

Lovely Professional University
Virtualization and Cloud Computing
Term Paper: Amazon EC2



Submitted By: Prince Paraste
Submitted To: Umer Iqbal Wani
Section: K1629
Roll No: B60
Registration No: 11604723

Amazon EC2

Prince Paraste

Lovely Professional University, Phagwara, Punjab, India

ABSTRACT: This Term Paper for Amazon Elastic Compute Cloud (Amazon EC2). In this I have explain the Features, Benefits of Amazon EC2, featured customers of Amazon EC2, Pricing for using Amazon EC2, EC2 Instance Type, Methods so that one can Migrate Amazon EC2. This Term Paper also explains about the Amazon EC2 Auto Scaling, Amazon EC2 CLI (Command Line Interface), Amazon EC2 API Actions. Networking and Security and finally Method to build an Amazon EC2 Instance.

I. INTRODUCTION

AWS Elastic Compute Service or EC2 is IaaS (Infrastructure as a Service). Amazon Elastic Compute Cloud (Amazon EC2) is a web service that has secure, resizable figure capability within the cloud. It is designed to create web-scale cloud computing easier for developers. Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you use. Amazon EC2 provides developers the tools to build failure resilient applications and isolate them from common failure scenarios.

II. FOLLOWING ARE THE FEATURES THAT AWS EC2 GIVES:

[A]. FUNCTIONALITY

Amazon EC2 presents a true virtual computing environment, allowing you to use web service interfaces to launch instances with a variety of operating systems, load them with your custom application environment, manage your network's access permissions, and run your image using as many or few systems as you desire.

To use Amazon EC2, you simply:

- Select a pre-configured, templated Amazon Machine Image (AMI) to get up and running immediately. Or create an AMI containing your applications, libraries, data, and associated configuration settings.
- Configure security and network access on your Amazon EC2 instance.
- Choose which instance type(s) you want, then start, terminate, and monitor as many instances of your AMI as needed, using the web service APIs or the variety of management tools provided.
- Determine whether you want to run in multiple locations, utilize static IP endpoints, or attach persistent block storage to your instances.
- Pay only for the resources that you consume, like instance-hours or data transfer.

Features:

Amazon EC2 provides several powerful features for building scalable, failure resilient, enterprise class applications.

BARE METAL INSTANCE

Amazon EC2 bare metal instances provide your applications with direct access to the processor and memory of the underlying server. These instances are ideal for workloads that require access to hardware feature sets (such as Intel® VT-x), or for applications that need to run in non-virtualized environments for licensing or support requirements. Bare metal instances are built on the Nitro system, a collection of AWS-built hardware offload and hardware protection components that come together to securely provide high performance networking and storage resources to EC2 instances. Bare metal instances are EC2 instances and thus offer the same robust security, reliability, capacity elasticity, and support for different operating systems and software packages as other virtual EC2 instances. You can also use bare metal instances with AWS services such as Amazon Virtual Private Cloud (VPC), Elastic Block Store (EBS), Elastic Load Balancing (ELB) and more.

COMPUTE PERFORMANCE AND COST WITH AMAZON EC2 FLEET

With a single API call, Amazon EC2 Fleet lets you provision compute capacity across EC2 instance types, Availability Zones, and purchase models to help optimize scale, performance and cost. You can specify how much On-Demand and Spot capacity to launch via EC2 Fleet. You can also define which instance types you prefer and whether to scale capacity based on cores, instances or memory. Now you can access EC2 Fleet capabilities via Amazon EC2 Auto Scaling to provision and automatically scale compute capacity across EC2 instance types, Availability Zones, and purchase options in a single Auto Scaling Group.

PAUSE AND RESUME INSTANCE

You can hibernate your Amazon EC2 instances backed by Amazon EBS and resume them from this state at a later time. Applications that take a while to bootstrap and persist state into memory (RAM) can benefit from this feature. Hibernation gives you all the benefits of Stop and Start, and additionally, data from memory (RAM) is also persisted between sessions. AWS not charged for instance usage while your instance is hibernated. Storage is charged at standard EBS rates.

GPU COMPUTE INSTANCE

Customers requiring massive floating-point processing power will benefit from the next-generation of general-purpose GPU compute instances from AWS, Amazon EC2 P3 instances with up to 8 NVIDIA® V100 Tensor Core GPUs. P3 instances provide up to 1 petaFLOPS of mixed-precision, 125 teraFLOPS of single-precision and 62 teraFLOPS of double-precision floating point performance. A 300 GB/s second generation NVLink interconnect allows GPU-to-GPU communication at high speed and low latency. P3 instances also feature up to 96 vCPUs based on custom Intel processors, 768 GB of DRAM, and 100 Gbps of dedicated aggregate network bandwidth using the Elastic Network Adapter (ENA). P3 instances are ideally suited for machine learning, high-performance computing, computational fluid dynamics, computational finance, seismic analysis, molecular modeling, genomics, and rendering workloads.

