NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

SkillsFuture Career Transition Program

Cloud Infrastructure Engineering - AWS Compute Instances

Nanyang Technological University
Skills Union

2022/2023

Elastic Compute Cloud (EC2)

- EC2 is one of the most popular AWS IaaS offerings.
- Main capabilities include:
  - Renting VMs from Cloud provider (EC2)
  - Storing data on virtual drives (EBS)
  - Distributing load across machines (ELB)
  - Scaling services using an auto-scaling group (ASG)

- One of the most important and basic services to understand

EC2 Considerations

- Operating System → Linux? Windows? MacOS? Others?
- RAM allocation?
- CPU allocation?
- Storage space needed?
  - Network-attached (EBS & EFS)?
  - Additional hardware (EC2 Instance Store)?
- Network card - Speed of card? Public IP address needed? Static IP?
- Firewall Rules via Security Groups?
- Bootstrap script (configured at first launch)?

EC2 User Data

- It is possible to bootstrap our instances using an EC2 User data script.
- **bootstrapping** means launching commands when a machine starts
- That script is only run once at the instance first start
- EC2 user data is used to automate boot tasks such as:
  - Installing updates
  - Installing software
  - Downloading common files from the internet
  - Anything you can think of
- The EC2 User Data Script runs with the root user

EC2 User Data

User data - *optional* **Info**

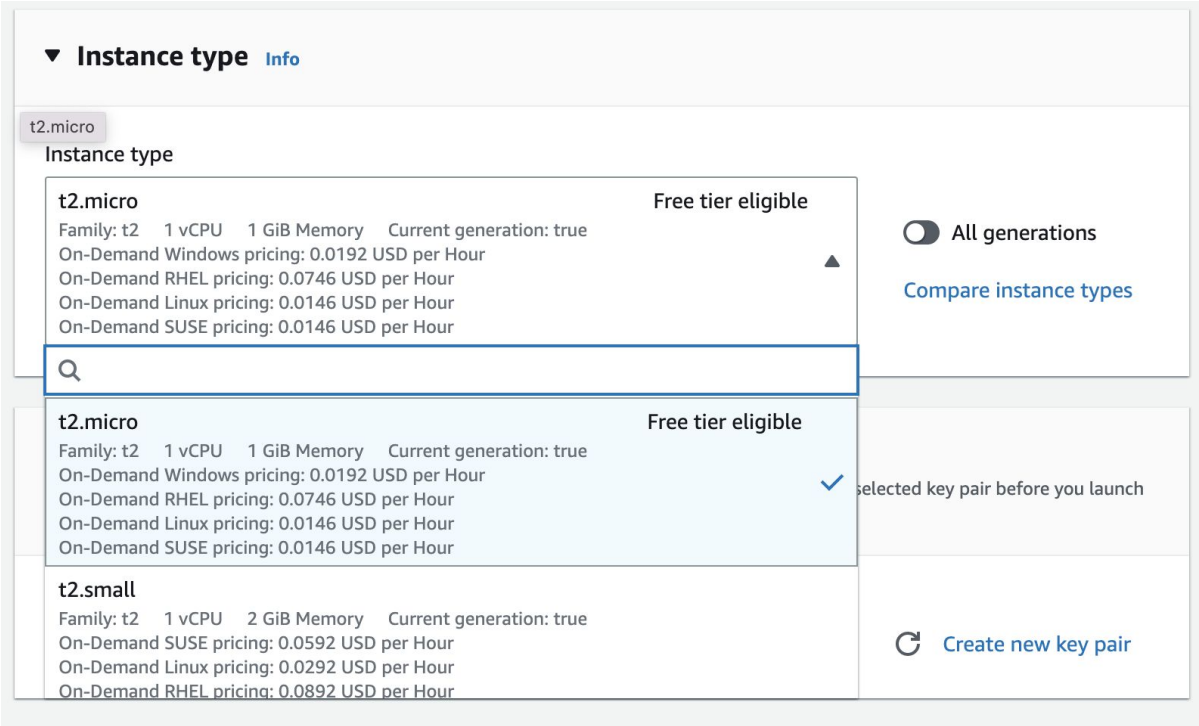
Enter user data in the field.

```
yum install -y
yum install docker -y
```

