and existing Internet connections and would like to implement a convenient, potentially low-cost hub-and-spoke model for primary or backup connectivity between these remote offices

Direct Connect

- Physical connection from the on-premise server to AWS
- Takes a long time to provision because AWS needs to build a physical line from on-premise to AWS cloud
- Private & secure & Expensive

Site-to-Site VPN

- Connect user's on-premise VPN to AWS
- Using internet connection (not private)
- The on-premise server must have a Customer Gateway
- AWS must use Virtual Private Gateway

Transit gateway

- Transitive between many VPC
- Connect, join all the above types of connect

IAM (Identity and Access Management)

Overview

- least privilege principle

Access Advisor

- Need more details

Cognito

- Identity management
- Let aws user add sign-in, signup to web and mobile apps quickly
- Support OIDC, SAML 2.0

Directory services

- Managed Microsoft active directory

Account alias

- Customize login url
- used as the account ID

IAM Root user

- created by default, should not share, has the most power
- MFA should be enabled

IAM Users

- can be part of 0 to n groups

IAM Identities

- Users
- Groups
- Roles

IAM Group

- Contains users only

IAM Policies

- Types
    - Identity-based: attach managed and inline policies to IAM identities → grant permission to identities
    - Resource-based policies: attach inline policies to resources.
    - Permissions boundaries:
        - Define maximum permissions that the identity-based policies can grant to an entity
        - Does not grant specific permission
        - Do not define the maximum permissions that a resource-based policy can grant
    - Organizations SCPs:
        - Use with AWS Organizations
        - Define the maximum permissions for account members of and organization or organization unit
        - Do not grant a specific permission
    - Access Control Lists (ACLs)
        - The only policy that doesn't use JSON structure
        - Similar to resource-based policies
    - Session policies
- Used to manage access by creating policy and attach to IAM identities
- Define permissions
- When creating a new IAM Policy, these are the required fields:
    - PolicyName

    - PolicyDocument
- AWS managed policies
    - Standalone policies created and managed by AWS
    - Provide permissions for common use cases
- Customer managed policies
    - Standalone policies that are created and managed by the user
- Inline policies
    - embedded in an IAM identity
- Inline vs managed policy

	Inline	**Managed**
Reusability	NO	YES
Central change management	NO	YES
Versioning and rollback	NO	YES
Delegating permissions management	NO	YES
maintain a strict one-to-one relationship between a policy and the identity to which it is applied	YES	NO

Roles

- Not used by users but by programs
- For example, assign a role to EC2 to do some tasks on aws

Strong password using password policy

- Require length
- prevent-reuse
- force rotation

CloudShell

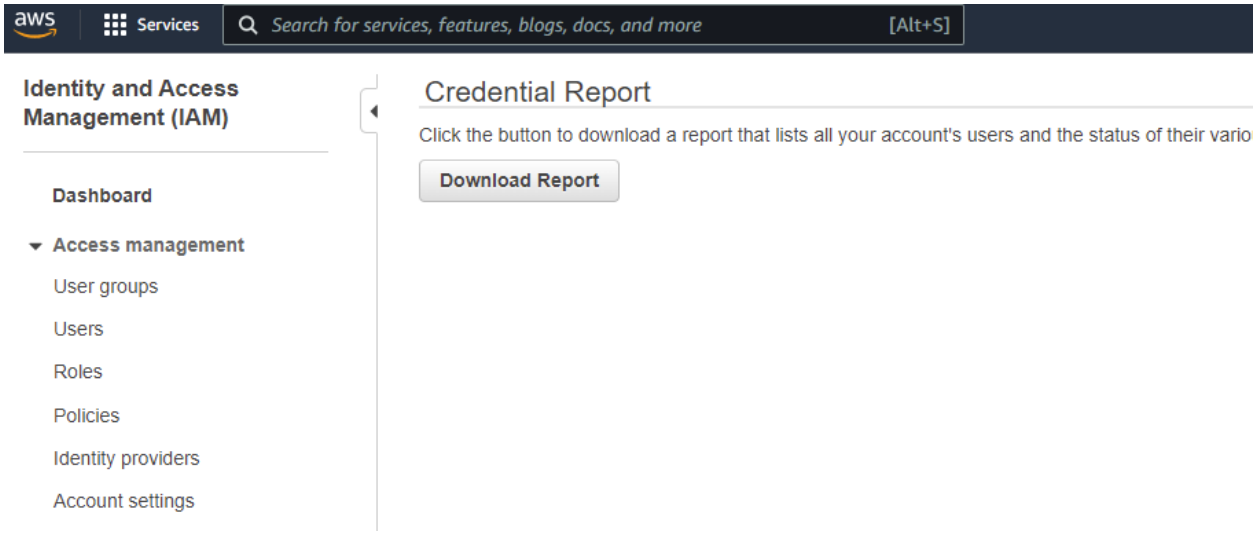
- Terminal in the cloud (similar to google's shell)
- Terminal open in browser

Role

- Consists of permissions
- Use to create custom combinations of permissions

IAM Credentials report

- Export all accounts to CSV highlighting which account has MFA enabled etc.