GPU GRAPHICS INSTANCES

Customers requiring high graphics capability will benefit from GPU graphics instances. The current generation GPU graphics instance, G3 instance, provides access to NVIDIA Tesla M60 GPUs, each with up to 2,048 parallel processing cores, 8 GiB of GPU memory and a hardware encoder supporting up to 10 H.265 (HEVC) 1080p30 streams and up to 18 H.264 1080p30 streams. With the latest driver releases, these GPUs provide support for OpenGL, DirectX, CUDA, OpenCL, and Capture SDK (formerly known as GRID SDK). GPU graphics instances are ideally suited for 3D visualizations, graphics-intensive remote workstation, 3D rendering, application streaming, video encoding, and other server-side graphics workloads.

HIGH I/O INSTANCES

Customers requiring very high, low latency, random I/O access to their data can benefit from High I/O instances. High I/O instances are an Amazon EC2 instance type that can provide customers with random I/O rates over 3 million IOPS. High I/O I3 instances are backed by Non-Volatile Memory Express (NVMe) based SSDs and are ideally suited for customers running very high-performance NoSQL databases, transactional systems, and Elastic Search workloads. High I/O instances also offers sequential disk throughput up to 16 GB/s, which is ideal for analytics workloads.

See Amazon EC2 Instance Types to find out more about High I/O instances.

DENSE STORAGE INSTANCES

Customers requiring very high storage density per instance, and high sequential I/O for data-intensive applications like Massively Parallel Processing (MPP) data warehouse, MapReduce and Hadoop distributed computing, and log and data processing can benefit from Dense Storage instances. Dense Storage instances are an Amazon EC2 instance type that can provide customers with sequential I/O throughout of up to 3.9 GB/s and provide customers with up to 48 TB of instance storage across 24 hard disk drives or offer a balance with lesser storage and memory per vCPU with ENA based networking for up to 25 Gbps of Network Bandwidth within a placement group. See Amazon EC2 Instance Types to find out more about Dense Storage instances.

OPTIMIZED CPU CONFIGURATIONS

Optimize CPUs feature gives you greater control of your Amazon EC2 instances on two fronts. First, you can specify a custom number of vCPUs when launching new instances to save on vCPU-based licensing costs. Second, you can disable Intel Hyper-Threading Technology (Intel HT Technology) for workloads that perform well with single-threaded CPUs, such as certain high-performance computing (HPC) applications.

FLEXIBLE STORAGE OPTIONS

Different Amazon EC2 workloads can have vastly different storage requirements. Beyond the built-in instance storage, we also have Amazon Elastic Block Store (Amazon EBS) and Amazon Elastic File System (Amazon EFS) to suit other cloud storage workload requirements.

Amazon EBS provides persistent, highly available, consistent, low-latency block storage volumes for use with Amazon EC2 instances. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability. It is designed for application managers who need to tune workloads for capacity, performance and cost.

Amazon EFS provides simple, scalable, persistent, fully managed cloud file storage for shared access. Designed for high availability and durability across multiple Availability Zones it delivers a file system interface with standard file system access semantics, grows and shrinks capacity automatically, and provides application managers with high throughput and consistently low latencies at petabyte scale.

PAYING FOR WHAT YOU USE

You will be charged at the end of each month for your EC2 resources consumed.

As an example, assume you launch 20 instances of the Small type costing \$0.085 per hour at some point in time. The instances will begin booting immediately, but they won't necessarily all start at the same moment. Each instance will store its actual launch time. Thereafter, each instance will charge for its hours (at \$.085/hour) of execution at the beginning of each hour relative to the time it launched. Each instance will run until one of the following occurs: you terminate the instance with the Terminate Instances API call (or an equivalent tool), the instance shuts itself down (e.g. UNIX "shutdown" command), or the host terminates due to software or hardware failure. Partial instance hours consumed are billed as full hours.

MULTIPLE LOCATIONS

Amazon EC2 provides the ability to place instances in multiple locations. Amazon EC2 locations are composed of Regions and Availability Zones. Availability Zones are distinct locations that are engineered to be insulated from failures in other Availability Zones and provide inexpensive, low latency network connectivity to other Availability Zones in the same Region. By launching instances in separate Availability Zones, you can protect your applications from failure of a single location. Regions consist of one or more Availability Zones, are geographically dispersed, and will be in separate geographic areas or countries. The Agreement

commitment is 99.99% availability for each Amazon EC2 Region. Please refer to Regional Products and Services for more details of our product and service availability by region.

