EC2

Dr Peadar Grant

September 21, 2022

Infrastructural services consist of compute, low-level storage and the networking required to connect them. Compute is provided on AWS using a service called EC2. Virtual Machines on EC2 are called Instances.

Instances are grouped into a virtual network called a VPC. VPCs are important — without them we wouldn't be able to connect to our virtual machines at all.

1 Compute instances

The backbone IaaS compute service offered by AWS is Elastic Compute 2. This allows you to create Virtual Machines on the AWS platform. Many of their other services rely on EC2. Other major cloud platforms (Azure, IBM, Google) all offer a basic virtual-machine compute service. Virtual Machines created on AWS are known as EC2 instances (or just instances).

1.1 EC2 instances

To launch an instance, we must decide:

Instance Type which sets the hardware configuration. Includes CPU family, numberof cores, clock speed, RAM.

Operating System template as available as an Amazon Machine Image (or AMI) that will be used to clone the image from.

Storage: The instance's virtual hard disk.

VPC and Subnet that the instance will be launched into.

Security group which defines what traffic is allowed in/out of individual EC2 instances. Try to use shared/template ones rather than one per instance.

1.2 Amazon Machine Image (AMI)

- On physical servers (and when dealing with on-site virtualisation systems) we often install an OS using an installer disk and work through the steps of the installer.
- Cloud compute instances are normally created by cloning an image and running some minor post-installation tasks.

- AWS uses the Amazon Machine Image (AMI) to clone a compute instance's virtual hard disk from.

1.3 Amazon Linux

There are a few different Linux distributions on AWS as AMIs, (as well as Windows). One possible Linux flavour is Amazon Linux which is stripped-down for use as a cloud server and is maintained by AWS:

- By default, a user named `ec2-user` is provisioned. You can of course set up other users as you wish as on any OS.
- It uses the RedHat `yum` package manager rather than `apt` as in Ubuntu / Debian. *(You should be comfortable at this stage working with similar non-identical tools to broaden your experience and horizon!)*

1.4 Remote Access

Unlike some virtualisation solutions (e.g. Xen server, Hyper-V), AWS provide no emulated Keyboard / Video / Mouse (KVM). All operations must be carried out using remote access protocols like Secure Shell (SSH) or Remote Desktop Protocol.

1.5 Key pairs

Some AMIs have a default username/password, but most including Amazon Linux use a key-pair:

1. You generate a public / private key pair using `ssh-keygen`
2. You use the public key when creating EC2 instances.
3. Log in using `ssh` using your private key. Username for Amazon Linux is `ec2-user`.

2 Security groups

Security groups control a per-instance firewall that limits traffic into or out of each instance. Each instance may have one or more security groups attached.

2.1 Default security group

Every instance created can have a default security group attached butthis leads to a few problems:

- Hard to get an overview of allowed/denied traffic to instances (security risk)
- Hard to reconfigure allowed/denied traffic to a number of instances (time consuming, nuisance)

Instead it is preferable to create a security group and attach it as needed to instances in our VPC.

2.2 Ingress rules

Ingress rules allow traffic in on a particular port (e.g. TCP 22 for SSH). We often use the IP range 0.0.0.0/0 as the source, meaning from anywhere on the internet. We can lock this down to specific IP addresses or IP ranges (e.g. your ISP).

2.3 Egress rules

By default, security groups allow egress of all traffic from instances, so this doesn't need to be set up.

3 Virtual private clouds

IaaS components (like EC2) are launched in Virtual Private Clouds (VPCs).

3.1 Virtual Private Cloud (VPC)

A VPC (Virtual Private Cloud) is a virtual software-defined network to which your compute instances connect. Essentially it's a software-defined multi-site data entre environment. Each VPC . . .

- . . . is associated with a single specific region.
- . . . is owned by a single AWS account.
- . . . has a CIDR block of addresses, such as 10.0.0.0/16.

3.1.1 IP addresses

Classless Inter Domain Routing is a way to specify a range of IP addresses in a standard format. An IPv4 address is a 32-bit identifier conventionally considered as 4 blocks of 8-bits, Figure 1

