**AWS – Solution Architect**

* Global Infrastructure
  + Region is a geographical area, contains 2 or more AZs
  + Availability Zone is data center
  + Edge location is CDN end point for Cloud Front, data cached in edge location
* Services
  + Compute
    - Elastic Compute Cloud (EC2) – Virtual Instance
      * EBS virtual disc installed in EC2
    - EC2 Container Service
    - Elastic Beanstalk – **AWS** Elastic **Beanstalk** is an orchestration service offered from **Amazon** Web Services for deploying infrastructure which orchestrates various **AWS** services, including **EC2**, S3, Simple Notification Service (SNS), CloudWatch, autoscaling, and Elastic Load Balancers.
    - Lambda – Trigger Events
    - Lightsail
    - Batch Computing in cloud
  + Storage
    - S3 (Simple Storage Service) – Virtual disc in cloud (store objects like documents, files, pictures, movies, text file etc.), **Object based storage**
    - Glacier – Archive file from S3 (retrieve time is 3 – 4 hours)
    - EFS (Elastic file Service) –**File/object based storage,** install DB, game
    - Storage Gateway - hybrid storage service that enables your on-premises applications to seamlessly use storage in the AWS Cloud
    - Snowball - petabyte-scale data transfer solution (large amount of data)
  + Database
    - RDS
    - Dynamo DB (No-SQL)
    - Redshift – AWS warehouse system
    - Elastic cache
  + Migration
    - DMS (Database Migration System)
    - Application Discovery Service
    - AWS Migration Service
    - Snowball
  + Network and Content Delivery
    - VPC (Virtual Private Cloud) – Multiple VPC in region, connect between VPC
    - Route 53 – DNS service, register domain name
    - Cloud Front – Content Delivery Network
    - Direct Connect – Connecting physical data center in local using wire connection
    - API Gateway
  + Developer Tool
    - Code commit – GIT hub
    - Code Build – Compile code
    - Code Deploy – Deploy code to EC2
    - Code Pipeline
    - Code Star
    - X-Ray
  + Management Tool
    - Cloud Watch – Monitor performance of EC2
    - Cloud Trial – Auditing AWS resources
    - Cloud Formation – Turning infrastructure into code (describes AWS environment)
    - OPS work – Automate deployment using chef
    - Config – Auditing environment like alert configuration etc.
    - Service Catalogs
    - Trust Advisor - An online resource to help you reduce cost, increase performance, and improve security by optimizing your AWS environment. It provides real time guidance to help you provision your resources following AWS best practices
    - System Manager
  + Media Services
    - Elastic Transcoder – Transcode video based on devices
    - Media Convert
    - Media Live
    - Media Store
    - Media Tailor
    - Media Package
  + Analytics
    - Athena
    - EMR (Elastic Map Reduce) – Process large amount of data (Big-data), uses Hadoop tech
    - Cloud Search
    - Elastic Search Services
    - Kinesis – Streaming and analyzing data (transaction, social media etc.)
    - Kinesis Video Streams
    - Data Pipeline
    - Quick Sight – Business analytic tool
    - Data Pipeline
    - Glue
  + Security & Identity
    - IAM (Identity Access Management)
    - Cognito
    - Guard Duty
    - Inspector – Agent install in VM to monitor EC2
    - Macie
    - Certificate Manager - SSL
    - Cloud HSM
    - Directory Service
    - Web Application Firewall (WAF) – Application level protection like CROSS, SQL Injection
    - Shield
    - Artifacts – Compliance Document in AWS console
  + Mobile Service
    - Mobile Hub
    - AWS App Sync
    - Device Farm
    - Mobile Analytics
    - Pinpoint
  + Application Services
    - Step Functions
    - Simple Work flow service (SWF)
    - SNS (Simple Notification Service)
    - SQS (Simple Queue Service) – Queue system
    - SES (Simple Email Service)
    - Amazon MQ
  + Customer Engagement
    - Connect
    - Simple Email Service
  + Business Productivity
    - Alexa for Business
    - Chime
    - WorkDocs
    - WorkMail
  + Desktop and App Streaming
    - Work Spaces - Desktop-as-a-Service
    - App Stream 2.0 - Streaming desktop application to user
  + Internet of Things (IOT)
  + Machine Learning
    - Alexa (Lex) – Amazon Eco
    - Polly – tool to convert text to media file
    - Machine Learning
    - Rekognition