ELASTIC IP ADDRESSES

Elastic IP addresses are static IP addresses designed for dynamic cloud computing. An Elastic IP address is associated with your account not a instance, and you control that address until you choose to explicitly release it. Unlike traditional static IP addresses, however, Elastic IP addresses allow you to mask instance or Availability Zone failures by programmatically remapping your public IP addresses to any instance in your account. Rather than waiting on a data technician to reconfigure or replace your host or waiting for DNS to propagate to all your customers, Amazon EC2 enables you to engineer around problems with your instance or software by quickly remapping your Elastic IP address to a replacement instance

AMAZON EC2 AUTO SCALING

Amazon EC2 Auto Scaling allows you to automatically scale your Amazon EC2 capacity up or down according to conditions you define. With EC2 Auto Scaling, you can ensure that the number of Amazon EC2 instances you're using scales up seamlessly during demand spikes to maintain performance, and scales down automatically during demand lulls to minimize costs. EC2 Auto Scaling is particularly well suited for applications that experience hourly, daily, or weekly variability in usage. EC2 Auto Scaling is enabled by Amazon CloudWatch and available at no additional charge beyond Amazon CloudWatch fees.

HIGH PERFORMANCE COMPUTING (HPC) CLUSTERS

Customers with complex computational workloads such as tightly coupled parallel processes, or with applications sensitive to network performance, can achieve the same high compute and network performance provided by custom-built infrastructure while benefiting from the elasticity, flexibility and cost advantages of Amazon EC2. Cluster Compute, Cluster GPU, and High Memory Cluster instances have been specifically engineered to provide high-performance network capability and can be programmatically launched into clusters – allowing applications to get the low-latency network performance required for tightly coupled, node-to-node communication. Cluster instances also provide significantly increased throughput making them well suited for customer applications that need to perform network-intensive operations.

ENHANCED NETWORKING

Enhanced Networking enables you to get significantly higher packet per second (PPS) performance, lower network jitter and lower latencies. This feature uses a new network virtualization stack that provides higher I/O performance and lower CPU utilization compared to traditional implementations. In order to take advantage of Enhanced Networking, you should launch an HVM AMI in VPC, and install the appropriate driver.

ELASTIC FABRIC ADAPTER (FAST INTERCONNECT FOR HPC CLUSTERS)

Elastic Fabric Adapter (EFA) is a network interface for Amazon EC2 instances that enables customers to run HPC applications requiring high levels of inter-instance communications, like computational fluid dynamics, weather modelling, and reservoir simulation, at scale on AWS. It uses a custom-built operating system bypass technique to enhance the performance of inter-instance communications, which is critical to scaling HPC applications. With EFA, HPC applications using popular HPC technologies like Message Passing Interface (MPI) can scale to thousands of CPU cores. EFA supports industry-standard libfabric APIs, so applications that use a supported MPI library can be migrated to AWS with little or no modification.

EFA is available as an optional EC2 networking feature that you can enable on C5n.9x1, C5n.18x1, and P3dn.24x1 instances.

AVAILABLE ON AWS PRIVATE LINK

Customers can privately access Amazon EC2 APIs from their Amazon Virtual Private Cloud (VPC) or over AWS Direct Connect, without using public IPs, and without requiring the traffic to traverse across the Internet. AWS Private Link is a purpose-built technology designed for customers to access Amazon services in a high performance and highly available manner, while keeping all the network traffic within the AWS network. To use Amazon EC2 with AWS Private Link, you will need to create an endpoint for EC2 in your VPC. Any traffic destined to this endpoint will get privately routed to the EC2 service.

AMAZON TIME SYNC SERVICE

The Amazon Time Sync Service provides a highly accurate, reliable and available time source to AWS services including EC2 instances. All instances running in a VPC can access the service at a universally reachable IP address. The service uses a fleet of redundant satellite-connected and atomic reference clocks in AWS regions to deliver highly accurate and reliable current time readings of the Coordinated Universal Time (UTC) global standard.

[B]. OPERATING SYSTEMS

Amazon Machine Images (AMIs) are preconfigured with an ever-growing list of operating systems. We work with our partners and community to provide you with the most choice possible. You are also empowered to use our bundling tools to upload your own operating systems. The operating systems currently available to use with your Amazon EC2 instances include:



[Amazon Linux](#)

Amazon

\$0.02 to \$5.67/hr include EC2 charges
+ other AWS usage fees



[Windows Server 2012](#)

Microsoft
\$0.02 to \$4.72/hr include EC2 charges + other AWS usage fees



CentOS 6.5

CentOS

\$0.00/hr for software + AWS usage fees



Debian 7.4

Debian

\$0.00/hr for software + AWS usage fees

[C]. SOFTWARE

AWS Marketplace features a wide selection of commercial and free software from well-known vendors, designed to run on your EC2 instances. A sample of products are below.



SAP BusinessObjects

SAP

\$150.00/month + \$0.50/hr for software + AWS usage fees



LAMP Stacks

LAMP

From \$0.00/hr for software + AWS usage fees



Drupal

From \$0.00/hr for software + AWS usage fees

III. BENEFITS OF AMAZON EC2:

ELASTIC WEB-SCALE COMPUTING

Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously. You can also use Amazon EC2 Auto Scaling to maintain availability of your EC2 fleet and

automatically scale your fleet up and down depending on its needs to maximize performance and minimize cost. To scale multiple services, you can use AWS Auto Scaling.

