Amazon Elastic Compute Cloud (EC2) is a web service that provides resizable compute capacity in the cloud. It allows users to run virtual servers (instances) on-demand, enabling scalable computing for a variety of applications, from web hosting to big data processing.

Key Features

- **Scalability**: Easily scale your capacity up or down as your computing requirements change.
- **Variety of Instance Types**: Choose from a wide range of instance types optimized for different use cases, including compute-optimized, memory-optimized, storage-optimized, and GPU instances.
- **Pay-As-You-Go Pricing**: Only pay for the compute capacity you use, with options for on-demand, reserved, or spot instances.
- **Integration with AWS Services**: Seamlessly integrate with other AWS services, such as Amazon S3, RDS, and VPC.

EC2 Instance Types

EC2 instances are categorized into several families, each designed for specific workloads:

- **General Purpose**: Balanced compute, memory, and networking resources (e.g., T3, M5).
- **Compute Optimized**: High-performance processors for compute-intensive tasks (e.g., C5).
- **Memory Optimized**: Optimized for memory-intensive applications (e.g., R5, X1).
- **Storage Optimized**: Designed for high storage throughput and low latency (e.g., I3, D2).
- **Accelerated Computing**: Includes GPU instances for high-performance computing tasks (e.g., P3, G4).

- **High I/O**: Ideal for applications that require high input/output operations per second (IOPS), such as large databases or data-intensive applications (e.g., I3en).

Creating an EC2 Instance: Step-by-Step Guide

Step 1: Sign In to AWS

- Log in to your AWS Management Console at AWS Management Console using your AWS account credentials.
- From the AWS Management Console, locate the **Services** dropdown and select **EC2** under the **Compute** section.