EC2

Types of instances

- General purposes
- Compute optimized (Prefix C)
- Memory optimized (Prefix R/X)
- Storage optimized (Prefix I/D)

Dedicated Host

- allow you to use your existing per-socket, per-core, or per-VM software licenses that are bound to VMs, sockets, or physical cores, subject to your license terms

EC2 On-demand

- Pay per second or hours
- Minimum 60 seconds

Purchase options

- On-demand (regular case)
- Spot instance (cheap but not reliable)
- Reserved instance (big discount but need long-term commitment: 1 year or 3 years)
    - 1 to 3 years
    - up to 72% discount
    - Options
        - Standard: Can change AZ, instance size (Linux), networking type
        - Convertible: Can change AZ, instance size, OS, payment options, have lower discount than standard
- Dedicated hosts (most expensive, physical server, can request location placement, not sharing hardware)
- Dedicated instance (virtual, cannot require location placement)
- Capacity reservation (no discount, only reserve resources)

- The minimum charge for an on-demand instance is 60 seconds (so if you use less than 60, you still charged 60s)

User data

- Bash script that run on 1st start-up (once) of EC2 instance

Security group

- Consider as firewall or access control rules
- Control access to ports
- Allowed IP ranges
- Control inbound/outbound traffic
- can be applied to multiple EC2
- cannot share between vpc/regions

Instance roles

- Attach IAM role to the instance
- Use to configure aws access inside ec2

EBS & EFS & AMI

Block storage	File storage	Object storage
EBS	EFS	S3
Instance store	FSx	

EBS

- to attach an EBS to an EC2, they must be on the same AZ
- Make a backup using a snapshot (Do not need to detach) – copy to other regions, takes a long time to recover if using cheap storage. Use recycle bin to avoid accidental delete
- Pricing
    - Pay for what user provision
    - Storage & IOPS
    - Pay for snapshots
    - Pay for snapshot restore on Archive (free for the standard)

AMI

- Amazon Machine image
- Pre-installed, customized EC2 images (like a docker image)
- Public AMI (AWS provide) – Custom API (create yourself) – Download from the marketplace
- To create an AMI, start an EC2, install all the packages and build the image. This also creates EBS snapshot
- AMI and EC2 must be in the same region

EC2 Image builder

- Free service
- Only pay for underlying services (storage)

Instance store

- Ephemeral storage, not suitable for persistent data
- Fast because resides on the same physical machine with EC2
- Cannot attach/reattach to different EC2
- Data lost when disk failure/instance hibernate

EFS

- Serverless file system
- You pay only for the storage you use, for read and write access to data stored in Infrequent Access storage classes, and for any provisioned throughput.
- can mount to many EC2
- Can share between AZ
- EFS-IA storage class for cheaper price. Suitable for files are not accessed every day (save up to 92%)
- EFS moves files to EFS-IA based on last time file is accessed (configurable using lifecyle Policy)

Edge Locations

- A type of hosting service
- Have a massive amount of storage devices, high-bandwidth networking
- Main usage are CloudFront and S3 Transfer Acceleration
- User can upload data to a Edge location and AWS will transfer the data using AWS internal networking for faster data transfer (not through public internet)

FSx

- Managed file system services
- Fast IO
- Backend connect directly to the FSx service

ECS & Fargate & ECR

ECS

- run tasks and services on a Amazon EC2 cluster that you manage

Fargate

- Serverless, plug and run
- No need EC2

ECR

- Container registry
- Like docker hub

Step Functions

- low-code, visual workflow service that developers use to build distributed applications
- Drag-n-drop to build multiple steps execution using Lambda and other AWS services

Lambda

- No need server
- Short execution
- Run on demand
- Automated scaling
- Pay per request & compute time
- Can run containers using Lambda Container Image
- Can be invoked by:
    - The lambda console
    - Function URL (https)
    - The Lambda API
    - AWS CLI
    - AWS SDK
    - AWS toolkits

API gateway

- Serveless
- Support REST and Websocket, security, authenticate, API limiting

Batch

- Launch EC2 to run many task
- User submit task in Batch queue
- Jobs are submitted using Docker image
- Resource provisioning is automatic

Lightsail

- For people with limited cloud experience
- Used for deploying simple applications (similar to cPanel?)
- No autoscaling
- Predictable pricing
- Keyword: blueprint

