4 Core Components for IAAS

**Storage**

ESB virtual hard drives

**Compute**

EC2 instances

**Networking**

VPC Private Cloud Network

**Databases**

RDS SQL databases

**Regions and AZ**

25 launched regions

81 availability zones

108 direct connection locations

275+ points of presence

17 wavelength zones

Regional services scopes into a selected region

Global services operate across multiple regions – Amazon 3, CloudFront

Availability zone made up of one or more datacenters

A region generally has 3 AZs

A subnet is associated with an availability zone and EC2 (virtual machines) are launched into those subnets

**Fault Tolerance**

Fault level – AWS region

Fault domain – AZ

Each region is designed to be completely isolated from other AZs

They are connected through low latency links

Each AZ is an independent failure zone

Multiple AZ for high availability

**AWS Global Infrastructure**

AWS Global interconnections between the Global Infrastructure

Edge locations on and off ramps to the AWS Global Network

Edge locations – AWS Global Accelerator, S3 Transfer Acceleration

Off ramp – Amazon Cloudfront – storage and commute near end user

VPC Endpoints

**POP – Point of Presence**

Location between AWS Region and end user - these will be either

Edge location

Regional Edge Caches

Cloudfront – Content Delivery Network – you can point your website to CloudFront

AWS Global Accelerator – finds optimal path from end user to web server.

**AWS Direct Connect**

Private connection between datacenter, office and AWS

Has two very fast network connections – lower and high bandwidth

Direct Connect Locations – trusted partnered datacenters from on-premise to AWS

**AWS Local Zones**

Datacenters close to densely populated areas

You need to opt in for Local Zones

**Wavelength zones**

Edge-computing on 5g networks

Ultra- low latency to be close as possible

**Data Residency**

Physical location of where the org or cloud resources reside

Need to know compliance boundaries or data sovereignty

**AWS Outposts**

Physical rack of servers that you can put in your datacenter, Fully managed service

**AWS Config**

Policy as a Code service – check the resources config

IAM Policies - To deny access for regions

Service Control Policy – SCP – permissions applied organization wide

**AWS Ground Station**

Full managed service that lets you control satellite communications

**Cloud Services Must Have**

**High Availability**

Ensuring there is no single point of failure

Running workloads over multiple Availability Zones

Elastic Load Balancer – allows you to evenly distribute traffic to multiple servers in one or more datacenter

**High Scalability**

Ability to increase your capacity

Vertical – upgrade to a bigger server

Horizontal scaling – scaling out to add more servers of the same size

**Elasticity**

Ability to automatically increase or decrease your capacity based on current demand of traffic, memory

Auto Scaling Groups (ASG)

**Highly Fault Tolerant**

Fail-overs is a plan to shift traffic to a redundant system in case the primary system fails

RDS Multi-AZ – is when you run a duplicate standby database

**High Durability**

Ability to recover from a disaster and to prevent the loss of data

CloudEndure Disaster Recovery

AWS API – Postman - This is a workspace of AWS collections, allowing developers to work with many of the AWS APIs.

**Amazon Resource Name - ARN**

The ARN breaks down into - ARN:partition:service:region:account-id:resource-id

Partition Examples-

AWS Global regions

Aws CN – China Regions

Aws-us-gov AWS GovCloud

Service Examples –

EC2, S3, IAM

**Software Development Kit**

Collection of software development tools in one package

AWS SDK – offered in various programming languages

Cloud 9 – debugs code, sets up an environment on an EC2 instance

Infrastructure as Code

AWS CloudFormation CFN - scripting

AWS Cloud Development Kit – programming

Aws Toolkit – open source plug in for VScode to create, debug and deploy AWS resources

AWS Explorer – explore resources

AWS CDK Explorer

Amazon Elastic Container Service

Serverless Applications

Access key – a user must be granted access to use access keys, you can have 2 active keys at the same time

**Shared Responsibility Model –**

AWS -

Hardware/Global Infrastructure – Regions, Availability Zones, Edge Locations, Physical Security

Software – Compute, Storage, Database, Networking

Customer –

Configuration of Managed Services – Platforms, Applications, Identity and Access Management (IAM),

Configuration of Virtual Infrastructure – Operating System, Network, Firewall

Security Configuration of Data – Client-Side Data Encryption, Server-Side Encryption, Networking Traffic Protection, Customer data,

Customer – Data and Configuration, IN the cloud

Amazon – Hardware, Operation of Managed Services, Global Infrastructure, Of the cloud

**Compute –**

Iaas – Bare Metal, Virtual Machine (EC2), Containers (ECS)