----------------------------------------------------------------------------------------------------------------

**Identity Access Management (IAM) - Global**

* Users
* Group
* Role – Interact between AWS resources
* Policies Documents
* IAM is global, not apply to regions
* Programmatic Access: Access key ID and Access Secret key
* Console Access: User ID and Password

----------------------------------------------------------------------------------------------------------------

**AWS Object Storage and CDN - S3, Glacier and CloudFront**

* Simple Storage Service (S3)
  + S3 is **Object based storage** like videos, flat file, photo etc.
  + File size can be 0 to 5 TB, largest object that can be uploaded in a single PUT is 5 gigabytes. larger than 100 megabytes, customers should consider using the [Multipart Upload](http://docs.amazonwebservices.com/AmazonS3/latest/dev/UploadingObjects.html) capability
  + Data Consistency Model
    - Read after Write Consistency for PUTs new Object
    - Eventual Consistency Model (ATOMIC) for overwrite PUTs and DELETES
  + Storage Tiers/ Classes
    - Standard S3
    - Standard S3 – IA (infrequently Accessed) – Charged retrieval fee, Minimum Object size 128KB, Minimum Storage duration – 30 days
    - Reduced Redundancy Storage
    - Glacier – Data Archival, 3 – 5 hours to restore, Minimum Storage duration – 90 days
  + URL Format:
    - Path-Style:
      * N.Virginia: [http://s3.amazonaws.com/[BucketName]/[Filename](http://s3.amazonaws.com/%5bBucketName%5d/%5bFilename)]
      * Other Regions: [http://s3-[region].amazonaws.com/[BucketName]/[Filename](http://s3-[region].amazonaws.com/%5bBucketName%5d/%5bFilename)]
    - Virtual-host:
      * N.Virginia: [http://[*BucketName*].s3.amazonaws.com/[*Filename*](http://[BucketName].s3.amazonaws.com/%5bFilename)]
      * Other region: [http://[*BucketName*].s3-[region].amazonaws.com/[*Filename*](http://[BucketName].s3-[region].amazonaws.com/%5bFilename)]
    - Static Website: <http://[BucketName].s3-website-[region].amazonaws.com/>

S3 Properties & Permission setting

* + Properties
    - Bucket level
      * Versioning
      * Logging
      * Static Website hosting
      * Tags
      * Transfer Accelerator
      * Events
      * Requester pay
    - Object Level
      * Storage Class – Standard, Infrequently Access, Reduced Redundancy Storage
      * Encryption
      * Metadata – AES 256
      * Tags
  + Permissions
    - Bucket Level (ACL/Bucket Policy/CORS configuration)
      * List object
      * Write object
      * Read bucket permission
      * Write bucket permission
    - Object level
      * Read Object
      * Read bucket permission
      * Write bucket permission
  + Cross Region Replication
    - Versioning should be enabled in source & destination bucket
    - Multiple chain of Cross Region replication will not happen
  + Life Cycle Management
    - Can be used in conjunction with Version
    - Can be applied to current and previous version
    - S3 – IA: 30 days, Glacier: 90 days
    - Minimum size for S3-IA is 128kb
* Cloud Front CDN
  + Edge location – Location where content will be cached and is connected to region
  + Origin – This is the origin of all the files that CDN will distribute. This can be S3 bucket, EC2 Instance, Elastic Load Balancer or Route 53
  + Distribution – Collection of Edge location
    - Web Distribution – Web site
    - RTMP – Adobe flash media Server’s RTMP protocol
  + TTL
  + Restrict Viewer Access
    - GEO Restriction- Restriction Type: White list / Black list
  + Invalidation: Remove object from cache
* Security and Encryption
  + 2 type of Encryption
    - In Transit: SSL/TLS
    - At Rest:
      * Server Side: SSE-S3, SSE-KMS, SSE-C
      * Client Side
  + Security
    - Bucket policy
    - Access control list
    - Logging
* Storage Gateway: Gateway b/w on premise infrastructure and AWS
  + File Gateway (NFS) – Store directly in S3, flat file storage
  + Volume Gateway (iSCSI): Block base storage
    - Stored Volume – Primary store locally (on-prem) & asynchronies backed up in S3
    - Cached Volume – Primary stored in S3 and most frequently data is cache in on site
  + Tape Gateway (VTL) – Backup application like NetBackup, Veeam etc.
* Snowball
  + Type of Snowball
    - Snowball – Import to S3 and Export from S3
    - Snowball Edge – Compute capability [Run lambda function]
    - Snow mobile
* S3 Transfer Accelerator – <https://[BucketName].s3-accelerator.amazonaws.com>
* Static Website are Server less
* Request Header - <https://docs.aws.amazon.com/AmazonS3/latest/API/RESTObjectPUT.html>