DB services

Deployments

- Read replicas:
    - distribute read load between servers
    - scalability

    - Can be within AZ to cross Regions
- Multi-Az
    - High availability
    - Aurora: async replication, non-Aurora: synchronous replication
    - Span at least 2 AZ within a region
    - Automatic failover to standby instance (Non-Aurora) or to read replica (Aurora)
- Multi-region
    - Disaster recovery & local performance
    - Async replication
    - Can read at all regions

RDS & Aurora

- Available options:
    - Aurora Mysql
    - Aurora PostgreSQL
    - MySQL
    - MariaDB
    - PostgreSQL
    - Oracle
    - SQL Server
    - on-premise with RDS and Outposts
- Aurora is not free
- RDS has free tier
- Aurora autoscale
- Can create, share and backup snapshot
- High availability
    - Can create replicas to load balance read/write
    - Multi-AZ to setup failover (only replication and read/write when the main instance dies)
    - Multiple-Region (Same as AZ but across region and has read-replica. good for disaster recovery and local quick access)

Elasticache

- In-memory database
- Managed redis/memcache
- Use for caching

DynamoDB

- Distributed NoSQL
- Fast & autoscaling
- DynamoDB Accelerator – DAX is an in-memory cache for DynamoDB
- Partition key
- All data in a single table (can create multiple tables)

Redshift

- based on postgresql
- use for analytics & data warehouse

- OLAP (Online analytical processing) not OLTP (transaction)
- Column-based, not row-based
- Has SQL console
- Not access every seconds

EMR

- Elastic Map-Reduce
- Hadoop
- Use for data processing, machine learning, big data...

Athena

- Works with S3
- Supports CSV, JSON, Avro, Parquet...
- Use for log processing
- Have SQL interface to query using SQL commands

Quicksight

- Business Intelligence tool
- Create dashboard (like grafana)

DocumentDB

- Actually Mongo

Neptune

- Graph database (like Neo4J)
- Highly available

QLDB

- Quantum Ledger DB
- For financial transaction
- Like blockchain but not decentralized

Managed Blockchain

- Hyperledger fabric
- Ethereum

DMS

- Database Migration Service
- Migrate this db to other db
- Support transporting different DB

Glue

- ETL service
- Takes input from S3, RDS...

ELB & ASG

ELB

- Forward request to multiple EC2
- Managed service
- Four types:
    - App load balancer (http level 7)
        - Supports path-based routing
        - Support dynamic host port mapping
    - Network Load Balancer (TCP – layer 4)
        - Handle millions of requests per second
        - Support dynamic host port mapping
    - Classic LB (layer 4 & 7) – about to retire soon
    - Gateway LB
        - Layer 3

Application LB (ALB)

- On layer 7 (application layer HTTP for example)
- Load balance traffic between EC2
- Auto-select healthy instances

Auto Scaling group (ASG)

- Auto scale number of EC2 under an ELB
- Three capacity: minimum/maximum/desired
- Can scale manually or dynamically
- Dynamic scaling:
    - Step scaling: base on resource usage (if CPU > x% ➜ increase)
    - Target tracking scaling: Keep Mem average usage < 50%
    - Schedule scaling
    - Predictive scaling: use machine learning to predict

Development services

CloudFormation

- Infrastructure as code (like terraform)
- Use to create resources
- There are templates to reuse
- Define code in YAML file (called templates)
- Free to use, the user pays for the infra

CDK

- Convert programming languages code to CloudFormation
- defines your cloud application resources using familiar programming languages.

Beanstalk

- For developer
- Helps developers run code faster by provisioning required infra (EC2/ELB/ASG...)
- Paas
- Supports many languages & deployment options. Also supports docker
- Can be used to monitor and check the health of an environment.
- Free to use, the user pays for the infra

CodeCommit

- Is a Git clone

CodeDeploy

- Works with EC2 or on-premises
- User must provision the computing service first (EC2 or Server)
- Deploy code to the connected server
- Deploy code automatically
- Need agents installed on the EC2 or on-prem servers

CodePipeline

- Connect CodeBuild and CodeDeploy
- CICD pipeline
- Compatible with many services

CodeBuild

- Build code
- Pay for build time

CodeArtifact

- Like maven repository
- or nexus repo
- Repository for dependencies

CodeStar

- UI for managing code-related issues
- UI for all Code* above
- Can be used to view pipeline
- Can integrate with Jira

Cloud9

- IDE on cloud

SSM (System manager)

- Manage EC2 & On-premise (fleet)
- Hybrid service

- Run command across all servers (Patch for example)
- To do this, all instances (EC2/on-premise) must have SSM agent installed)
- gives visibility and control of your infrastructure on AWS.
- provides a unified user interface so you can view operational data from multiple AWS services
- automate operational tasks across AWS resources.