Paas –Managed Platform – AWS Elastic Beanstalk

Saas – Amazon WorkDocs

Faas – Functions – AWS Lambda

**Cost and Capacity Management Computing Services**

EC2 Spot Instances, Reserved Instances and Savings Plan

AWS Batch – plans, schedules and executes your batch computing workloads

AWS Compute Optimizer – suggests how to reduce costs and improve performance by using machine learning

Ec2 Autoscaling Groups – automatically adds or remove EC2 servers to meet the current demand of traffic

Elastic Load Balance – distributes traffic to multiple instance

AWS Elastic Beanstalk - deploying web applications without developers

**Types of Storage Services**

Elastic Block Store – Block – linking virtual hard drive attached to a VM

Elastic File Storage – File – file-share where multiple users or VMs need to access the same drive

Amazon Simple Storage Service – Object – upload files

**S3 Rundown**

Provides unlimited storage

S3 Objects

S3 buckets – buckets hold objects, can have folders, must have unique name

S3 Storage Classes – most expensive to cheapest

S3 Standard Default – Across at least 3 AZs

S3 Intelligent Tiering

S3 Standard-IA – Additional retrieval fee

S3 One-Zone-IA – only in one AZ

S3 Glacier – long term cold storage – minutes to hours

S3 Glacier Deep Archive – 12 hours for data retrieval

**AWS Snow Family**

Storage and computer devices to physically move data in or out of the cloud

Snowcone – two sizes, 8tb or 14tb

Snowball Edge – 80tb or 39.5tb

Snowmobile - 100 Pb

**Storage Services**

Simple Storage Service is a serverless object storage service

S3 Glacier – cold storage service

EBS is a persistent block storage service

EFS – cloud-native NFS file system service

Storage Gateway – hybrid cloud storage

File Gateway

Volume Gateway

Tape Gateway

AWS Back up – managed backup service

CloudEndure Disaster Recovery

Amazon FSx – file system

**NoSQL Database Service**

DynamoDB – flagship database service

DocumentDB – noSQL document database

Amazon Keyspaces – fully managed Apache Cassandra database

**Relational Database Services**

Relational Database Service – supports SQL engines – MySQL, MariaDB, PSQL, Aurora, Microsoft SQL Server