COMPLETELY CONTROLLED

You have complete control of your instances including root access and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition, and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

FLEXIBLE CLOUD HOSTING SERVICES

You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating system and application. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

INTEGRATED

Amazon EC2 is integrated with most AWS services such as Amazon Simple Storage Service (Amazon S3), Amazon Relational Database Service (Amazon RDS), and Amazon Virtual Private Cloud (Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

RELIABLE

Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers. The Amazon EC2 Service Level Agreement commitment is 99.99% availability for each Amazon EC2 Region.

SECURE

Cloud security at AWS is the highest priority. As an AWS customer, you will benefit from a data centre and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your computer resources.

INEXPENSIVE

Amazon EC2 passes on to you the financial benefits of Amazon's scale. You pay a very low rate for the compute capacity you consume.

EASY TO START

There are several ways to get started with Amazon EC2. You can use the AWS Management Console, the AWS Command Line Tools (CLI), or AWS SDKs. AWS is free to get started.

IV. FEATURED CUSTOMERS

Following are the some of the famous companies that uses AWS:



V. AMAZON EC2 PRICING

Amazon EC2 is free to try. There are four ways to pay for Amazon EC2 instances: On-Demand, Reserved Instances, and Spot Instances. You can also pay for Dedicated Hosts which provide you with EC2 instance capacity on physical servers dedicated for your use.

[A]. FREE TIER

AWS Free Tier includes 750 hours of Linux and Windows t2.micro instances each month for one year. To stay within the Free Tier, use only EC2 Micro instances.

(i). On-Demand

With On-Demand instances, you pay for compute capacity by per hour or per second depending on which instances you run. No longer-term commitments or upfront payments are needed. You can increase or decrease your compute capacity depending on the demands of your application and only pay the specified per hourly rates for the instance you use.

On-Demand instances are recommended for:

- Users that prefer the low cost and flexibility of Amazon EC2 without any up-front payment or long-term commitment
- Applications with short-term, spiky, or unpredictable workloads that cannot be interrupted
- Applications being developed or tested on Amazon EC2 for the first time

(ii). Spot instances

Amazon EC2 Spot instances allow you to request spare Amazon EC2 computing capacity for up to 90% off the On-Demand price.

Spot instances are recommended for:

- Applications that have flexible start and end times
- Applications that are only feasible at very low compute prices
- Users with urgent computing needs for large amounts of additional capacity

(iii). Reserved Instances

Reserved Instances provide you with a significant discount (up to 75%) compared to On-Demand instance pricing. In addition, when Reserved Instances are assigned to a specific Availability Zone, they provide a capacity reservation, giving you additional confidence in your ability to launch instances when you need them.

For applications that have steady state or predictable usage, Reserved Instances can provide significant savings compared to using On-Demand instances.

Reserved Instances are recommended for:

- Applications with steady state usage
- Applications that may require reserved capacity
- Customers that can commit to using EC2 over a 1 or 3 year term to reduce their total computing costs

(iv). Dedicated Hosts

A Dedicated Host is a physical EC2 server dedicated for your use. Dedicated Hosts can help you reduce costs by allowing you to use your existing server-bound software licenses, including Windows Server, SQL Server, and SUSE Linux Enterprise Server (subject to your license terms), and can also help you meet compliance requirements.

- Can be purchased On-Demand (hourly).
- Can be purchased as a Reservation for up to 70% off the On-Demand price.

[B]. PER SECOND BILLING

With per-second billing, you pay for only what you use. It takes cost of unused minutes and seconds in an hour off of the bill, so you can focus on improving your applications instead of maximizing usage to the hour. Especially, if you manage instances running for irregular periods of time, such as dev/testing, data processing, analytics, batch processing and gaming applications, can benefit.

EC2 usage are billed on one second increments, with a minimum of 60 seconds. Similarly, provisioned storage for EBS volumes will be billed per-second increments, with a 60 second minimum. Per-second billing is available for instances launched in:

- On-Demand Reserved and Spot forms
- All regions and Availability Zones
- Amazon Linux and Ubuntu

VI. AMAZON EC2 INSTANCE TYPES

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your applications. Each instance type includes one or more instance sizes, allowing you to scale your resources to the requirements of your target workload

Some of them are as follows:

1. **A1:** Amazon EC2 A1 instances deliver significant cost savings and are ideally suited for scale-out and Arm-based workloads that are supported by the extensive Arm ecosystem. A1 instances are the first EC2 instances powered by AWS Graviton Processors that feature 64-bit Arm Neoverse cores and custom silicon designed by AWS.