SSM Session manager

- Allows the user to start a secure shell (terminal) to connect to EC2 or on-premise server without SSH

OpsWorks

- Managed Chef & Puppet service

Route53 & CloudFront & S3 Acceleration & Outposts & WaveLength & Local Zones

Route53

- DNS
- A record → IPv4
- AAAA Record → IPv6
- Can enable health check or not
- Routing policies
    - Simple routing policy: Point traffic to a single resource
    - Failover routing policy: Configure for active-passive failover
    - Geolocation routing policy: Route based on user's location (for example, US users go to this, EU users go to another...)
    - Latency routing policy: Route users based on the minimum latency
    - IP-based routing policy: Route users based on the IP
    - Weighted routing policy: Route to resources with a percentage (70% go to resource A, 30% go to B...)
    - Multivalue answer routing policy:
    - Geo Proximity routing policy:
- Features:
    - domain registration
    - DNS
    - Traffic flow
    - health checking
    - failover

CloudFront

- CDN
- Serves static content
- Edge locations
- DDoS protection

- Integrated with Cloud Shield & Application Firewall
- Origin could be S3 or EC2 server (http)
- Automatic replication to serve static content everywhere

Outpost

- An extension of AWS running on on-premise cloud
- Placed in client's data center

Local Zones

- Place AWS resources closer to client
- For example, one AZ in the city center, then there are Local Zones at surrounding districts
- Another AWS extension, not enabled by default
- Managed by AWS

WaveLength

- For 5G device

Global Accelerator

- No caching (unlike CloudFront)
- Has DDoS protection using Shield
- direct users to the nearest AWS Region that contains an endpoint for an application
- Users are provided with two global static public IPs as fixed entry point to applications

Cloud integrations

SQS

- Simple queue service
- Fully managed queue
- pull-based

SNS

- Pub/Sub
- Simple notification service
- Subscribers get all messages
- Send message to topics
- push-based

Kinesis

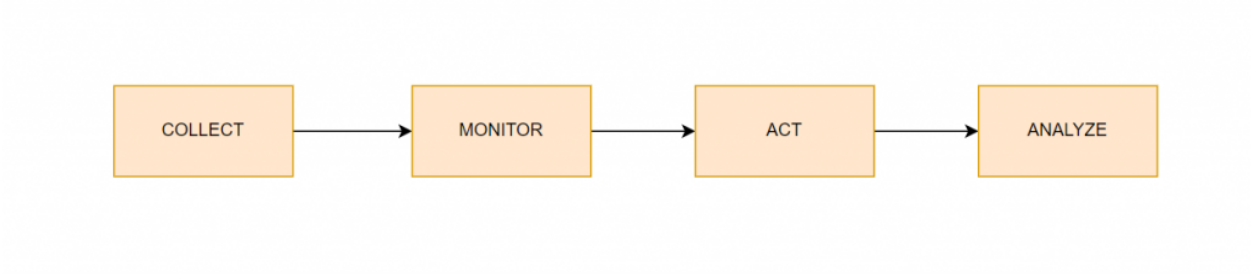
- Realtime big data streaming
- Collect process, analyze real-time streaming at any scale/amount
- Logs analysis, IOT...

Amazon MQ

- Managed Apache ActiveMQ
- Not scale as well as SNS or SQS

Cloud Monitoring

CloudWatch



- Basically a metrics repository. Metrics for all services. Users can put custom metrics
- Users can use the metrics to calculate statistics and view graphs in the Cloudwatch dashboard
- use alarms, logs, and events data to take automated actions and reduce mean time to resolution (MTTR)
- Can be used to create billing alarms
- Has alarms to trigger on any metrics (autoscaling, EC2 start-stop, SNS...)
- Logs: CloudWatch has a log agent on EC2 to collect logs
- Events have rules (for example, trigger when someone logs in). React to events in AWS, configurable, can trigger on schedule
- CloudWatch Container Insights to monitor, troubleshoot, and alert your containerized applications and microservices
- CloudWatch collects, aggregates, and summarizes compute utilization information such as CPU, memory, disk, and network data, as well as diagnostic information such as container restart failures, to help DevOps engineers isolate issues and resolve them quickly

CloudWatch alarms

- Automatically initiate actions on user's behalf
- Threshold: amount that trigger the alarm. Action: What to do when threshold is reached. Action is invoked on sustained changes. That means the changes has been over a certain period of time.
- Action is a notification sent to SNS topic or auto scaling topics

EventBridge

- connects applications using events
- define a filtering rule to filter events and route events to AWS service targets and API destinations (via HTTP endpoints)