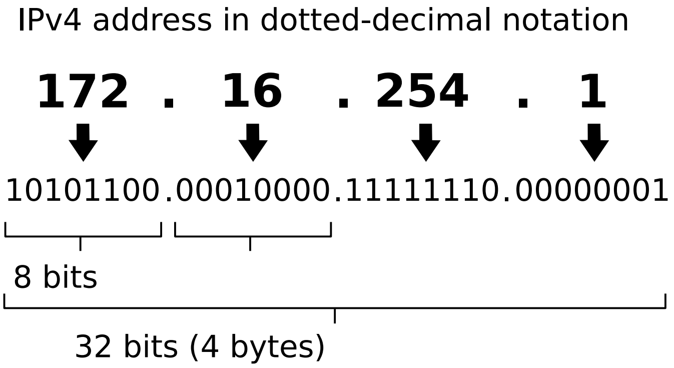
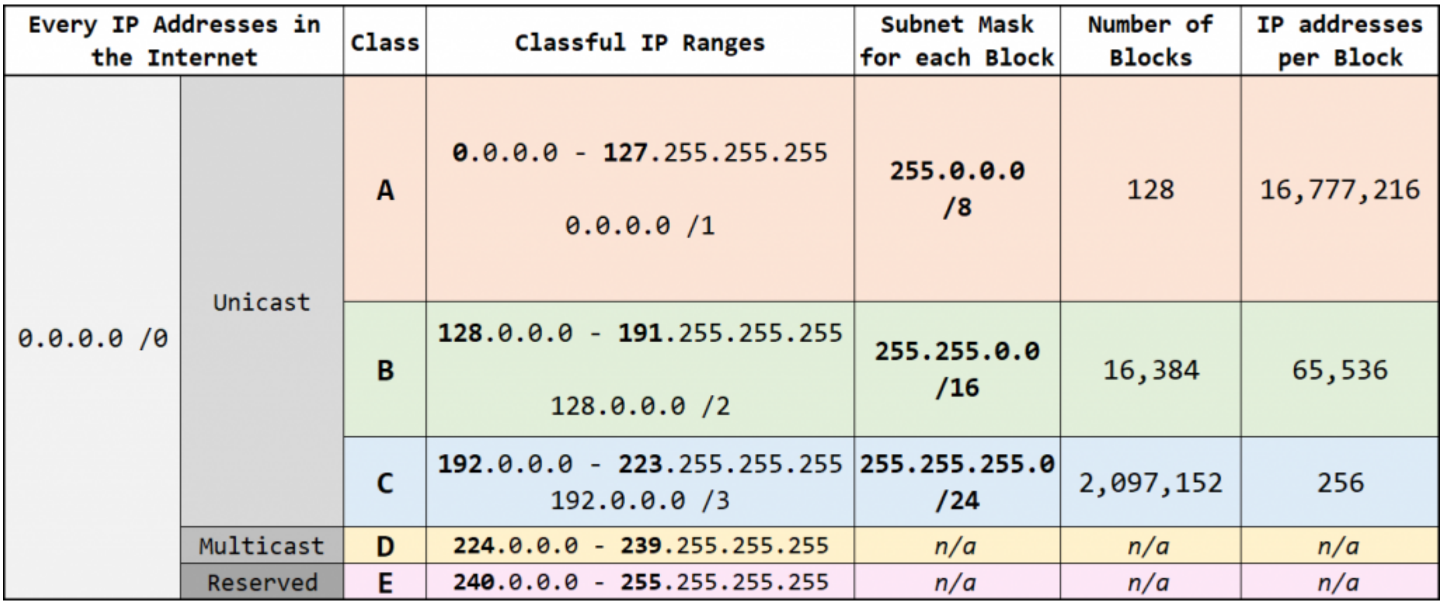


Figure 1: IPv4 address

3.1.2 CIDR blocks

Common CIDR ranges are shown Figure 2.



Every IP Addresses in the Internet		Class	Classful IP Ranges	Subnet Mask for each Block	Number of Blocks	IP addresses per Block
0.0.0.0 /0	Unicast	A	0.0.0.0 - 127.255.255.255 0.0.0.0 /1	255.0.0.0 /8	128	16,777,216
		B	128.0.0.0 - 191.255.255.255 128.0.0.0 /2	255.255.0.0 /16	16,384	65,536
		C	192.0.0.0 - 223.255.255.255 192.0.0.0 /3	255.255.255.0 /24	2,097,152	256
	Multicast	D	224.0.0.0 - 239.255.255.255	n/a	n/a	n/a
	Reserved	E	240.0.0.0 - 255.255.255.255	n/a	n/a	n/a

Figure 2: Common CIDR ranges

3.2 Subnets

A VPC will normally contain one or more subnets. Each subnet:

- is associated with a single specific Availability Zone.
- has a single CIDR-block of addresses, such as 10.0.1.0/24.
- may contain one or more EC2 instances. Each EC2 instance therefore belongs to a particular VPC.

3.3 Internet Gateway

When a VPC (and its subnets) are created, it's actually isolated from the internet. An Internet Gateway connects a VPC to the public internet.

There are some circumstances we'll see later on where we actually don't want a VPC to connect to the internet, but generally we do. Therefore, except in some unusual configurations, every VPC will have an internet gateway attached to it.

3.4 Route table(s)

Route tables control how traffic is directed among the different subnets and into/out of the VPC. They're not difficult to understand, but a misconfigured route table will generally cause problems. Every VPC has a default route table that normally specifies that:

- Traffic for IP ranges within the VPC is local.
- Traffic for IP ranges anywhere (0.0.0.0/0) must go through the internet gateway.

4 Names and IDs

Names are important to identify components. We already know that AWS use text code names for the different regions and AZs. As we create and use resources it's important to note the distinction between names we assign and Ids that are assigned by AWS:

Names are assigned by you when creating components like VPCs, subnets, internet gateways, ec2 instances.

Ids are assigned by AWS when you create the same resources.

You can name things any way you like, but I tend to suggest you follow a pattern and avoid spaces! Table 1 shows suggested suffix patterns I use (with the corresponding AWS prefixes for the ids)

Component type	Suggested name suffix	AWS name prefix
VPC	`_VPC`	`vpc-`
Subnet	`_SN`	`sn-`
Internet gateway	`_IGW`	`igw-`
Route table	`_RTB`	`rtb-`

Table 1: Naming suggestions and AWS Id prefixes