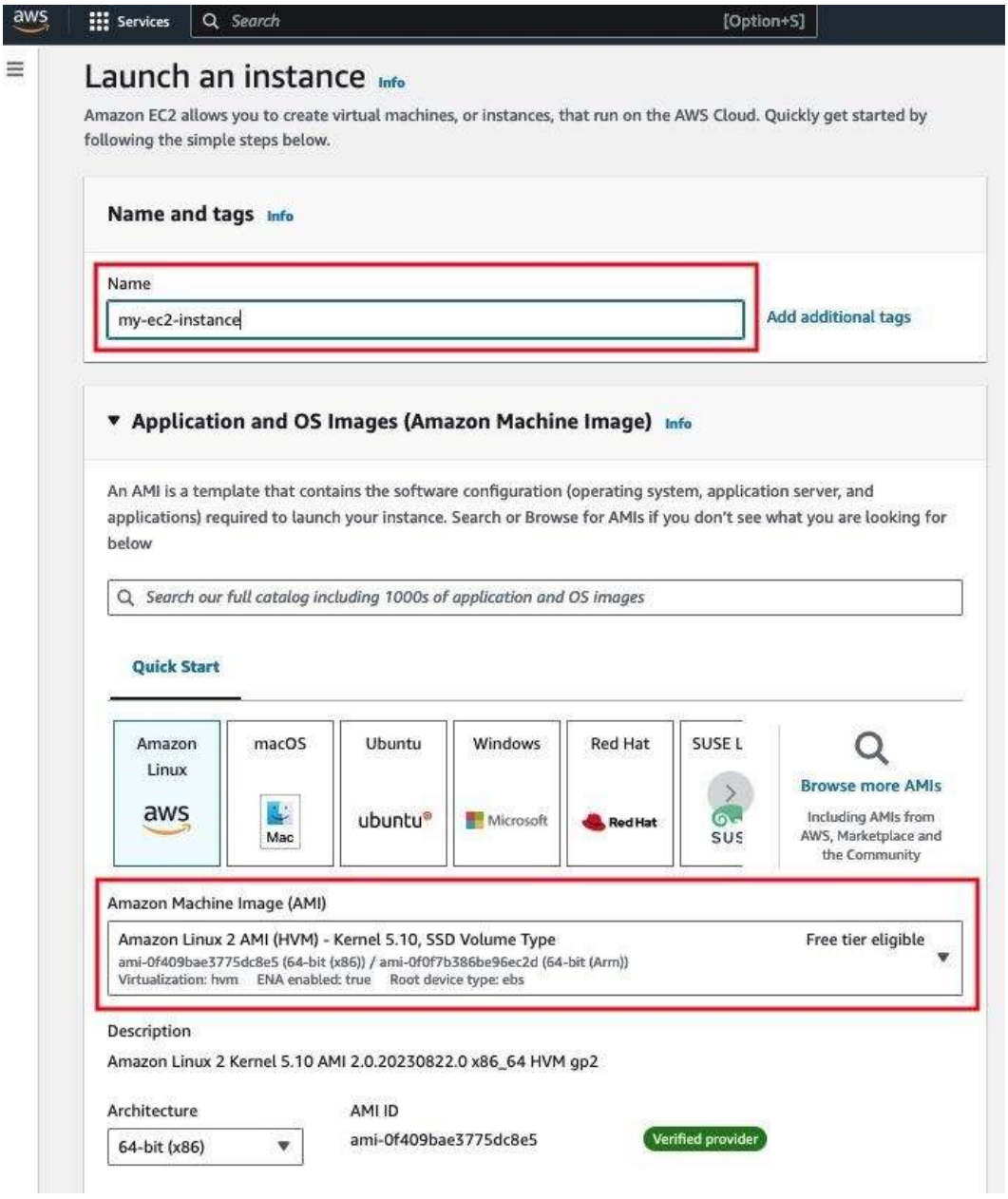
Step 2: Launch an Instance

- In the EC2 Dashboard, click the **Instances** link in the left navigation pane, then click the orange **Launch Instance** button.

Step 3: Choose an Amazon Machine Image (AMI)

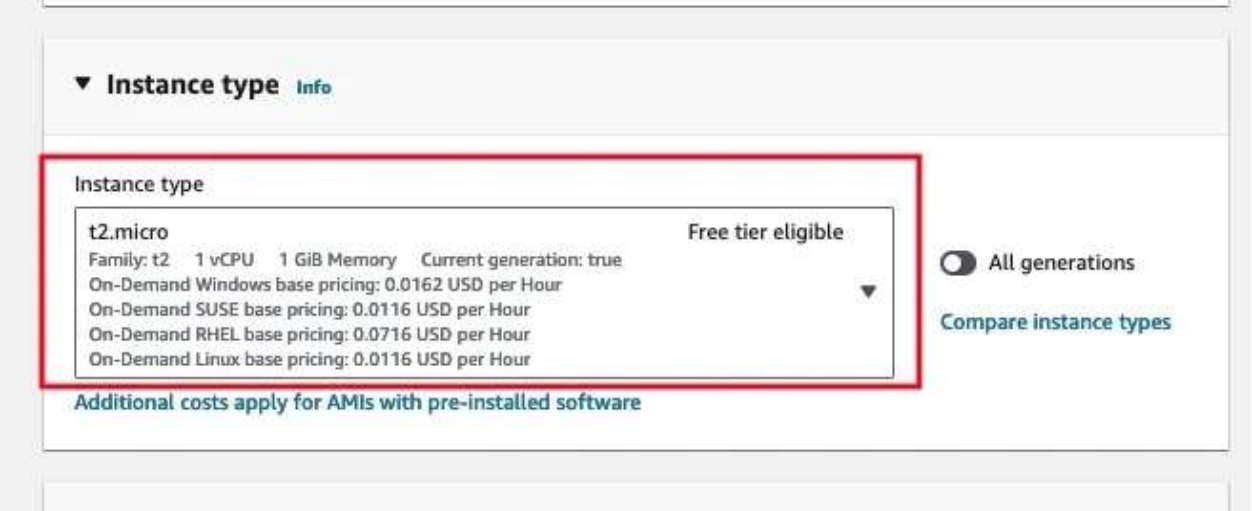
- Choose an AMI that matches your requirements.