CloudTrail

- Ensure compliance
- Enable by default
- Logs all activity across AWS (including action on console)

- AWS CloudTrail is a web service that records activity made on your account
- CloudTrail records user activity and API usage across AWS services as Events.
- CloudTrail Events help you answer the questions of “who did what, where, and when?”
- Events are stored for 90 days. Longer than that, a user needs to store it in S3 or Athena
- AWS CloudTrail Insights helps AWS users identify and respond to unusual activity associated with write API calls by continuously analyzing CloudTrail management events.
- CloudTrail can record the history of events/API calls made within you AWS account
- Enable encryption for all logs file
- Types of events:
    - Management events: actions on resources (like deleting an S3 bucket)
        - Read
        - Write
    - Data events: actions on data object such as deleting an S3 file object
        - S3 Read/Put/list
    - Insights events
        - Unusual events

CloudTrail Insights

- Identify unusual activities in AWS account

X-Ray

- Distributed Tracing
- Similar to zipkin
- gain insights into that data to identify issues and opportunities for optimization
- Mainly to identify performance issues in multiple services model (microservices, SOA...)

CodeGuru

- Automated code review
- performance recommendations
- CodeGuru Profiler detects rogue code

AWS Service Health Dashboard

- Monitor health of all services across regions

Personal Health Dashboard

- Provides alerts and guide to remedy the issues when there are problems with AWS may affect the user because she uses certain services.
- provides alerts and remediation guidance when AWS is experiencing events that may impact user

Networking

VPC

- Virtual private cloud
- Bound to region

Security

DDoS protection

- AWS Shield standard is available for all users, free of charge
- AWS Shield advanced: premium service with support

Security Token Service

- web service that enables you to request temporary, limited-privilege credentials for AWS Identity and Access Management (IAM) users or for users you authenticate (federated users)

WAF (Web Application Firewall)

- protect your web applications or APIs against common web exploits and bots
- monitor web requests that are forwarded to Amazon CloudFront distributions or an Application Load Balancer
- block common attack patterns, such as SQL injection or cross-site scripting
- pricing is based on how many rules you deploy and how many web requests your application receives

KMS (Key manager service)

- Data at rest (lying in storage)
- Data in transit (Moving from one location to another)
- AWS manages the encryption key
- Services that have encyption enabled by default:
    - CloudTrail logs
    - S3 Glacier
    - Storage Gateway

Cloud HSM (Hardware Security Module)

- AWS provides hardware
- User manage keys
- Tampering resistant
- Users can add and remove HSMs on-demand using the AWS Management Console and AWS API

Customer Master Keys

- Customer managed CMK
    - Managed by user
    - Users can define their own keys
- AWS managed CMK
    - AWS creates, and manages on the user's behalf
- AWS owned CMK
    - Used by AWS to encrypt cross accounts

ACM (AWS Certificate Manager)

- SSL certificate

AWS Secret Manager

- Encrypts secrets at rest using keys that customer owns and store in KMS
- Automatic secret rotation (for example, rotate database password)
- To store secrets
- integrate with RDS
- Rotating using the Lambda function

Artifact

- AWS Artifact is your go-to, central resource for compliance-related information that matters to you
- No cost
- self-service

GuardDuty

- Threat discovery using ML
- Anomaly detection
- Logs are the input (CloudTrail, K8s audit log...)
- Work with CloudWatch to create trigger/event in cases of finding
- Prevent cryptocurrency mining
- Supported findings on these services:
    - EC2
    - EKS
    - S3
    - IAM

Inspector

- vulnerability management service that continuously scans your AWS workloads for vulnerabilities and unintended network exposure
- Only for EC2 and container-related resources (ECR)
- Integrate with AWS security hub and findings are reported to Amazon event bridge

AWS Config

- assess, audit, and evaluate the configurations of your AWS resources to check for compliance
- Monitor and record configuration changes
- Receive SNS (Simple notification service) when changes made
- Need to enable and configure

Macie

- uses machine learning and pattern matching to discover and protect your sensitive(personal data for example) data in AWS
- Discover sensitive data
- HIIPA

Security Hub

- Centralized tool to manage security across AWS accounts
- Automated security checks
- Security Hub collects security data from across AWS accounts, services
- Analyze & identify threats

AWS Detective

- Analyze, investigate and identify the root cause of security issues or suspicious activities
- Input: AWS CloudTrail, VPC Flow Logs, and Amazon GuardDuty findings, and maintains up to a year of aggregated data for analysis.

