EC2

QUALITY THOUGHT

INFO SYSTEMS (INDIA) PVT. LTD

Amazon EC2 (Elastic Compute Cloud)

- Virtual Servers In The Cloud
  - EC2 is the instance computing platform for servers in the cloud. This service makes up several different features such as Elastic Load Balancer, Auto Scaling, EBS Volumes, and EC2 instances.

  - A deprecated version of Amazon EC2 is known as EC2-classic and is no longer available on new accounts created after December 2013.

EC2 Instances

- Instance Storage
  - EC2 instances can be created with two “types” of storage
  - Instance-store volumes
    - Instance store volumes are considered ephemeral data, the data on the volumes only exists for the duration of the instance life
    - Once the instance is “stopped” or “shutdown” the data is erased
    - The instance can be rebooted and still maintain its ephemeral data
    - Instance-store volumes are virtual devices whose underlying hardware is physically attached to the host computer for the instance
    - AMI

  - EBS backed volumes
    - EBS backed volumes are network attached storage
    - Provide persistent data across EC2 instances even if they are shutdown
    - AMI

EC2 EBS Volumes

- EBS Volumes measure input/output operations in IOPS
- IOPS are input/output operations per second
- AWS measures IOPS as 256KB or smaller
- Operations that are greater than 256KB are separated into 256KB units
- A 512KB operation would count as 2 IOPS
- The type of EBS volume you specify greatly influences the I/O performance or IOPS your device will receive. It is important as architects to understand if our application requires more I/O to the EBS volume.
- Even volumes with provisioned IOPS may not produce the performance you expect. If this is the case an EBS optimized instance is required which prioritizes EBS traffic over the network OR an instance with higher network traffic capacity.

EBS Volume Types

- General Purpose SSD
  - Commonly used as the "root" volume on a system
  - Use on dev/test environments and smaller DB instances
  - 3 IOPS/GiB (burstable with baseline performance)
  - Volume size of 1GiB to 16TiB
  - Considerations when using T2 instances with SSD root volumes (burstable vs. baseline performance)
- Provisioned IOPS
  - Mission critical applications that require sustained IOPS performance
  - Large database workloads
  - Volume size of 4GiB to 16TiB
  - Performs at provisioned level and can provision up to 20,000 IOPS
- Magnetic
  - Low storage cost
  - Workloads where performance is not important or data is infrequently accessed
  - Volume size of Min 1GiB Max 1024 GiB

Note: Pre-warming Volumes

EBS Snapshots

- Pay attention, has proven to be a tough concept for some students

- Snapshots are incremental in nature
  - A snapshot only stores the changes since the most recent snapshot thus reducing costs and only having to pay for storage for the "incremental changes" between snapshots
  - What happens when the original snapshot is deleted?
    - The data is still available, snapshot storage might only charge you as an incremental snapshot but the prior data is still there
    - Think about it like this you have "snapshots" point in time but the actual source file is dynamically growing. If you delete old snapshots the data in the source location still exists.

- Frequent snapshots of your data increases data durability

- When a snapshot is being taken against the EBS volume it can degrade performance so snapshots should occur during non-production or non-peak load hours

EC2

- Starting, stopping, and terminating instances
- IAM users with proper permissions
- While an instance is stopped you are not paying for compute time only for storage
- Termination and termination protection
- AWS Command Line Interface / SDK
- User-Data/Cloud-init

- Access an instances user-data and meta-data by opening this URL WITHIN the instance
  - http://169.254.169.254/latest/meta-data or http://169.254.169.254/latest/user-data
    - More examples in the course
    - Information such as user-data and ami-launch-index if launched as a group
    - Use to register an instance-id as part of a cluster or application suite AUTOMATION!

Commands For Checking user 'N' META DATA

curl http://169.254.169.254/latest/meta-data/

curl http://169.254.169.254/latest/user-data/

EC2-Classic

- YES! It's still important
- Instances are assigned a public IP address and cname
- Each instance receives a private IP address but is NOT part of a VPC
- This means if an instance is shut down It will lose its private IP and will change on next boot up

Security Groups

- Security groups are used as a firewall in front of an EC2 instance
- An instance can belong to multiple security groups
- Security groups can reference themselves as "source" traffic in firewall rules

Placement Groups

- A placement group is a cluster of instances within the same availability zones
- Instances within a placement group have a low-latency, 10 Gbps network connections between them.
- Used for instances that run applications whose requirements are an extremely low latency network between them.
- Instances that are in the placement group need to have enhanced networking in order to maximize placement groups.

Troubleshooting Placement Groups

- If an instance in a placement group is stopped once it is started again it will continue to be a member of the placement group

- It is suggested to launch all the required instances within a placement group in a single request and that the same instance type is used for all instances within the placement group
  - AWS attempts to place all the instances as close as physically possible to reduce latency

- It is possible, if more instances are added at a later time to the placement group OR if a placement group instance is stopped and started again, to receive an "insufficient capacity error".

  - Resolve the capacity error by stopping all instances in the member group and attempting to start them again.

Troubleshooting Placement Groups

- Placement group keys
  - Instances not originally launched/created in the placement group cannot be moved into the placement group
  - Placement groups cannot be merged together
  - A placement group cannot span multiple availability zones
  - Placement group names must be unique within your own AWS account
  - Placement groups can be "connected"
  - Instances must have 10 gigabit network speeds in order to take advantage of placement groups

AMI CREATION

Creation of AMI From Running Instance

http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/creating-an-ami-ebs.html

http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/creating-an-ami-instance-store.html