- AMIs are pre-configured templates that include an operating system and other software. Most servers use Linux because it's open-source, reliable, and efficient, commonly used for web servers, databases, and more. For this guide, we'll choose a Linux AMI.



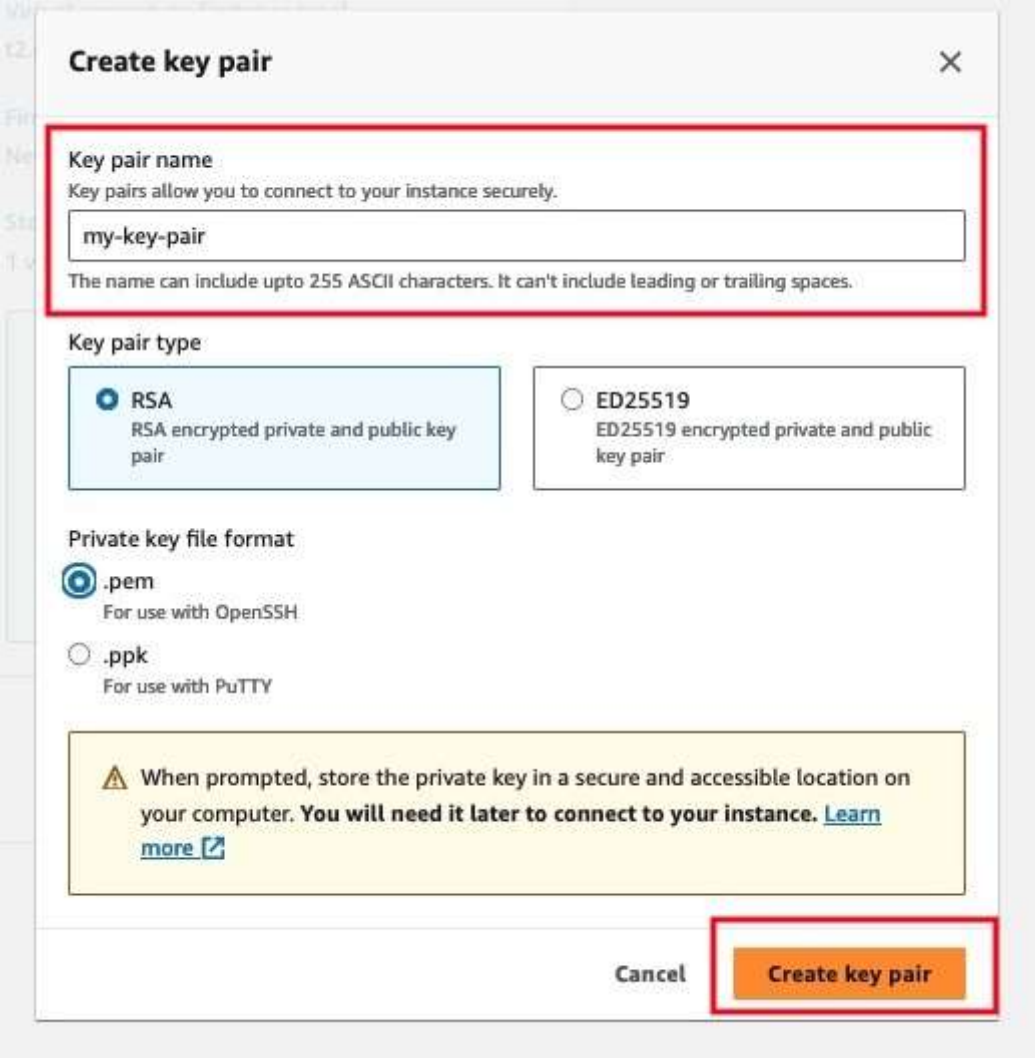
Step 4: Choose an Instance Type

- Select the instance type based on your computing needs. Consider factors like CPU, memory, storage, and network capacity.