AWS abuse

- Report abusive actions to AWS

Root user privileges

- Close account
- Change support plan
- Register as a seller in the marketplace

Machine learning services

Rekognition

- Face recognition
- OCR

Transcribe

- Speech to text

Polly

- Text to speech

Translate

- Translation
- Translate website

Lex

- Speech to text
- NLP
- Chatbot
- Build a contact center with Connect

Comprehend

- For NLP
- Extract meaning from text

SageMaker

- Build ML models

Forecast

- Build forecasting models

Kendra

- Document search service
- Extract text and make them searchable
- Built-in natural language understanding

Personalize

- Recommender system based on user's data

Textract

- Extract text from scanned documents
- PDF/Images included

Account management & Billing

Services that have reserve models

- EC2
- RDS
- Elsticache
- OpenSearch
- RedShift
- DynamoDB

Organizations

- Master account controls other accounts
- Billing for all sub-accounts
- API to create accounts
- Restricts other accounts' privileges
- Use tagging for billing purposes

Service Control policies

- Applied at organization account or account level
- Whitelist and blacklist IAM account
- Does not apply to the master account
- Applies to all users and Roles, including root
- Does not affect service-linked roles
- Denied by default, allow must be explicit
- Denied over allow, if parent policy denied but child policy allows, still denied

Consolidated billing

- Combined usage for better pricing
- A simple account with 1 bill

Control tower

- Setup and govern a secure and compliant multi-account AWS environment
- Automate set up an environment
- Automate ongoing policy management
- Detect violations
- Monitor compliance

Pricing models

- Pay as you go
- Save when reserving
- Use more cost less
- Pay less as AWS grows

Cost Explorer

- View and analyze your costs and usage
- Can identify underutilized EC2 instances
- Create a report to analyze cost and usage
- view data for up to the last 12 months, forecast how much you're likely to spend for the next 12 months
- Offer two saving plan reports

Cost and Usage Report

- Break down cost by the hour, day, month, by product, product resource, or by tags
- Publish billing reports to S3
- Update the report once a day in CSV
- Provides estimated charges
- Has API to create, retrieve and delete reports

Compute optimizer

- Use ML to reduce cost, improve performance
- analyzes the configuration and utilization metrics of AWS resources
- generates optimization recommendations to reduce the cost and improve the performance
- provides recent utilization metric data
- Supported resources:
    - EC2
    - EC2 ASG
    - EBS
    - Lambda

AWS budgets

- Send alarm when actual usage reaches a budget limit
- Recommend workload (which one is over/under optimization)
- Supported EC2, EC2 ASG, EBS, Lambda

Trusted Advisor

- inspects your AWS environment, and then makes recommendations when opportunities exist to save money, improve system availability and performance, or help close security gaps.
- Can identify underutilized EC2 instances
- Analyze account and give recommendations on 5 categories
    - Cost optimization
    - Performance
    - Security
    - Fault tolerance
    - Service limits
- 7 core checks of the trusted advisor (Basic & Developer support plan)
    - S3 Bucket permission
    - Security groups
    - IAM use
    - MFA on root account
    - EBS public snapshots
    - RDS public snapshots
    - Service limits
- Checks security groups for unrestricted access
- Full check (Business & Enterprise support plan)
    - Full checks on 5 categories
    - Set cloudwatch alarms when reaching limits, Access AWS Support API

Support plans

- Basic is free
- 5 plans
    - Basic
    - Developer
    - Business
    - Enterprise-on-ramp
    - Enterprise (Business-critical system down < 15 minutes)
- Full Trusted Advisor check starts at the Business plan
- The developer plan can open unlimited support cases but has only 1 primary contact while from the business plan, there is no limit on the number of contacts.
- Email support starts from the Developer plan (Business hour)
- 247 phone, email, and chat start from the business plan
- Third-Party Software Support only available from the business plan
- Only enterprise plan has dedicated TAM and training
- API access from the business plan
- The Concierge Support Team is the primary point of contact for billing or account inquiries. Available on enterprise-level only
- Severity levels:
    - General guidance
    - System impaired
    - Product system impaired
    - Product system down

<table>
<tr><th>Basic</th><th>Developer</th><th>Business</th><th>Enterprise</th></tr>
<tr><td>Email Support only for Billing and Account</td><td colspan="3">Tech support via Email ~24 hrs until reply</td></tr>
<tr><td></td><td>NO 3rd Party Support</td><td colspan="2">Tech support via Chat, Phone anytime 24 x 7</td></tr>
<tr><td></td><td colspan="3">General Guidance <24 hrs</td></tr>
<tr><td></td><td colspan="3">System Impaired < 12hrs</td></tr>
<tr><td></td><td></td><td colspan="2">Production System Impaired 😟< 4hrs</td></tr>
<tr><td></td><td></td><td colspan="2">Production System Down 😨 < 1hr</td></tr>
<tr><td></td><td></td><td></td><td>Business Critical System Down 😨 < 15mins</td></tr>
<tr><td></td><td></td><td></td><td>Personal Concierge 😎</td></tr>
<tr><td></td><td></td><td></td><td>TAM 🤓</td></tr>
<tr><td colspan="2">7 Trusted Advisor Checks</td><td colspan="2">All Trusted Advisor Checks ✅</td></tr>
</table>