Features:

- Custom built AWS Graviton Processor with 64-bit Arm Neoverse cores
- Support for Enhanced Networking with Up to 10 Gbps of Network bandwidth
- EBS-optimized by default
- Powered by the AWS Nitro System, a combination of dedicated hardware and lightweight hypervisor.

SPECS OF A1 INSTANCE:

Model	vCPU	Mem (GiB)	Storage	Network Performance (Gbps)
a1.medium	1	2	EBS-Only	Up to 10
a1.large	2	4	EBS-Only	Up to 10
a1.xlarge	4	8	EBS-Only	Up to 10
a1.2xlarge	8	16	EBS-Only	Up to 10
a1.4xlarge	16	32	EBS-Only	Up to 10

2. **T3:** T3 instances are the next generation burstable general-purpose instance type that provide a baseline level of CPU performance with the ability to burst CPU usage at any time for as long as required. T3 instances offer a balance of compute, memory, and network resources and are designed for applications with moderate CPU usage that experience temporary spikes in use.

Features:

- Burstable CPU, governed by CPU Credits, and consistent baseline performance
- Unlimited mode by default to ensure performance during peak periods and Standard mode option for a predictable monthly cost

- Powered by the AWS Nitro System, a combination of dedicated hardware and lightweight hypervisor
- AWS Nitro System and high frequency Intel Xeon Scalable processors result in up to a 30% price performance improvement over T2 instances

SPECS OF T3 INSTANCE:

Model	vCPU*	Mem (GiB)	Storage	Network Performance (Gbps)
t3.nano	2	0.5	EBS-Only	Up to 5
t3.micro	2	1	EBS-Only	Up to 5
t3.small	2	2	EBS-Only	Up to 5
t3.medium	2	4	EBS-Only	Up to 5
t3.large	2	8	EBS-Only	Up to 5
t3.xlarge	4	16	EBS-Only	Up to 5
t3.2xlarge	8	32	EBS-Only	Up to 5

3. **T2:** T2 instances are Burstable Performance Instances that provide a baseline level of CPU performance with the ability to burst above the baseline.

T2 Unlimited instances can sustain high CPU performance for as long as a workload needs it. For most general-purpose workloads, T2 Unlimited instances will provide ample performance without any additional charges. If the instance needs to run at higher CPU utilization for a prolonged period, it can also do so at a flat additional charge of 5 cents per vCPU-hour.

Features:

- High frequency Intel Xeon processors
- Burstable CPU, governed by CPU Credits, and consistent baseline performance
- Lowest-cost general purpose instance type, and Free Tier eligible*
- Balance of compute, memory, and network resources

SPECS OF T2 INSTANCE:

Model	vCPU*	Storage	Network Performance
t2.nano	1	EBS-Only	Low
t2.micro	1	EBS-Only	Low to Moderate
t2.small	1	EBS-Only	Low to Moderate
t2.medium	2	EBS-Only	Low to Moderate
t2.large	2	EBS-Only	Low to Moderate
t2.xlarge	4	EBS-Only	Moderate
t2.2xlarge	8	EBS-Only	Moderate

4. **C5n:** C5n instances are ideal for high compute applications (including High Performance Computing (HPC) workloads, data lakes, and network appliances such as firewalls and routers) that can take advantage of improved network throughput and packet rate performance. C5n instances offers up to 100 Gbps

network bandwidth and increased memory over comparable C5 instances.

Features:

- 3.0 GHz Intel Xeon Platinum processors with Intel Advanced Vector Extension 512 (AVX-512) instruction set
- Run each core at up to 3.5 GHz using Intel Turbo Boost Technology
- Larger instance size, c5n.18xlarge, offering 72 vCPUs and 192 GiB of memory
- Requires HVM AMIs that include drivers for ENA and NVMe
- Network bandwidth increases to up to 100 Gbps, delivering increased performance for network intensive applications.
- 33% higher memory footprint compared to C5 instances

Model	vCPU*	Mem (GiB)	Storage (GiB)
c5n.large	2	5.25	EBS-Only
c5n.xlarge	4	10.5	EBS-Only
c5n.2xlarge	8	21	EBS-Only
c5n.4xlarge	16	42	EBS-Only
c5n.9xlarge	36	96	EBS-Only
c5n.18xlarge	72	192	EBS-Only

5. **C4:** C4 instances are optimized for compute-intensive workloads and deliver very cost-effective high performance at a low price per compute ratio.

Features:

- High frequency Intel Xeon E5-2666 v3 (Haswell) processors optimized specifically for EC2
- Default EBS-optimized for increased storage performance at no additional cost
- Higher networking performance with Enhanced Networking supporting Intel 82599 VF
- Requires Amazon VPC, Amazon EBS and 64-bit HVM AMIs

Model	Storage	Dedicated EBS Bandwidth (Mbps)	Network Performance
c4.large	EBS-Only	500	Moderate
c4.xlarge	EBS-Only	750	High
c4.2xlarge	EBS-Only	1,000	High
c4.4xlarge	EBS-Only	2,000	High
c4.8xlarge	EBS-Only	4,000	10 Gigabit

INSTANCE FEATURES

Amazon EC2 instances provide a number of additional features to help you deploy, manage, and scale your applications.