Aurora – fully managed database of either MySQL and PSQL

Aurora Serverless – serverless on demand version of Aurora

RDS on VMware – allows you to deploy RDS supported engines

**Other Database Services**

Redshift – petabyte-size data-warehouse

ElastiCache – managed database of the in-memory and caching open-source database

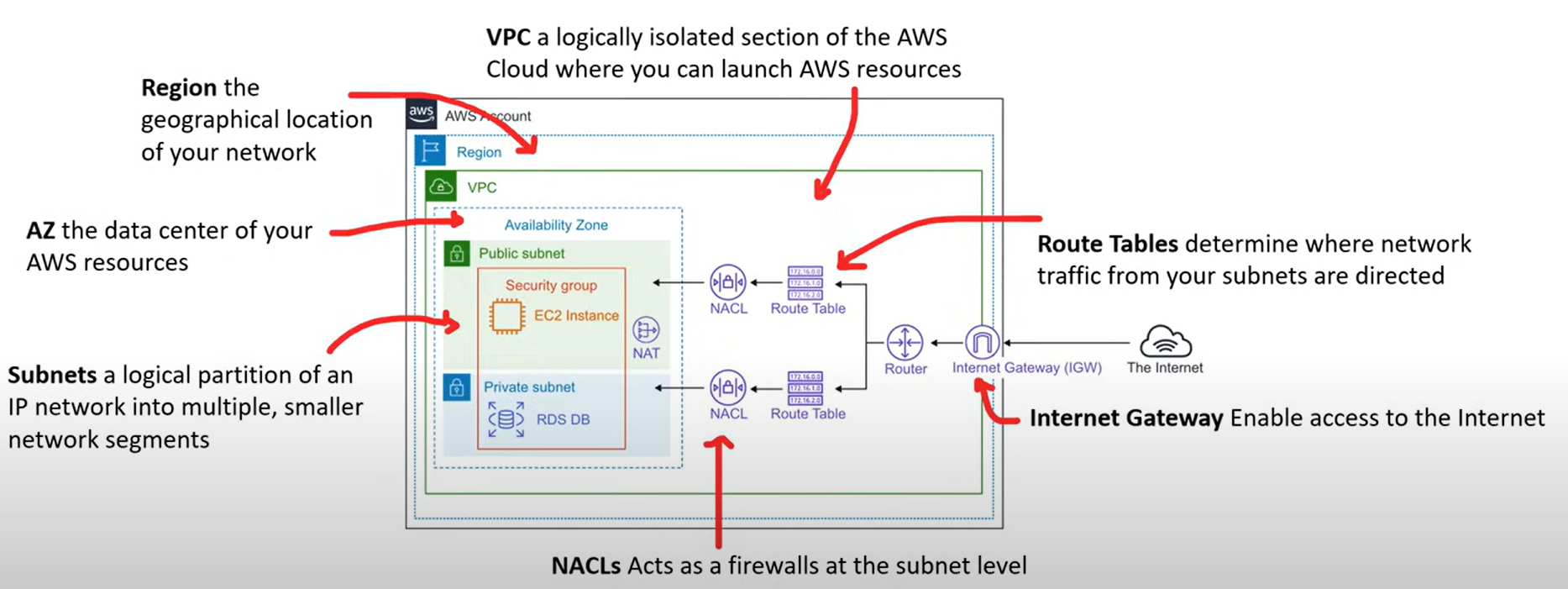
Neptune – managed graph database

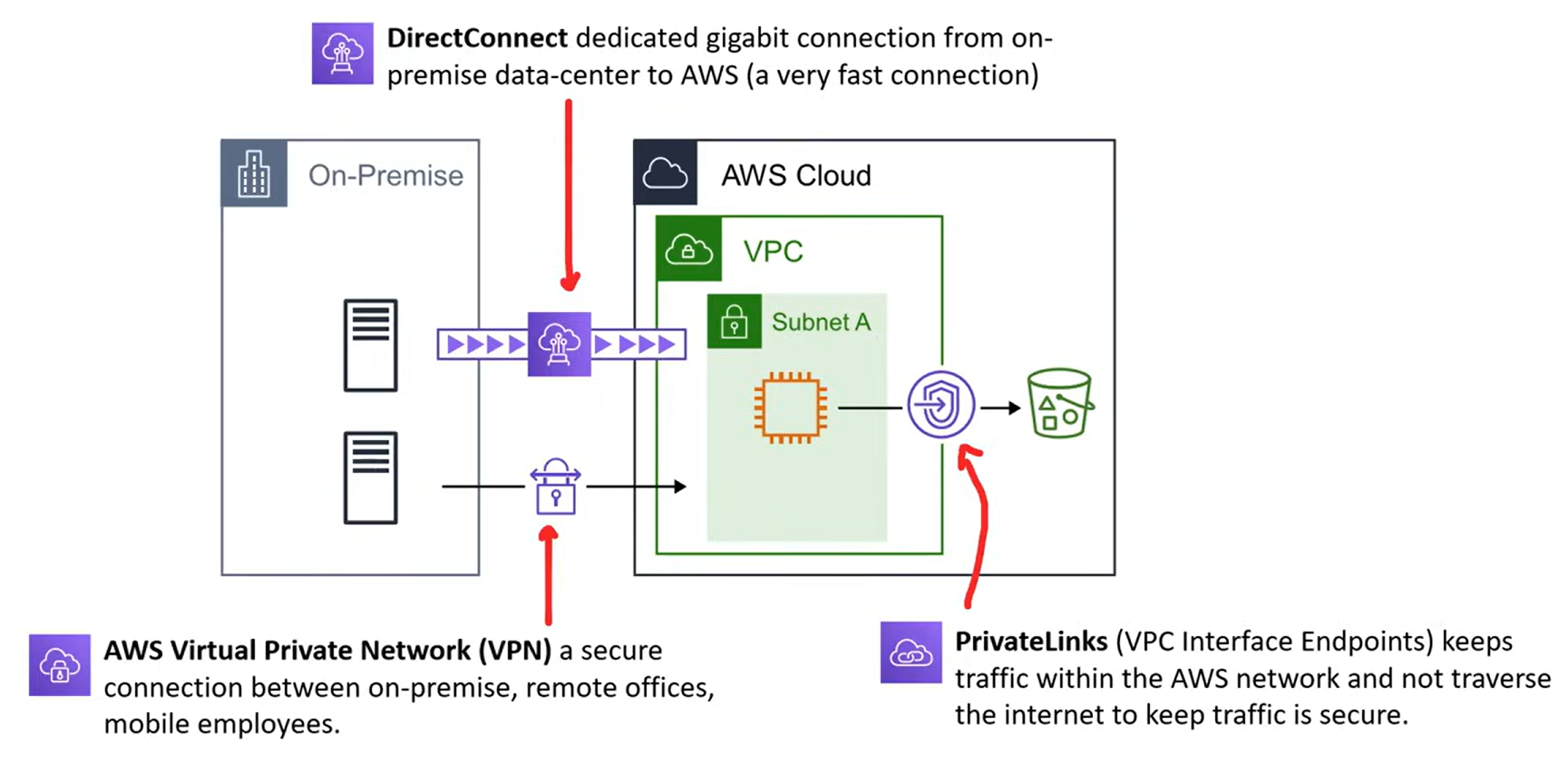
Amazon Timestreams – fully managed time series