Disaster recovery plans

- RTO: Recovery Time Objective: Time expected to recovery from when the disaster happens
- RPO: Recovery Point Objective: maximum time since the last data recovery point (how much data is allowed to lose)
- Both RTO and RPO are defined by the organization

Plans:

	Backup and restore	Pilot light	Warm standby	Multi-region active-active
RTO	hours	minutes	seconds	~0
RPO	<24 hours	hours	Minutes	~0

Plan details

- Backup and restore
    - Cheapest
    - Using point in time to backup into the DR region
- Pilot light
    - DR system off, need to turn on
    - Data replication and backup always on (DB, object storage such as S3)
    - Application servers are loaded with correct configs but turned off
- Warm standby
    - Backup servers running at scaled-down level
    - Business-critical systems are fully duplicated and always on (scaled down)
    - When disaster happens, the DR system is scaled up to full scale → hot standbay
- Multi-region active-active
    - Application running concurrently in multiple regions
    - Data sync across regions
    - User Route53 or Global accelerator to route traffic to where workload is healthy

WhitePapers Well-Architected

6 pillars of a well-architected framework

- Operational Excellence
    - Making frequent, reversible and continuous changes
    - Infrastructure as code
    - Refine operations procedures frequently
    - Anticipate failure

    - Learn from all operational failures
    - Related services:
        - CloudFormation
        - AWS Config
        - CloudTrail
        - CloudWatch
        - X-Ray
        - Code*
- Security
    - Implement a strong identity foundation
    - Enable traceability
    - Apply security at all layers
    - Automate security best practices
    - Protect data in transit and at rest
    - Keep people away from data
    - Prepare for security events
- Reliability
    - Resilient to failure
    - Automatically recover from failure
    - Test the recovery procedures
    - Scale horizontally
    - Stop guessing capacity
    - Manage change in automation
    - Including services:
        - VPC
        - Trusted Advisor
        - Service quotas
        - Auto scaling
        - CloudWatch
        - CloudTrail
    - Quickly recover from failure
    - Better to replace than fix
- Performance Efficiency
    - Make advanced technologies accessible for everyone
    - Ensure best performance while cost-optimized
    - Select the correct size of resources
    - Monitor resources to maintain performance
    - Go global in minutes
    - Use serverless architecture
- Cost Optimization
    - Avoid unnecessary costs
    - Using cost explorer, cost and usage report to optimize cost
    - Use serverless
    - Implement cloud financial management
    - Measure overall efficiency
    - Analyze expenditure
- Sustainability

    - Understand your overall impact on environment/society
    - Establish sustainability goals
    - Maximize utilization (Sharing resources between accounts in organizations)
    - Adopt new/efficient technologies
    - Use managed services

Best practices

- Scalable
- Disposable resources
- Automation
- Loose coupling
- Services not servers

Amazon S3

Pricing

You are charged by

- Storage
- Request & Data retrieval
- Data transfer IN and out
    - IN from internet is free
    - Out to internet is free for the first 100GB per month
    - Data transfer between S3 buckets or other services in the same region
    - Data transferred out to cloudfront
- Management & analytics
- Replication
- S3 Object Lambda

S3 Transfer Acceleration

- fast, easy, and secure transfers of files over long distances between your client and an S3 bucket.
- uses the globally distributed edge locations in Amazon CloudFront.
- As the data arrives at an edge location, data is routed to Amazon S3 over an optimized network path.

Buckets

- bound to a region

- name must be unique

Object

- files
- Object's key is its full path
- Max size is 5TB
- Max upload size is 5GB

Bucket policy

- User-based: Create
- Resource-based: Object ACL & Bucket ACL
- Encryption

S3 Versioning

- similar to Git

Access Logging

- Disabled by default
- Must select bucket to log
- Log access to a bucket

S3 replication

- Disabled by default
- Can be across regions or in the same region
- Must enable versioning
- Copy is async
- Buckets can be from different accounts