BURSTABLE PERFORMANCE INSTANCE

Amazon EC2 allows you to choose between Fixed Performance Instances (e.g. M5, C5, and R5) and Burstable Performance Instances (e.g. T3). Burstable Performance Instances provide a baseline level of CPU performance with the ability to burst above the baseline.

MULTIPLE STORAGE OPTION

Amazon EC2 allows you to choose between multiple storage options based on your requirements. Amazon EBS is a durable, block-level storage volume that you can attach to a single, running Amazon EC2 instance. we can use Amazon EBS as a primary storage device for data that requires frequent and granular updates.

CLUSTER NETWORKING

Select EC2 instances support cluster networking when launched into a common cluster placement group. A cluster placement group provides low-latency networking between all instances in the cluster. The bandwidth an EC2 instance can utilize depends on the instance type and its networking performance specification. Inter instance traffic within the same region can utilize up to 5 Gbps for single-flow and up to 25 Gbps for multi-flow traffic in each direction (full duplex). Traffic to and from S3 buckets in the same region can also utilize all available instance aggregate bandwidth.

MEASURING INSTANCE PERFORMANCE

Amazon EC2 allows you to provision a variety of instances types, which provide different combinations of CPU, memory, disk, and networking. Launching new instances and running tests in parallel is easy, and we recommend measuring the performance of applications to identify appropriate instance types and validate application architecture. We also recommend rigorous load/scale testing to ensure that your applications can scale as you intend.

WHAT IS VM IMPORT/EXPORT?

VM Import/Export enables you to import virtual machine (VM) images from your existing virtualization environment to Amazon EC2, and then export them back. This enables you to migrate applications and workloads to Amazon EC2, copy your VM image catalog to Amazon EC2, or create a repository of VM images for backup and disaster recovery.

VII. MIGRATING VIRTUAL MACHINE

To migrate a VM to Amazon EC2, use vCenter with the connector. The connector can migrate up to four VMs concurrently.

To migrate your virtual machine to Amazon EC2

1. From vCenter, click **Home** and then click **VMs and Templates**.
2. Select the VM.
3. Right-click the VM, and then click **Migrate VM to EC2**. If your administrator did not grant you permission to migrate VMs, you'll see a message to ask your administrator to grant you permission.

4. Complete the form as follows:

1. Select the operating system running on the VM.
2. Select the region and environment for the resulting EC2 instance. The list of environments contains only the environments to which your administrator has granted you permission.
3. Select a subnet, instance type, and security group for the instance.
4. Select a security group. The list of security groups contains only the security groups associated with the environment you've selected.
5. Click Begin migration to Amazon EC2.
6. If the connector displays a warning that there are already four active migration tasks and that this will affect the speed of these tasks, you can either continue or cancel the migration task.

5. After the migration begins, we display the import task ID if the migration task started immediately or the queued task ID otherwise. Note the ID if you want to monitor the migration task. Otherwise, you can close the import window and your vSphere client after the connector notifies you that the import task was created or queued, and the migration will continue.

6. (Optional) To monitor the status of the migration, do the following:

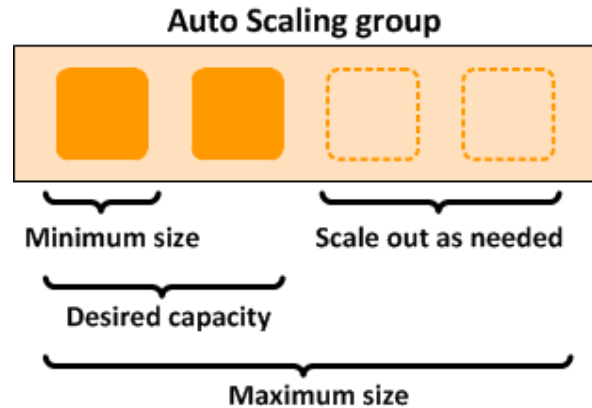
1. From vCenter, click **Home** and then click **AWS Management Portal**.
2. Expand the region for the instance, select the environment, and then click the **VM-to-EC2 Migrations** tab.
3. Find the entry with the import task ID or queued task ID that you noted earlier. The ID of the instance is shown in the **Instance ID** field.

7. To start the EC2 instance after the migration has completed, expand the environment, expand **Imported Instances**, select the instance, and then click the **Summary** tab. The ID of the instance should be the instance ID that you noted from the **VM-to-EC2 Migrations** tab. On the **Getting Started** tab, click **Start instance**.