Amazon Quantum Ledger Database – fully managed ledger database

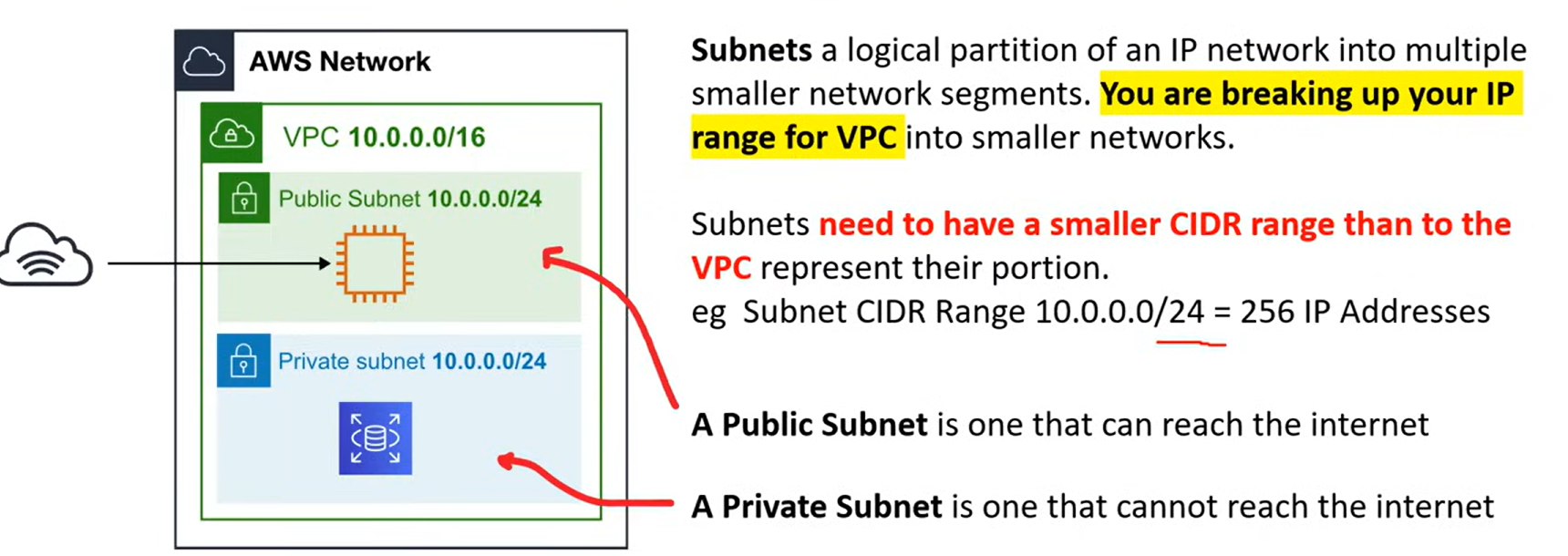
Database Migration Service – migrate from on premise, from SQL no to NoSQL database