Storage class

- Standard: 99.99% availability
- Amazon S3 Intelligent-Tiering (Automatically move objects between classes to save cost)
- Standard-IA: Access infrequently. Access instantly, cross AZ
- One Zone-IA: 99.5% availability. Access instantly, single AZ (Cheaper than standard IA)
- Glacier Instant retrieval: Milliseconds retrieval
- Glacier Flexible retrieval:
    - Ideal for backup and disaster recovery use cases when large sets of data occasionally need to be retrieved in minutes
    - Configurable retrieval times, from minutes to hours, with free bulk retrievals
- Glacier Deep Archive:
    - Lowest cost storage class designed for long-term retention of data that will be retained for 7-10 years
    - Retrieval time within 12 hours

S3 Glacier

- Automatic encryption enabled

Glacier vault lock

- Lock for future edit

S3 Object Lock

- Write once read many
- Block a specific version deletion

S3 Encryption

- No encryption
- Server-side encryption
- Client size encryption

AWS Snow

- Physical devices
- AWS sends actual devices so the client can copy big data and send it back to AWS to upload to WAS
- Snowball edge: TB to PB
    - Supports EC2 natively
- Snowball edge storage optimized: 80TB
- Snowball Edge compute Optimized: 42TB
- Snowcone: smaller device, 8TB. Can send to aws offline or over the internet using datasync
- Snowmobile: transfer EiB data using a truck

AWS edge computing

- Snowball edge
- Machines to do computing when a connection to AWS is not possible

AWS OpsHub

- Snow devices management dashboard

Storage Gateway

- Connect on-premise to the cloud storage
- Can use direct gateway or over the internet
- Encryption enabled
- Pay for storage and data transfer out. Data transfer in is free
- Types:
    - Tape Gateway (Used for backup, compatible with most backup software)
    - Amazon S3 File Gateway
    - Amazon FSx File Gateway

    - Volume Gateway (backup and used on the cloud as EBS)

AWS Quick starts

- deploy popular technologies on AWS according to AWS best practices.
- You can reduce hundreds of manual procedures to just a few steps so that you can build and start using your environment within minutes.

Other services

Elastic Transcoder

- Amazon Elastic Transcoder is media transcoding service
- convert (or “transcode”) media files from their source format into versions that will playback on devices like smartphones, tablets and PCs.

Marketplace

- The AWS Marketplace enables qualified partners to market and sell their software to AWS Customers
- Two ways to sell on marketplace:
    - Amazon Machine Image
    - SaaS

Penetration testing

- Users can carry out penetration testing without prior permission on these 8 permitted services:
    - Amazon EC2 instances, NAT Gateways, and Elastic Load Balancers
    - Amazon RDS
    - Amazon CloudFront
    - Amazon Aurora
    - Amazon API Gateways
    - AWS Fargate
    - AWS Lambda and Lambda Edge functions
    - Amazon Lightsail resources
    - Amazon Elastic Beanstalk environments
- These actions are not allowed
    - DNS Zone walking
    - DDoS, DoS
    - Port flooding
    - protocol Flooding
    - Request flooding

Workspaces

- VDI
- Desktop in the cloud

- Virtual Desktop

AppStream

- Desktop app streaming service
- App in browser (Like running Eclipse)

Sumerian

- Run VR, AR, 3D apps
- Create, view 3D models in browser

IoT Core

- Connect IoT Devices to AWS cloud
- Serverless

Elastic Transcoder

- Convert media files in S3 to format playable on tablets/smartphones/PC

Device farms

- Test mobile/web apps across devices

AWS Backup

- Manage and automate backup
- Support Point-in-time backup
- Supported services
    - Amazon Elastic Compute Cloud (Amazon EC2) instances
    - Windows Volume Shadow Copy Service (VSS) supported applications (including Windows Server, Microsoft SQL Server, and Microsoft Exchange Server) on Amazon EC2
    - Amazon Elastic Block Store (Amazon EBS) volumes
    - Amazon Simple Storage Service (Amazon S3) buckets
    - Amazon Relational Database Service (Amazon RDS) databases (including Amazon Aurora clusters)
    - Amazon DynamoDB tables
    - Amazon Neptune databases
    - Amazon DocumentDB (with MongoDB compatibility) databases
    - Amazon Elastic File System (Amazon EFS) file systems
    - Amazon FSx
    - AWS Storage Gateway volumes
    - VMware workloads on premises, on Amazon Outposts, and in VMware Cloud on AWS TM

Disaster Recovery Strategies

- Backup and restore
- Pilot light
- Warm standby

- Multi-site/hot-site

DataSync

- move large data from on-premise to AWS
- Data can transfer through the internet or direct-connect
- Data is encrypted if using the internet
- Can transfer data from
    - S3 on OutPost
    - Snowcone
    - DataSync agent
- Can transfer data to aws cloud
    - S3
    - EFS
    - FSx

Fault Injection Simulator

- Simulate disaster -> Chaos engineering
- Helps discovering hidden bugs