VIII. WHAT IS AMAZON EC2 AUTO SCALING?

Amazon EC2 Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application. You create collections of EC2 instances, called Auto Scaling groups. You can specify the minimum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never

goes below this size. we can specify the maximum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes above this size. If you specify the desired capacity, either when you create the group or at any time thereafter, Amazon EC2 Auto Scaling ensures that your group has this many instance. If you specify scaling policies, then Amazon EC2 Auto Scaling can launch or terminate instances as demand on your application increases or decreases.



IX. WHAT IS AWS COMMAND LINE INTERFACE?

The AWS Command Line Interface (AWS CLI) is an open source tool that enables you to interact with AWS services using commands in your command-line shell. With minimal configuration, you can start using functionality equivalent to that provided by the browser-based AWS Management Console from the command prompt in your favourite terminal program:

- **Linux shells** – Use common shell programs such as bash, zsh, and tsch to run commands in Linux, macOS, or Unix.
- **Windows command line** – On Windows, run commands in PowerShell or at the Windows command prompt.
- **Remotely** – Run commands on Amazon Elastic Compute Cloud (Amazon EC2) instances through a remote terminal such as PuTTY or SSH, or with AWS Systems Manager.

All IaaS (infrastructure as a service) AWS administration, management, and access functions in the AWS Management Console are available in the AWS API and CLI. New AWS IaaS features and services provide full AWS Management Console functionality through the API and CLI at launch or within 180 days of launch.

Example Upload a file to Amazon S3

aws s3 cp provides a shell-like copy command, and automatically performs a multipart upload to transfer large files quickly and resiliently.

```
$ aws s3 cp myvideo.mp4 s3://mybucket/
```

Performing the same task with the low-level commands (available under aws s3api) would take a lot more effort.

X. NETWORK & SECURITY

Amazon EC2 provides the following network and security features.

Features

- **Amazon EC2 Key Pairs:** Amazon EC2 uses public-key cryptography to encrypt and decrypt login information. Public-key cryptography uses a public key to encrypt a piece of data, such as a password, then the recipient uses the private key to decrypt the data. The public and private keys are known as a *key pair*.
To log in to your instance, you must create a key pair, specify the name of the key pair when you launch the instance, and provide the private key when you connect to the instance. On a Linux instance, the public key content is placed in an entry within `~/.ssh/authorized_keys`. This is done at boot time and enables you to securely access your instance using the private key instead of a password.
- **Amazon EC2 Security Groups for Linux Instances:** A security group acts as a virtual firewall that controls the traffic for one or more instances. When you launch an instance, you can specify one or more security groups; otherwise, we use the default security group. You can add rules to each security group that allow traffic to or from its associated instances. You can modify the rules for a security group at any time; the new rules are automatically applied to all instances that are associated with the security group. When we decide whether to allow traffic to reach an instance, we evaluate all the rules from all the security groups that are associated with the instance
- **Controlling Access to Amazon EC2 Resources:** Your security credentials identify you to services in AWS and grant you unlimited use of your AWS resources, such as your Amazon EC2 resources. You can use features of Amazon EC2 and AWS Identity and Access Management (IAM) to allow other users, services, and applications to use your Amazon EC2 resources without sharing your security credentials. You can use IAM to control how other users use resources in your AWS account, and you can use security groups to control access to your Amazon EC2 instances. You can choose to allow full use or limited use of your Amazon EC2 resources.
- **Amazon EC2 Instance IP Addressing:** Amazon EC2 and Amazon VPC support both the IPv4 and IPv6 addressing protocols. By default, Amazon EC2 and Amazon VPC use the IPv4 addressing protocol; you can't disable this behaviour. When you create a VPC, you must specify an IPv4 CIDR block (a range of private IPv4 addresses). You can optionally assign an IPv6 CIDR block to your VPC and subnets and assign IPv6 addresses from that block to instances in your subnet. IPv6 addresses are reachable over the Internet.

- **Elastic IP Addresses:** An Elastic IP address is a static IPv4 address designed for dynamic cloud computing. An Elastic IP address is associated with your AWS account. With an Elastic IP address, you can mask the failure of an instance or software by rapidly remapping the address to another instance in your account.
An Elastic IP address is a public IPv4 address, which is reachable from the internet. If your instance does not have a public IPv4 address, you can associate an Elastic IP address with your instance to enable communication with the internet; for example, to connect to your instance from your local computer. Amazon EC2 currently do not support Elastic IP addresses for IPv6.
- **Enhanced Networking on Linux:** Enhanced networking uses single root I/O virtualization (SR-IOV) to provide high-performance networking capabilities on supported `instance_types`. SR-IOV is a method of device virtualization that provides higher I/O performance and lower CPU utilization when compared to traditional virtualized network interfaces. Enhanced networking provides higher bandwidth, higher packet per second (PPS) performance, and consistently lower inter-instance latencies. There is no additional charge for using enhanced networking.
- **Placement Groups:** You can launch or start instances in a *placement group*, which determines how instances are placed on underlying hardware. When you create a placement group, you specify one of the following strategies for the group:
 - **Cluster** – clusters instances into a low-latency group in a single Availability Zone
 - **Partition** – spreads instances across logical partitions, ensuring that instances in one partition do not share underlying hardware with instances in other partitions
 - **Spread** – spreads instances across underlying hardware
 - There is no charge for creating a placement group.