----------------------------------------------------------------------------------------------------------------

**AWS Elastic Compute Cloud (EC2)**

* EC2 Option:
  + On Demand: Fixed Rate per sec/hour
  + Reserved: Capacity reservation (1 – 3-year term)
  + Spot – Bid price, Application have flexible start & end time
  + Dedicated Host: Physical Host EC2 Server (Server-bound software license, Regulator body)
* EC2 Instance Type
  + DR Mc GIFT PX
    - D for Density
    - R for RAM
    - M for main choice for general purpose app
    - C for Compute
    - G for Graphics
    - I for IOPS
    - F for FPGA
    - T for cheap general purpose
    - P for Graphics (thick pics)
    - X for Extreme Memory
* EBS: block base storage
  + EBS Volume Type
    - General Purpose SSD (GP2): 3 IOPS per GB up to 10000 IOPS
    - Provisioned IOPS SSD (IO1): more than 10000 IOPS up to 20000 IOPS
      * Used for No SQL, database
    - Throughput Optimized HDD (ST1)
      * Can’t be boot volumes
      * Used for Big Data, Log processing, Data warehousing
      * Frequently accessed workload
    - Cold HDD (SC1)
      * Can’t be boot volumes
      * File Server
      * Lowest Cost Storage for infrequently accessed workload
    - Magnetic (Standard)
      * Lowest storage cost
      * Infrequently accessed
  + Cannot mount 1 EBS volume to multiple EC2 instance; used EFS for shared instance
  + SSD-backed storage for transactional workloads (performance depends primarily on IOPS)
  + HDD-backed storage for throughput workloads (performance depends primarily on throughput, measured in MB/s)