EC2 Instance Types

- You can use different types of EC2 instances that are optimised for different use cases (https://aws.amazon.com/ec2/instance-types/)
- Types of Instances:
  - General Purpose
  - Compute Optimized
  - Memory Optimized
  - Accelerated Computing
  - Storage Optimized
  - HPC Optimized

EC2 Instance Types



EC2 Instance Types

- For example, let's look at the below instance naming convention:

m5.8xlarge

- m - Instance class
- 5 - Generation
- 8xlarge - Size within the instance class

EC2 Instance Types - GP

- Great for a diversity of workloads such as web servers or code repositories
- Balance between:
  - Compute
  - Memory
  - Networking
- We will be using the t2.micro a lot, which is a General Purpose EC2 instance, for our lessons

EC2 Instance Types - Compute

- Great for compute-intensive tasks that require high performance processors:
  - Batch processing workloads
  - Media transcoding
  - High performance web servers
  - High performance computing (HPC)
  - Scientific modeling & machine learning
  - Dedicated gaming servers

EC2 Instance Types - Memory

- Fast performance for workloads that process large data sets in memory
- Use cases:
  - High performance, relational/non-relational databases
  - Distributed web scale cache stores
  - In-memory databases optimized for BI (business intelligence)
  - Applications performing real-time processing of big unstructured data

EC2 Instance Types - Storage

- Great for storage-intensive tasks that require high, sequential read and write access to large data sets on local storage
- Use cases:
  - High frequency online transaction processing (OLTP) systems
  - Relational & NoSQL databases
  - Cache for in-memory databases (for example, Redis)
  - Data warehousing applications
  - Distributed file systems

EC2 Instance Types

It’s great that t2.micro is part of our free tier (Up to 750 hours per month)

Instance	vCPU*	CPU Credits / hour	Mem (GiB)	Storage	Network Performance
t2.nano	1	3	0.5	EBS-Only	Low
t2.micro	1	6	1	EBS-Only	Low to Moderate
t2.small	1	12	2	EBS-Only	Low to Moderate
t2.medium	2	24	4	EBS-Only	Low to Moderate
t2.large	2	36	8	EBS-Only	Low to Moderate
t2.xlarge	4	54	16	EBS-Only	Moderate
t2.2xlarge	8	81	32	EBS-Only	Moderate

All instances have the following specs:

- Intel AVX†, Intel Turbo†
- t2.nano, t2.micro, t2.small, t2.medium have up to 3.3 GHz Intel Xeon Scalable processor
- t2.large, t2.xlarge, and t2.2xlarge have up to 3.0 GHz Intel Scalable Processor

EC2 Security Groups

This portion will be covered in 2.6

EC2 Instances Savings Plans

- **On-Demand Instances**: short workload, predictable pricing
- Reserved: (MINIMUM 1 year)
  - **Reserved Instances**: long workloads
  - **Convertible Reserved Instances**: long workloads with flexible instances
  - **Scheduled Reserved Instances**: example – everyThursday between 3 and 6 pm
- **Spot Instances**: short workloads, cheap, can lose instances (less reliable)
- **Dedicated Hosts**: book an entire physical server, control instance placement
- **Dedicated Instances**: no other customers will share your hardware

EC2 - On Demand

- Pay for what you use
  - Linux - billing per second, after the 1st minute.
  - All other OS - billing per hour
- Highest Flexibility - you can start and stop and terminate the instance anytime
- No long-term commitment

Recommended for users who need short-term & uninterrupted servers and those who can’t predict how the application will behave.