**Cloud-Native Networking Services**

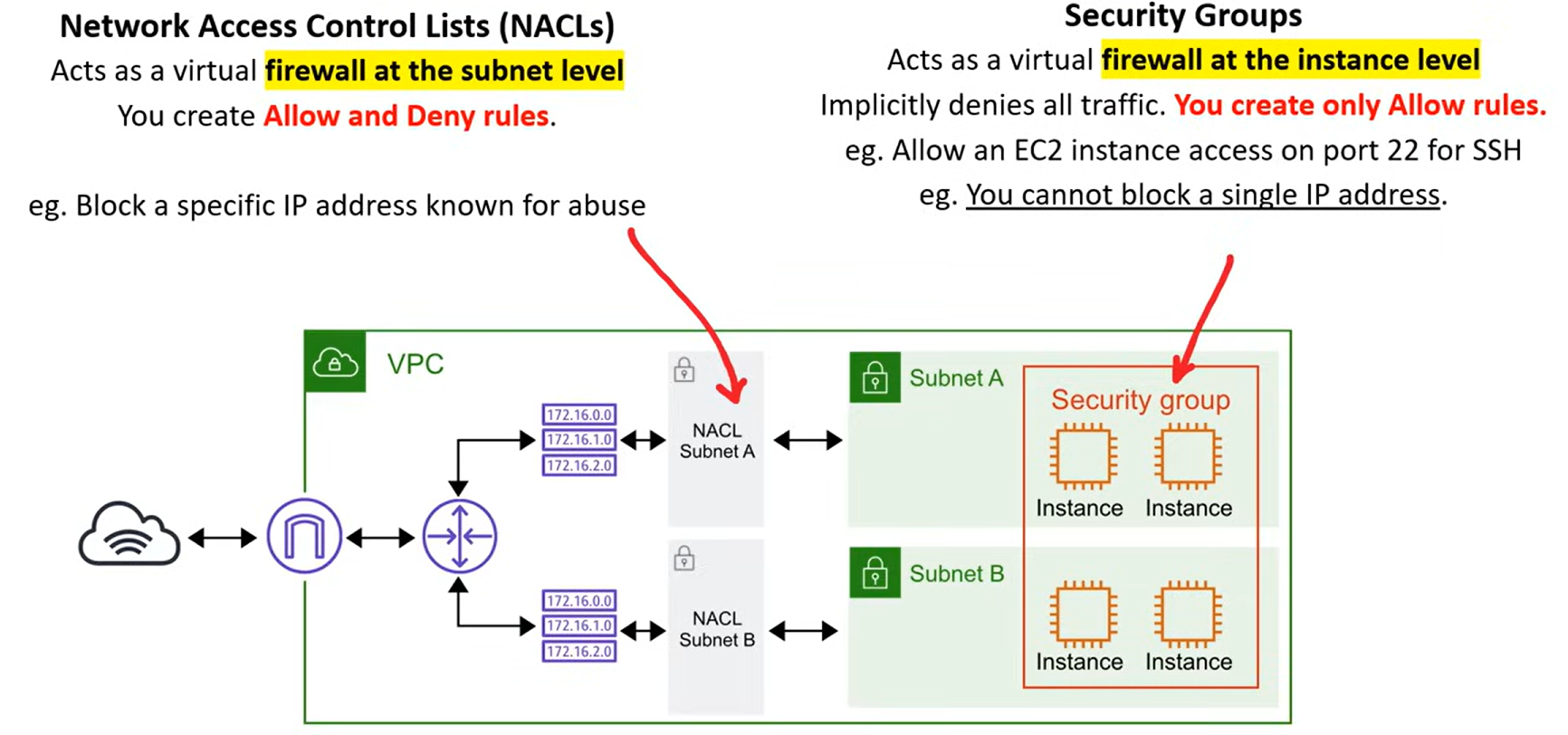


**Enterprise/Hybrid Networking**

**Virtual Private Cloud** – logically isolated section of the AWS network where you choose a range of IPs using CIDR range



**Security Groups vs NACLs**



**EC2 Instances**

These are the virtual machines AWS uses.

Instance families are different combos of CPU, Memory, Storage and Networking capacity. Instance families are:

General Purpose – balance of compute, memory and networking resources

Compute Optimized- ideal for compute bound applications

Memory Optimized – fast performance for workloads that process large data sets in memory

Accelerated Optimized – hardware accelerators

Storage Optimized – high sequential read and write access to large data sets on local storage

EC2 has three levels of tenancy.

Dedicated host - Dedicated Host are single-tenant Ec2 instance are designed to let you bring you own licence based on machine characteristics.

Dedicated instance

Default

**EC2 Pricing Models**

**On-Demand** – low cost and flexible, pay per hour, short term workload

This is the default pricing. No up-front payment and no long-term commitment.

**Spot** – most savings, flexible start and end times, must handle interruptions,

**Reserved** – steady state, committing over a 1- or 3-year term, can resell unused reversed instances, best long-term savings. Reduced pricing is based on Term x Class Offering x RI Attributes X Payment Option. The reserved instances do not renew automatically, they will move to on-demand. It is cheaper to pay more upfront.

RI Attributes are instance type, region, tenancy and platform. When you purchase an RI you need to pick a scope – either Regional or Zonal RI. There is a limit to the number of reserved instances you can purchase per month. 20 Regional reserved instances per region and 20 zonal reserved instances per AZ. RI is also split into Standard and Convertible RI. Standard can be modified but can’t be exchanged and Convertible has the opposite.

RI Marketplace allows you to sell your unused Standard RI to recoup your RI spend. You must have a US bank account and have at least one month remaining on the reserved instances.

**Dedicated** – dedicated servers can be on-demand or reserved or spot. Multi-tenant (everyone living in an apartment) or single tenant (everyone having their own house) are offered to meet regulatory requirements. If this isn’t selected it will be shared.

**Capacity Reservations**

Allows you to request to reserve an EC2 type in a specific region or AZ when it isn’t currently available

**AWS Savings Plans**

This can be 1 or 3 years plans.

Compute Savings Plans – these are automatically applied to the EC2 instances. Most flexible.

EC2 Instance Savings Plans – this provide the lowest price and automatically reduces your cost on the selected instance family.