* EC2 Instance
  + Termination Protection will be turned off by default, you must turn it on
  + On EBS-backed instance, the default action is for the root EBS volume to be deleted when the instance is terminated.
  + EBS Root Volumes of your DEFAULT AMI’s cannot be encrypted.
  + Additional volume can be encrypted.
* Security Group are STATEFULL
  + All Inbound traffic is Blocked by Default
  + All Outbound traffic is Allowed
  + You can specify allow rules, but not deny rules.
  + Cannot block specific IP address; it can be achieved through Network Access Control List
  + If you creating inbound rule allowing traffic in, that traffic is automatically allowed back out again
* Network ACL is STATELESS
* EBS Volume
  + Volume exist in EBS
  + Snapshot exist in S3
  + Snapshot is incremental
  + EBS volume can only by mounted by an EC2 instance in the same AZ
  + Modify Volume – GP2, ST1, SC1, IO1. Can’t modify Magnetic
  + Image from ec2 root volume directly
    - Select Instance > Action > Image > Create Image
  + Image from Volume
    - Create snapshot from Volume > Choose Snapshots > Action > Create Image
  + Move EC2/volumes from one AZ to another AZ
    - Create snapshot from Volume > Choose Snapshots > Action > Create Volume – Select AZ
  + Move EC2/volume from one region to another
    - Create snapshot from Volume > Choose Snapshots > Action > Copy
    - Create AMI from EC2 > Choose AMI > Action > Copy AMI
* Instance Store Volume
  + Instance Store Volumes are sometime called as Ephemeral Storage
  + Instance Store Volume cannot be stopped. If the underlying host fail, you will lose your data
  + local instance store only persists during the life of the instance. This is an inexpensive way to launch instances where data is not stored to the root device
* Elastic Load Balancer
  + Classic Load balancer – HTTP, HTTPS (Secure HTTP), SSL (Secure TCP) and TCP protocols
  + Application Load Balancer – HTTP, HTTPS protocol
  + Network load balancer – TCP (Layer-4) listener.
* Cloud watch
  + Basic monitoring is 5 min interval
  + Detail monitoring is 1 min interval
  + EC2 metrics
    - CPU
    - Network
    - Disk
    - Status Check
* Meta-data – curl <http://169.254.169.254/latest/meta-data>
* Placement Group
  + Placement Group can’t span multiple AZ
  + Low network latency, high network throughput
* EFS (Elastic File System)
  + Support NFSv4 protocol
  + Can scale up to petabytes
  + Can support thousands of concurrent NFS connection
  + Block base storage (Similar to S3)
  + Data are stored across multiple AZ’s within a region
  + Read after write consistency
* Lambda
  + Language: Java, GO, C#, Python, NodeJS
  + Trigger: API Gateway, AWS IoT, Alexa, Cloud Front, Cloud Watch, Code Commit, Cognito, Dynamo DB, Kinesis, S3 and SNS
  + Duration is max 5 min
* Billing
  + If data is transferred between these two instances, it is charged at "Data Transfer Out from EC2 to Another AWS Region" for the first instance and at "Data Transfer In from Another AWS Region" for the second instance
  + Regional Data Transfer rates apply if at least one of the following is true, but is only charged once for a given instance even if both are true:
    - The other instance is in a different Availability Zone, regardless of which type of address is used.
    - Public or Elastic IP addresses are used, regardless of which Availability Zone the other instance is in.
* Elastic IP
  + All accounts are limited to 5 Elastic IP addresses per region
  + By default, every instance comes with a private IP address and an internet routable public IP address. The private address is associated exclusively with the instance and is only returned to Amazon EC2 when the instance is stopped or terminated. The public address is associated exclusively with the instance until it is stopped, terminated or replaced with an Elastic IP address.

----------------------------------------------------------------------------------------------------------------

**Route53**

* DNS
  + “A” record – It is used by computer to translate the name of the domain to IP address
  + Elastic Load Balancer doesn’t have pre-defined IPV4 Address, you resolve to them using DNS name
  + TTL (Time To Live) in seconds – The length that a DNS records are cached on either the Resolving Server or the user local machine
  + Canonical Name (CName) – It can be used to resolve one domain name to another
* Routing Policies
  + Simple
  + Weighted
  + Latency
  + Failover
  + Geo Location

----------------------------------------------------------------------------------------------------------------

**Databases**

* RDS (OLTP)
  + MySQL
  + PostgreSQL
  + Oracle
  + Aurora
  + MariaDB
  + Microsoft SQL Server
* Dynamo DB (No SQL)
* Redshift (Data Warehouse) (OLAP)
* Elastic Cache – In Memory Cache: Memcached, Redis
* DMS (Database Migration Service)
* RDS
  + All database storage will be SSD (GP2, IO1) volumes
  + Different Backup Type
    - Automated Backups: Allow you to recover your database to any point in time with in a retention period ( 7 days to 35 days). Take full daily snapshot. Backup data is stored in S3
    - Database Snapshot
  + Multi AZ is for Disaster Recovery only
  + Read Replica is for performance improvement
    - Used to Scaling, not for DR
    - Must have automatic backups
    - You can have up to 5 read replica
    - Each read replica will have its own DNS end point
    - You can have read replica in second region
* DynamoDB:
  + Stored on SSD storage
  + Spread across 3 geographically distinct data center
  + Eventual Consistent Read (Read)
  + Strong Consistent Read
  + Push button scaling
* Redshift (OLAP) - Data Warehouse service
  + Current available in 1 AZ
* Elasticache (in-memory cache)
  + Memcached
  + Redis
* Aurora
  + 2 copies of your data is contained in each AZ, with a minimum of 3 AZ. 6 copies of each data
  + 2 Type of replica available
    - Aurora Replica – Fail over automatically occurs
    - My SQL Read Replicas

----------------------------------------------------------------------------------------------------------------