XI. CREATE YOUR AMAZON EC2 INSTANCE:

Step 1: Create Your EC2 Resources and Launch Your EC2 Instance

Before you can launch and connect to an Amazon EC2 instance, you need to create a key pair, unless you already have one. You can create a key pair using the Amazon EC2 console and then you can launch your EC2 instance.

Note: Using Amazon EFS with Microsoft Windows Amazon EC2 instances is not supported.

To create a key pair

- Follow these steps to create a key pair. If you already have a key pair, you do not need to create a new one

and you can use your existing key pair for this exercise.

To launch the EC2 instance

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
2. Choose **Launch Instance**.
3. In **Step 1: Choose an Amazon Machine Image (AMI)**, find an Amazon Linux AMI at the top of the list and choose **Select**.
4. In **Step 2: Choose an Instance Type**, choose **Next: Configure Instance Details**.
5. In **Step 3: Configure Instance Details**, choose **Network**, and then choose the entry for your default VPC. It should look something like vpc-xxxxxxx (172.31.0.0/16) (default).
 1. Choose **Subnet**, and then choose a subnet in any Availability Zone.
 2. Choose **Next: Add Storage**.
6. Choose **Next: Tag Instance**.
7. Name your instance and choose **Next: Configure Security Group**.
8. In **Step 6: Configure Security Group**, review the contents of this page, ensure that **Assign a security group** is set to **Create a new security group**, and verify that the inbound rule being created has the following default values.
 - **Type:** SSH
 - **Protocol:** TCP
 - **Port Range:** 22
 - **Source:** Anywhere 0.0.0.0/0

Step 6: Configure Security Group

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. [Learn more](#) about Amazon EC2 security groups.

Assign a security group: ☒ Create a new security group ☐ Select an existing security group

Security group name:

Description:

Type	Protocol	Port Range	Source
SSH	TCP	22	Anywhere 0.0.0.0/0

Add Rule

Warning
Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

[Cancel](#) [Previous](#) [Review and Launch](#)

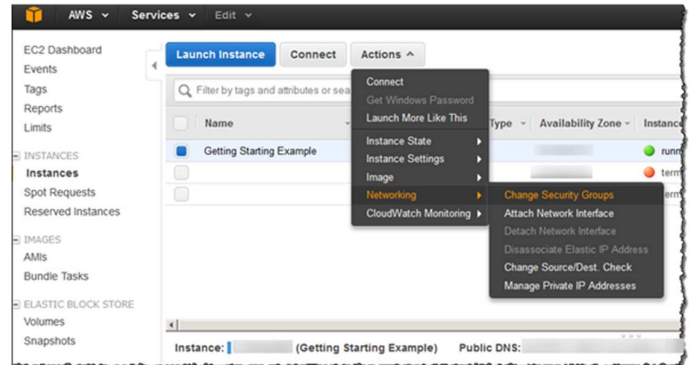
Note: You can configure the EFS file system to mount on your EC2 instance automatically.

9. Choose **Review and Launch**.
10. Choose **Launch**.
11. Select the check box for the key pair that you created, and then choose **Launch Instances**.

12. Choose **View Instances**.

13. Choose the name of the instance you just created from the list, and then choose **Actions**.

0. From the menu that opens, choose **Networking** and then choose **Change Security Groups**.



1. Select the check box next to the security group with the description **default VPC security group**.
2. Choose **Assign Security Groups**.

Change Security Groups

Instance ID:

Interface ID:

Select Security Group(s) to associate with your instance

Security Group ID	Name	Description
<input checked="" type="checkbox"/>	default	default VPC security group
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		

[Cancel](#) [Assign Security Groups](#)

Note: In this step, you assign your VPC's default security group to the Amazon EC2 instance. Doing this ensures that the instance is a member of the security group that the Amazon EFS file system mount target authorizes for connection in Step 2: Create Your Amazon EFS File System.

14. Choose your instance from the list.
15. On the **Description** tab, make sure that you have two entries listed next to **security groups**—one for the default VPC security group and one for the security group that you created when you launched the instance.

REFERENCE:

- **AWS Documentation:**
https://docs.aws.amazon.com/#lang/en_us
- **Getting started with AWS:**
<https://aws.amazon.com/documentation/gettingstarted/>
- **Amazon Web Service:**
<https://www.youtube.com/user/AmazonWebServices>
- **Wikipedia:**
https://en.wikipedia.org/wiki/Amazon_Web_Services
- **BOOK:** Getting Started with AWS Kindle
Edition by **Amazon Web Services**