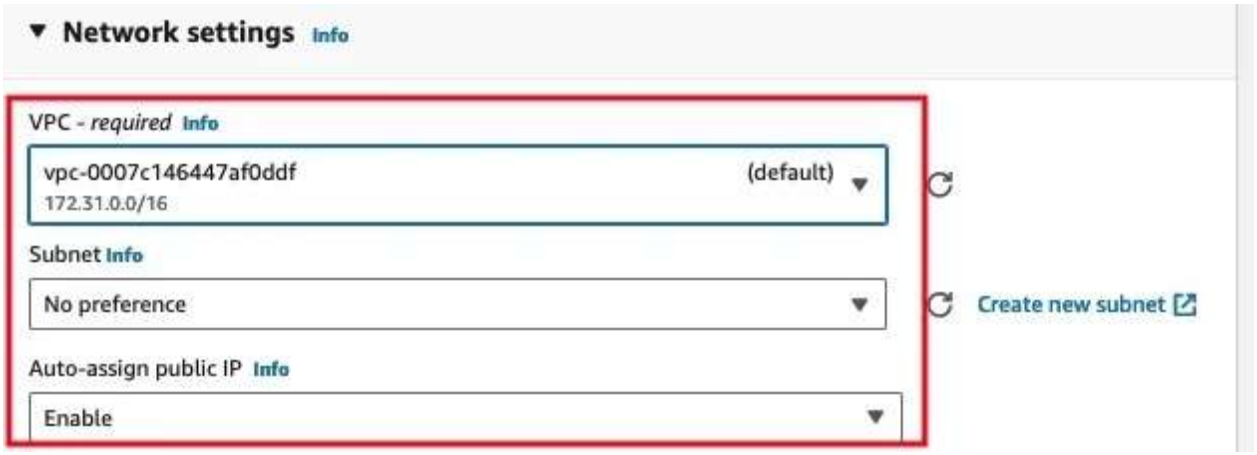
Step 5: Create or Select a Key Pair

- You need a key pair to securely connect to your instance using SSH. Create a new key pair or select an existing one.
- Make sure to safely store the private key file, as it's the key to accessing your server. If you lose it, you won't be able to access the instance. You can add other key pairs later if needed.
- Select **Create a new key pair**, give a name to your .pem key, and click the **Create key pair** orange button.
- A .pem key file will be downloaded to your local computer.



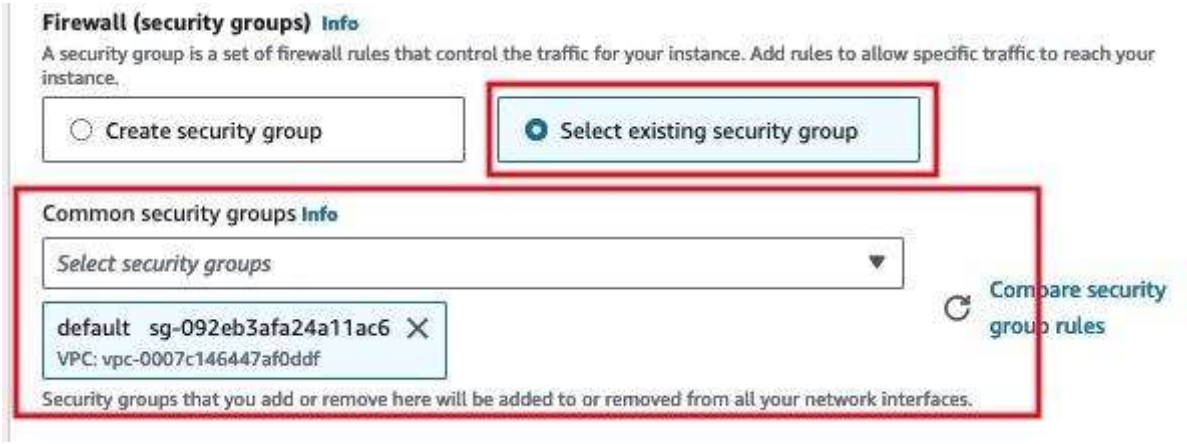
Step 6: Configure Network Details

- Choose the network settings (like the VPC and subnet). For now, we'll use the default VPC and subnet. Think of VPC as a private network, and we'll dive deeper into it later.