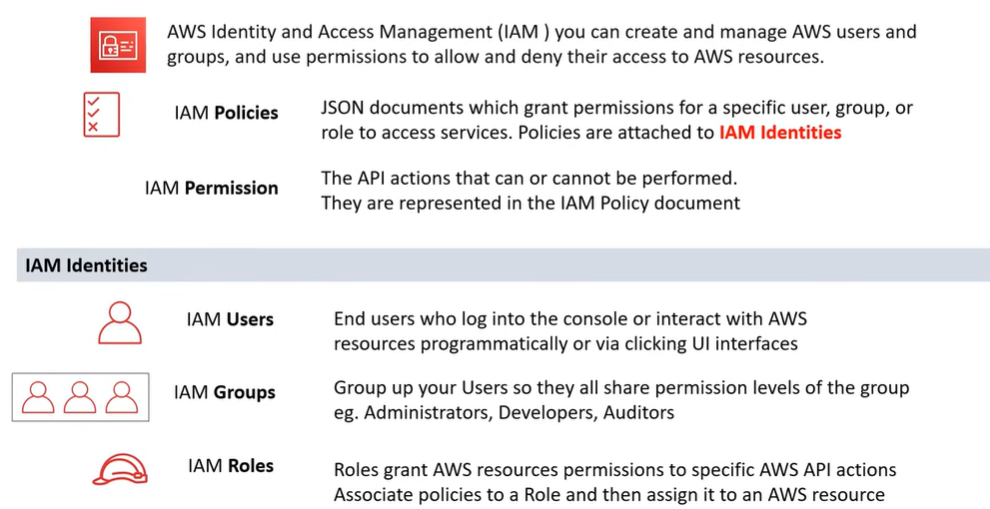
Sagemaker Savings Plans – for ML (machine learning).

**Zero Trust Model**

Operates on the principle of trust no one, verify everything.

Identity Security Controls you can implement on AWS to meet this model.

AWS IAM – IAM Policies, permission boundaries, Service Control Policies, IAM Policy Conditions,



**AWS Root Account Tasks**

Change your account settings

Restore IAM user permissions

Activate IAM access to the Billing and Cost Management Console

Sign up for GovCloud

Edit or delete a S3 bucket policy

View certain tax invoices

Close your AWS account

Change or Cancel AWS Support Plan

**Application Integration Services**

Amazon API Gateway – Creating secure APIs in the cloud environment at any scale.

AWS Step Functions – this is a state machine; an abstract model which decides how one state moves to another.

Event Bridge – serverless event bus service by streaming real-time data. Three types of event busses- default event bus, custom event bus and a SaaS event Bus. 100 rules per bus and 5 targets per rule.

Simple Queueing Service – queue up transaction emails to be sent. It is not real-time.

Amazon Kinesis – AWS fully managed solution for streaming data in the cloud.

Simple Notification Service (SNS) - Publish/Subscribe – this is sending the message to an event bus and the receivers of the messages subscribe to these groups. A subscriber could be SQS.

Amazon MQ – managed message broker services

Managed Kafka Service -fully managed Apache Kafka service

AppSync – fully managed GraphQL service

**Containers**

Primary Services – Elastic Container Service, AWS Fargate, Elastic Kubernetes Services, AWS Lambda

Provisioning and Deployment – Elastic Beanstalk, App Runner, AWS Copilot CLI

Supporting Services – Elastic Container Registry, X-Ray, Step Functions

AWS Organizations includes Root Account User, Organization Units, Service Control Policies. Must be switched on and cannot be switched off afterwards.

AWS Control Tower – helps Enterprises setup a secure AWS multi account

AWS Config – Compliance-as-Code framework on a per region basis.

**Business Centric Services**

Amazon Connect – virtual call center service

Workspaces – virtual remote desktop service

WorkDocs – like Sharepoint

Chime – AWS Skype or Zoom

WorkMail – managed business email

Pinpoint – marketing campaign management service

Simple Email Service – transaction email service

QuickSight – Business Intelligence

**Provisioning Services**

Elastic Beanstalk – Platform as a Service

It is powered by a CloudFormation template setups.

AWS OpsWorks – configuration management service – Chef and Puppet

CloudFormation – IAC – infrastructure modelling and provisioning service

AWS Quick Starts – automated reference deployments built by AWS

AWS Marketplace – digital catalogue of thousands software listings from independent software vendors

AWS Amplify – mobile and web application framework – for serverless services

AWS App Runner – fully managed service to deploy web apps and APIs

AWS CoPilot – CLI

AWS Codestar – easily launch types of stacks

AWS Cloud Development CDK – allows you to use favourite programming language to provision

**Serverless Services**

DynamoDB – NoSQL – can scale to billions of records

Simple Storage Service S3 – serverless object storage service

ECS Fargate – serverless orchestration container service

AWS Lambda – serverless functions service, run code without provisioning

Step Functions – state machine service

Aurora Serverless – on-demand version of Aurora

**Logging Services**

Cloudtrail – logs all API calls between AWS services by default, last 90 days via Event History, any longer must be a Trail and use Athena to analyze the Trail. Able to see who is doing what in the different instances.

