**AWS Assignment 4**

1. Explain the Snowball concept.

**Answer:** AWS Snowball is an edge computing, data migration, and edge storage appliance with two options. Snowball Edge Storage Optimized devices provide both block storage and Amazon S3 compatible object storage and 40 vCPUs. Ideal for local storage and transfer of large amounts of data. The Snowball Edge Compute Optimized device offers 52 vCPUs, block and object storage, and an optional GPU for use cases such as advanced machine learning and full-motion video analytics in disconnected environments. These devices can be used in environments (manufacturing, industry, transportation, etc.) or very remote locations (military or maritime operations, etc.) where data acquisition, machine learning and processing, and intermittent connectivity are possible before they are returned to AWS. ) Can be used for saving. These devices can also be rack-mounted and clustered to create larger temporary installations.

1. Make a distinction between NAT Gateways and NAT Instances.

**Answer:**

|  |  |  |
| --- | --- | --- |
| **Attribute** | **NAT gateway** | **NAT instance** |
| Availability | High availability. The NAT gateway for each Availability Zone is redundantly implemented. Create a NAT gateway in each Availability Zone to ensure a zone-independent architecture. | Use a script to manage failover between instances. |
| Bandwidth | Scale up to 45 Gbps. | Depends on the bandwidth of the instance type. |
| Maintenance | Managed by AWS. You do not need to perform any maintenance. | Managed by you, for example, by installing software updates or operating system patches on the instance. |
| Performance | Software is optimized for handling NAT traffic. | A generic AMI that's configured to perform NAT. |
| Cost | Charged depending on the number of NAT gateways you use, duration of usage, and amount of data that you send through the NAT gateways. | Charged depending on the number of NAT instances that you use, duration of usage, and instance type and size. |
| Type and size | Uniform offering; you don’t need to decide on the type or size. | Choose a suitable instance type and size, according to your predicted workload. |
| Public IP addresses | Choose the Elastic IP address to associate with a public NAT gateway at creation. | Use an Elastic IP address or a public IP address with a NAT instance. You can change the public IP address at any time by associating a new Elastic IP address with the instance. |
| Private IP addresses | Automatically selected from the subnet's IP address range when you create the gateway. | Assign a specific private IP address from the subnet's IP address range when you launch the instance. |
| Security groups | You cannot associate security groups with NAT gateways. You can associate them with the resources behind the NAT gateway to control inbound and outbound traffic. | Associate with your NAT instance and the resources behind your NAT instance to control inbound and outbound traffic. |

1. Describe the essential components of Amazon Web Services (AWS).

**Answer: The essential components of AWS are:**

1. **Data management and data transfer:** To run HPC applications in the AWS cloud, you need to move the required data to the cloud. There are several data transfer solutions designed to securely transfer large amounts of data. This overcomes issues such as long transmission times, high network costs, and security concerns. You can also automate the movement of data between the AWS cloud and local storage. There are several ways to connect to AWS privately from your area. This increases bandwidth, increases throughput, reduces network costs, and provides a consistent network experience.
2. **Computer and network:** There are several compute instance types that you can customize to suit your needs. It also monitors applications and adjusts their capacity to maintain constant predictable performance at an affordable cost. Setting up application scaling across multiple services for multiple resources also takes a few minutes. AWS's advanced network options allow you to reduce latency and increase bandwidth between instances.
3. **Storage:** When looking for an HPC solution, you need to consider storage options and costs. AWS services have several flexible block, object, and file storage options that allow persistent and temporary data storage. You can allocate storage volumes according to the size you want. You can store and access data types via the cloud without having to run a data migration project. You can also use AWS services to move your workload from on-premises to the cloud.
4. **Automation and orchestration:** To use the infrastructure efficiently, you need to automate the scheduling of submitted jobs and the job submission process. AWS services can run thousands of batch computing jobs by dynamically allocating computing resources based on demand.
5. **Operation and management:** The system administrator is responsible for preventing cost overruns and monitoring the infrastructure. There are several management and monitoring services that optimize resource usage, monitor applications, get a complete view of operational status, and adapt to performance changes.
6. **Visualization:** AWS services make it easy to visualize the results of engineering simulations without having to move large amounts of data. You can now remotely access your interactive application over a standard network and deliver your application session to any workstation.
7. **Security and compliance:** To run your application in the cloud, you need to understand regulatory compliance and security management. There are several quick launch templates and security-related services provided by AWS that help you protect your data and privacy by establishing strong security safeguards for your AWS infrastructure.
8. When should you utilize a spin-up server? Use example to demonstrate your point.

**Answer:** We utilize the spin-up server very often when we want to perform a service using machines. Using the Amazon Web Services, we have a simple step to follow, and a server is spined up for performing a service.

**Example:** When working with systems and application working on Big Data, Low-latency response application and websites which have a peak load all time or some specific time of the day and we can just quickly follow some steps and scale up and scale down our servers for performing our services. So, spinning up a server in a matter of minutes really help resolve many use cases around the industries.

1. Explain the concept of outlier car scaling.

**Answer:** Outliers are observations that are anomalous distances from other values ​​in a random sample of the population. The difference between a good machine learning model and an average model is often the ability to cleanse the data. One of the biggest challenges in data cleansing is identifying and processing outliers. Simply put, outliers are observations that are significantly different from other data points. The most effective way to find all outliers is to use the interquartile range (IQR). The IQR contains the central part of the data, so once you know the IQR, you can easily find the outliers.