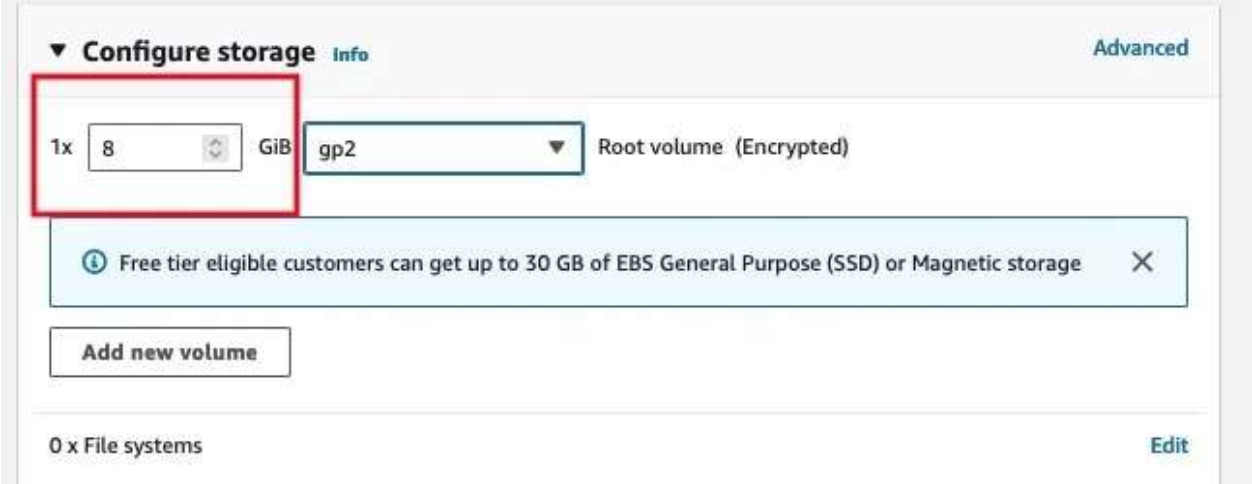
Step 7: Configure Security Group

- Security groups act as virtual firewalls for your instance. Define inbound and outbound rules to control network traffic. For now, use the default security group. We can delve deeper into security groups later.



Step 8: Add Storage

- Configure the amount and type of storage for your instance. You can add additional storage volumes if needed.



Step 9: Review and Launch

- Review your instance configuration settings. If everything looks good, click the **Launch** button.

Step 10: Launch Status

- Once you click **Launch**, your instance will start launching. You'll see its status change to **running** in the EC2 Dashboard.
- As you delve deeper into the EC2 service, you will gradually grasp and gain insights into additional settings, expanding your understanding and knowledge over time.

Accessing Your Instance

Access Your Instance using SSH

- Once your instance is running, go to the EC2 Dashboard, select your instance, and click the **Connect** button. You'll find instructions for both Linux and Windows connections below:

Connecting from Mac and Linux:

- Open a terminal on your local machine.
- Navigate to the directory where you saved your private key file (.pem).
- Use the following command to set the appropriate permissions on the key file:

```
Bash Code :
        chmod 400 <your-key-file>.pem
```

- Copy the SSH command from the EC2 instance connect page.
- Paste the command into your terminal and press Enter.

Connecting from Windows:

- Download and install an SSH client such as **PuTTY**.
- Convert your .pem key file to a .ppk key file using **PuTTYgen**.
- Open **PuTTY** and enter the public IP address of your instance in the "Host Name" field.
- Load your .ppk key file in the "Connection > SSH > Auth" settings.
- Click **Open** to start the SSH session.

Access Your Instance using AWS Session Manager

- You can access your instance through the AWS console using **Session Manager**. This provides a browser-based session similar to EC2 Instance Connect, eliminating the need to open port 22.