Cloudwatch – collection of multiple services – logs, metrics, events, alarms and dashboard, Cloudwatch Alarm monitors a metric based on a defined threshold and then it triggers an action. Log streams is a sequence of events from an application or instance being monitored. Cloudwatch Metric represents a time-ordered set of data points.

AWS X-Ray – distributed tracing system, testing between apps

**Machine Learning and AI Services**

Amazon SageMaker – build, train and deploy machine learning models at scale

Amazon Augmented AI – human-intervention review service

Amazon Sagemaker Ground Truth – data-labeling service

Amazon CodeGuru – machine-learning code analysis service

Amazon Lex – conversion interface service

Amazon Personalize – real-time recommendations service

Amazon Polly – text to speech

Amazon Rekognition – image and video recognition service

Amazon Transcribe – speech to text service

Amazon Textract – extract text from scanned documents

Amazon Translate – language translator

Amazon Comprehend – Natural Language Processor service

Amazon Forecast – time-service forecasting service

AWS Deep Learning AMIS – Amazon Ec2 instances that are pre-installed with popular deep learning frameworks

AWS Deep Learning Containers – Docker images instances pre-install with deep learning frameworks

AWS DeepComposer – machine learning enabled musical keyboard

AWS DeepLens – video camera uses deep learning

AWS DeepRacer – toy race care can be powered with machine learning

Amazon Elastic Inference – allows you to attach low cost GPU powered acceleration to EC2 instances

Amazon Fraud Detector – fully managed fraud detection service

Amazon Kendra – enterprise machine learning search engine service

**Big Data**

Used to describe massive volumes of structured/unstructured data

Amazon Athena – serverless interactive query service – query CSV or JSON files

Amazon CloudSearch – adding search to your website

Amazon Elasticsearch Service – managed Elasticsearch cluster. More robust than CloudSearch

Amazon Elastic MapReduce – data processing and analysis

Kinesis Data Streams – real time streaming data service, create producers which send data to a stream

Kinesis Firehose – serverless and simpler version of Data streams

Amazon Kinesis Data Analytics – allows you to run queries against data

Amazon Kinesis Video Streams – allows you to analyze real time streaming video

Managed Kafka Service – fully managed Apache Kafka service

Redshift – petabyte data warehouse – quickly generate analytics or reports from a large amount of data

