**IAM (Identity Access Management) – Summary**

TBC

**Amazon EC2 (Elastic Compute Cloud) – Summary**

EC2 provides new server instances within minutes.

It helps to pay for the services that you actually use.

**EC2 Pricing Options:**

**On Demand:** Allows user to pay by the hour or by the second without any commitment.

**Reserved**: Provides a capacity reservation with commitment of 1 to 3 years.

**Spot:** Enables user to bid for an instance capacity. It’s beneficial for application with flexible executions.

**Dedicated Servers:** These are Physical dedicated servers, which allow existing server-bound software licenses.

**EC2 Instance Types**

\*\* Detailed knowledge not required for Associate level exam \*\*

**FIGHT DR MC PX**

F1 – Field Programmable Gate Array

I3 – High Speed Storage

G3—Graphics Intensive

H1 – High Disk Throughput

T2 – Lowest Cost General Purpose

D2 – Dense Storage

R4 -- Memory Optimized (RAM) -- Memory Intensive Apps/DBs

M5- General Purpose

C5 – Compute Optimized

P3 – Graphics/General Purpose GPU

X1—Memory Optimized – SAP HANA/Apache Spark etc (Extreme Memory)

Command to install Apache and make EC2 Server a web server

**Service httpd start**

**EBS—Elastic Block Storage**

It allows you to create block storage and attach them to EC2 Instances.

**EBS Volume Type:**

**General Purpose SSD (GP2)**

Balance in price and performance, suitable for less than 10,000 IOps

**Provisioned IOps SSD (IO1)**

Designed for I/O intensive applications, use if you need more than 10,000 IOps

**Throughput Optimized HDD (ST1)**

Big data and cannot be root volume

**Cold HDD ()**

File Server, lowest cost for infrequently accessed workloads, cannot be boot volume

**Magnetic Standard**

Lowest storage cost, can be boot volume

**Elastic Load Balancers**

**Application Load Balancers –** Works on **OSI layer 7,** can make clever routing decisions.

**Network Load Balancers –** Fast speed

**Classic Load Balancers –** Legacy ELB

**X-Forwarded** Provides private address from DNS to EC2 instance

**X-Forwarded For** providespublic IP address

**RDS – Backups, Multi AZ (Availability Zone) and Read Replicas.**

Automated backup can be configured for 1 to 35 days.

Multi Availability Zones are for Disaster Recovery.

Read Replicas are for performance improvement.

**Amazon S3 (Simple Storage Service) – Summary**

S3 is ideal for files storage, image storage but not for OS or Database storage.

Object Based storage.

Data is spread across multiple facilities.

File size can be from 0B to 5TB, storage size is unlimited.

S3 has universal namespace, it must be unique globally.

**Data consistency Model**

Read after write consistency is provided for PUTS of new objects.

Eventual consistency for overwrite PUTS and DELETES (propagates after some time)

S3 is object based.

Key is name of the file

Value is data of file

Version Id

Metadata

**S3 Storage Tiers/Classes**

S3: 99.99% availability, 99.(11 -9s) durability

S3 IA (Infrequently Accessed) – Lower feed than s3 but retrieval in charged.

S3 One Zone IA: 20% lower cost but 99.5% availability

Reduced Redundancy Storage: 99.99% durability

Glacier: Archival, take 3-5 hours to retrieve data (no real time access)

**S3 Charges:**

Storage per GB

Requests (Get, Put, Copy etc)

Storage Management Pricing

Data Management Pricing

Transfer Acceleration

**S3 Security:**

By default all newly created buckets are PRIVATE.

**Bucket Policies:** Applied at bucket level.

**Access Control List:** Applied at object level.

S3 buckets can be configured to create access logs.

Encryption:

In Transit:

SSL/TLS (HTTPS)

At Rest (Server Side Encryption):

S3 Managed Keys: **SSE-S3**

AWS Key Management Service, Managed Keys, **SSE-KMS**

Customer Managed Keys **SSE- C**

CORS Configuration:

Used for inter bucket access, need to provide Origin of request.

<https://aws.amazon.com/s3/faqs/> good read before exam.

**Serverless Computing**

AWS Lambda is a compute service where you can upload your code and create lambda function.

AWS Lambda takes care of provisioning and managing servers that you need to run your code.

AWS Lambda can be used in following ways

Even Driven Compute Service – Runs a function on change even in s3 bucket or DB.

Request Driven Compute Service – Runs a function as response to HTTP request.

White Paper link: <https://d1.awsstatic.com/whitepapers/serverless-architectures-with-aws-lambda.pdf>

**Dynamo DB**

For applications that need single digit millisecond latency at any scale.

Supports both document and key-value data models.

Data is stored on SSD storage

Consistency Models

Eventual consistent reads

Strongly consistent reads

Amazon Dynamo DB is low latency NOSQL database

Consists of Tables items and attributes

Supports two models document and Key-Value pair

JSON, HTML and XML document formats are supported

Two types of PK

Partition Key

Composite Key (Partition Key + Sort Key)

Access is controlled using IAM policies

Dynamodb: LeadingKeys – Fine grained access control can be provided using IAM conditional parameter.

**KMS – Key Management Service**

**Other AWS Services**

**SQS – Simple Queue Service**

SQS is distributed message queuing system.

It allows decoupling of components to keep them independent

Pull based not push based

Standard Queue: Best effort ordering, Message delivered at least once.

FIFO Queue: Ordering is strictly preserved; messages are delivered only once, no duplicates.

**SNS – Simple Notification Service**

It is scalable and highly available service which allows application to push notifications from cloud.

Variety of message types are supported

Push Mechanism

**Elastic Beanstalk**

It is fastest and easiest way to deploy application on AWS.

Full Control over resources used by application can be retained by user or elastic beanstalk can do it.

Elastic BeanStalk Deployment Policies:

* All at once : deployment of everything at once and has downtime

Not ideal for mission critical system, if failed rollback needed to revert back to original version

* Rolling : Add new version in batches, capacity is reduced, instances are updated in batches

Not ideal for performance sensitive systems.

* Rolling with additional Batch: Launches additional batch of instances, deploy new versions in batches.

It maintains full capacity during deployment.

* Immutable Deployment Updates: Deploy new version in fresh group of instances in their own auto-scaling groups. Maintains full capacity, rollback is very easy. Preferable for mission critical applications.

**Kinesis**

There are three core Kinesis services.

Kinesis Video Streams: Securely stream video from connected devices to AWS for analytics and ML

Kinesis Data Streams: Build custom applications to process data in near real time.

Kinesis Firehose: capture, transform, load data streams into AWS data store for near real time analytics with BI tools.