**Virtual Private Cloud (VPC)** - <https://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_Subnets.html#vpc-subnet-basics>

* Classless Inter-Domain Routing (CIDR) block
* The first four IP addresses and the last IP address in each subnet CIDR block are not available for you to use, and cannot be assigned to an instance
* VPC can have 3 types of Subnet
  + Public Subnet - If a subnet's traffic is routed to an internet gateway
  + Private Subnet - If a subnet doesn't have a route to the internet gateway
  + VPN-only Subnet - If a subnet doesn't have a route to the internet gateway, but has its traffic routed to a virtual private gateway for a VPN connection
* Network Address Translation (NAT)
  + When creating a NAT instance, Disable Source/Destination check on the instance.
  + NAT Instance should be in public subnet.
  + Compare of NAT Instance and NAT Gateway: <https://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/vpc-nat-comparison.html>
  + NAT Gateway – IPv4
  + Egress Only Internet Gateway – IPv6
* Flow logs can be created 3 level
  + VPC
  + Subnet
  + Network Interface level
* FAQ - <https://aws.amazon.com/vpc/faqs/>
* VPC Peering
  + VPC Peering connection between your own VPCs or with a VPC in another AWS account within a **single region**
  + Transitive Peering not supported
  + You cannot create a VPC peering connection between VPC that have same matching or overlapping CIDR block
  + You cannot create a VPC peering connection between VPC in different regions

----------------------------------------------------------------------------------------------------------------

**Application Services**

* SQS (Simple Queue Service) – message oriented API
  + SQS are poll based system
  + Message are 256 KB in size
  + Messages are kept in the queue from 1 minute to 14 days; default is 4 days.
  + Visibility Time out is the amount of time that the message is invisible in queue after a reader pick up the message
  + Visibility time out max is 12 hours
  + Guarantees Message will be processed at least once.
  + Type
    - Standard Queue (Default queue type) - Standard queues provide at-least-once delivery, which means that each message is delivered at least once.
    - FIFO Queue – FIFO queues provide exactly-once processing, Limited to 300 Transaction per second
* Simple Work Flow Service (SWF) – Task oriented API
  + Three type of Actors
    - Workflow Starter
    - Decider
    - Activity Worker
  + Task is assigned only once and never duplicated
  + Retention period of 1 year for work flow executions
  + API – RegisterDomain
* Simple Notification Service (SNS)
  + Publish-subscribe mechanism paradigm
  + Push based system
  + Protocol Types
    - HTTP
    - HTTPS
    - Email
    - Email-JSON
    - Lambda
    - Application
    - SQS
* Kinesis (Analysis of Stream Data)
  + Kinesis Streams – it consist of Shard, 24 – 7 day retention
  + Kinesis Firehose
  + Kinesis Analytics – SQL queries to Streams and Firehose

----------------------------------------------------------------------------------------------------------------

**AWS White Paper**

* Security Processes
  + <https://aws.amazon.com/compliance/shared-responsibility-model/>
  + -Xen hypervisor
* Well Architected Framework
  + Security
    - Data Protection
    - Privilege Management
    - Infrastructure Protection
    - Detective Controls
  + Reliability
    - Foundation
    - Change Management
    - Failure Management
  + Performance Efficiency
    - Compute
    - Storage
    - Database
    - Space-time Trade off
  + Cost Optimization
    - Matched supply and demand
    - Cost Effective resources
    - Expenditure Awareness
    - Optimizing Over Time
  + Operational Excellence
    - Preparation
    - Operation
    - Response

----------------------------------------------------------------------------------------------------------------

* Security Token Service (STS)
  + Federation (Typically Active Directory)
    - SAML
  + Federation with Mobile App
    - Facebook, Google
  + Cross Account Access
  + Steps
    - Web Application 🡺 Identity Broker 🡺 LDAP Directory (Active Directory) 🡺 AWS Security Token Service 🡺 Identity Broker 🡺 Web Application 🡺 AWS Service (S3)
    - Identity Broker calls GetFederationToken function using IAM credentials
    - Security Token Services return four value to Application - Access key, Secret Access key, token, duration (token life time)

----------------------------------------------------------------------------------------------------------------

**Exam Reference Blog**

* <http://jayendrapatil.com/>