Amazon QuickSight – business intelligence dashboard

AWS Data Pipeline – automates movement of data

AWS Glue – Extract, Transform and Load

AWS Lake Formation – centralized, curated and secured repository

**AWS Well Architected Framework**

Operational Excellence

Security

Reliability

Performance Efficiency

Cost Optimization

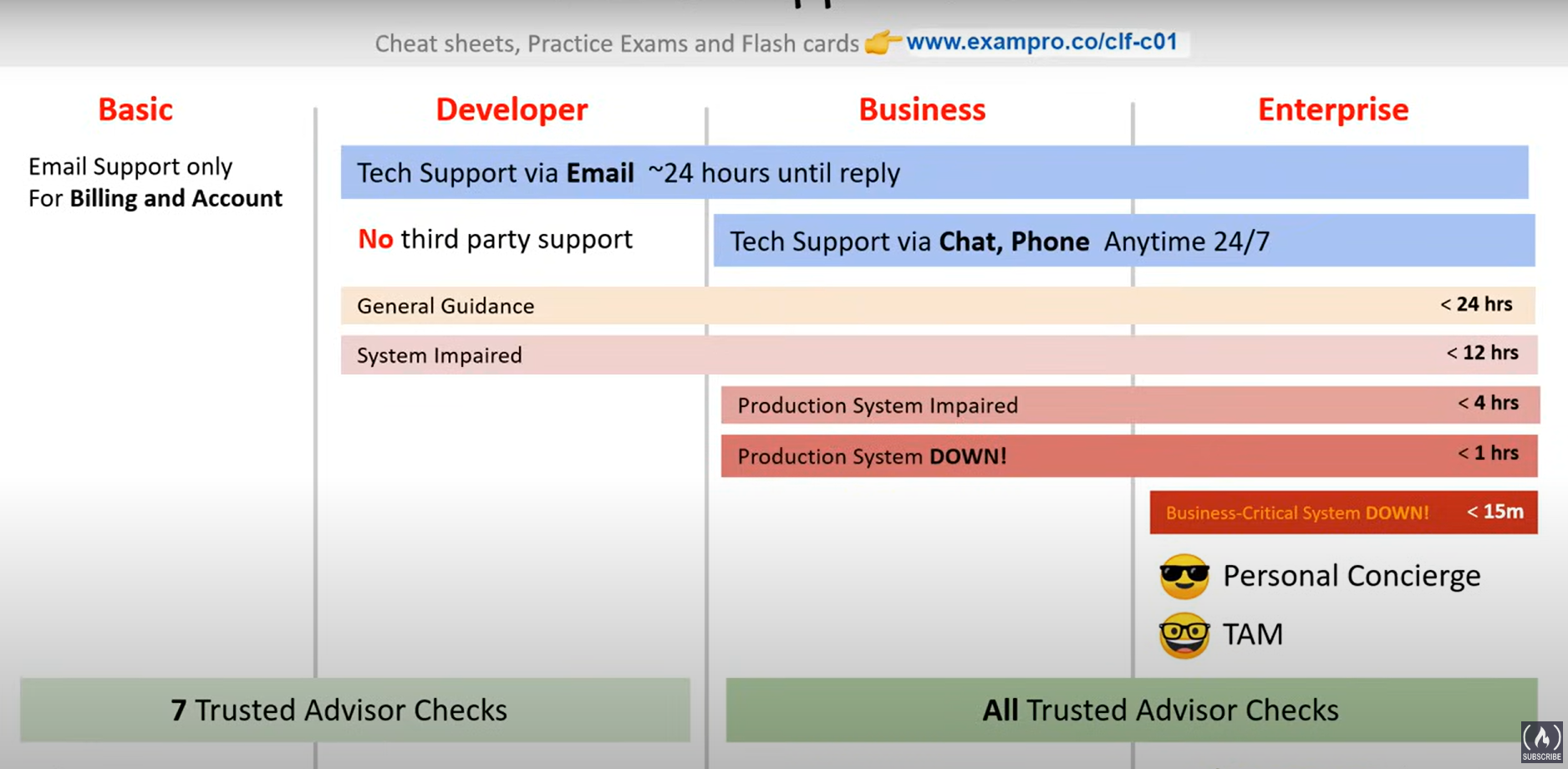
**CAPEX – Capital Expenditure vs OPEX – Operational Expenditure**

Capex – storage, network costs, technical personal

OPEX – Shifted cost to service provider, only need to be concerned with non-physical costs

AWS pricing calculator – within web browser to estimate cost of AWS services

**AWS Support Plans**

****

**Technical Account Manager**

Technical account manager – build solutions, provide technical guidance, check environments, work closely with AWS customers – at an Enterprise level account

**AWS Trusted Advisor**

Recommendation tool to provide actional recommendations

Only 7 checks for Basic and Developer

All checks for Business and Enterprise

**AWS Abuse** – to report improper use of AWS services

**AWS Budgets –** Set up alerts if you exceed or are approaching your defined budget, support EC2, RDS, Redshift, Elasticache – tracked at monthly, quarterly or yearly levels

**AWS Cost and Usage Reports** – Better analyze and understand your AWS Costs

**AWS Inspector** – runs a security benchmark against specific EC2 instances

**AWS Artifact** – self server portal for AWS Compliance reports

**AWS Shield** – managed DDoS protection service, Shield standard is free, Advanced has a cost per year,

**Amazon Guard Duty –** threat detection service, uses machine learning to analyze AWS logs

**Amazon Macie** – full managed service that continuously monitors S3 data access activity for anomalies

**AWS VPN** – secure and private tunnel from your network to the AWS global network

**AWS WAF –** web application firewall, protecting the apps, attached to Cloudfront or Application Load Balancer

**AWS Health** - AWS Health provides personalized information about events that can affect your AWS infrastructure, guides you through scheduled changes, and accelerates the troubleshooting of issues that affect your AWS resources and accounts.

**Route 53** – DNS, can perform health checks

**Hardware Security Module (HSM)**

HSMs – multi tenant- AWS KMS

HSMS - single tenant – AWS CloudHSM

**AWS Key Management Service** – managed service to create and control encryption keys, uses Envelope Encryption.

**CloudHSM** – automates hardware provisioning, software patching and backups,

**AWS Config** – governance tool for Compliance as a Code

**AWS AppConfig** – automate the process of deploying application configuration variable to your web applications

**Amazon Inspector** – inspects a single EC2 instance – generates a report

**Trusted Advisor** – doesn’t generate out a PDF report, does give a holistic view of recommendations