EC2 - Reserved Instances

- Up to 72% discount compared to On-Demand
- Reserve for 1 year or 3 years (more discount)
- Purchasing options: 0 or partial upfront & Full upfront (more savings)
- Reserve a specific instance type

Recommended for users who have steady-state usage applications (DB servers)

EC2 - Reserved Instances

- **Convertible Reserved Instance**
  - Able to change the EC2 instance type
  - Save up to **45% discount**
- **Scheduled Reserved Instances**
  - launch within time window you reserve
  - Require a fraction of day / week / month
  - Commitment for 1 year only

EC2 - Spot Instances

- Save up to **90% compared to On-Demand**
- However, these instances **can be stopped and terminated at any point** if your max price is less than the current spot price
- Most **cost-efficient instance** in AWS

EC2 - Spot Instances

- Useful for workloads that are **resilient to failure**
  - Batch jobs
  - Data analysis
  - Image processing
  - Any distributed workloads
  - Workloads with a flexible start and end time
- • Not suitable for critical jobs or databases

EC2 - Dedicated Hosts

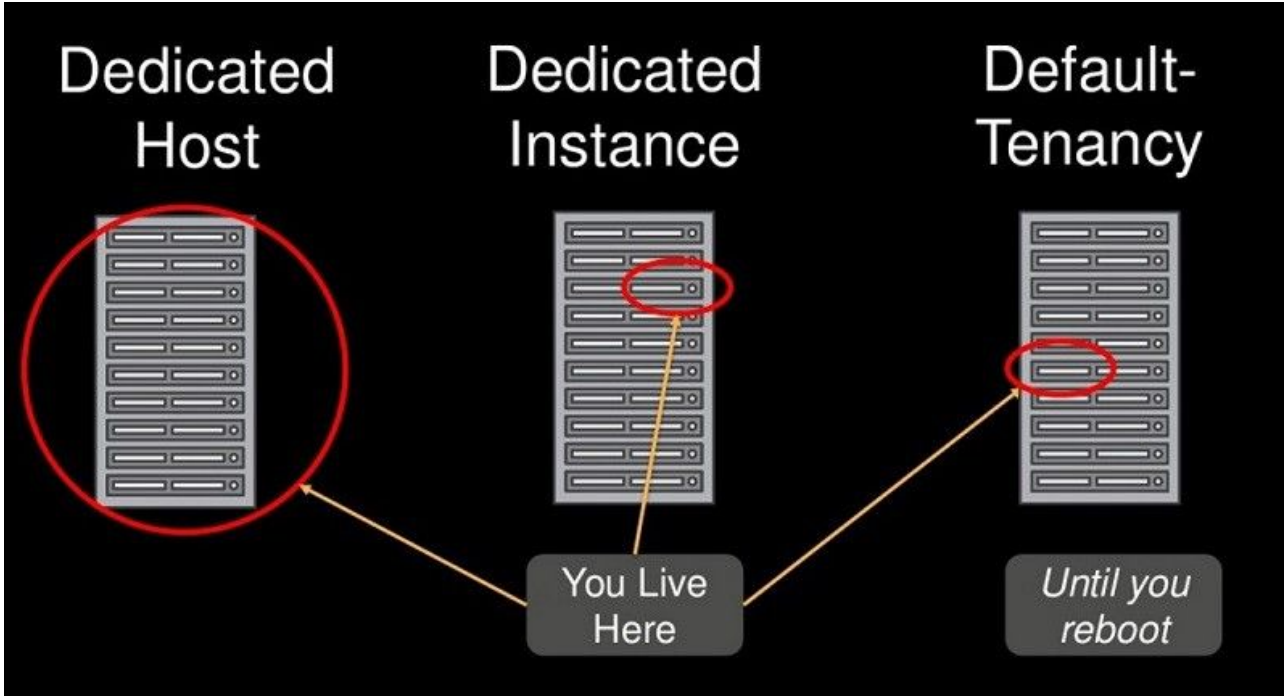
- An Amazon EC2 Dedicated Host is a physical server with EC2 instance capacity fully dedicated to your use. Dedicated Hosts can help you **address compliance requirements** and **reduce costs** by allowing you to use your existing server-bound software licenses.
- Allocated for your account for a **3-year period reservation**
- **More expensive**

- Useful for software that have complicated licensing model (BYOL – Bring Your Own License) or for companies that have strong regulatory or compliance needs

EC2 - Dedicated Instances

- Instances running on hardware that's dedicated to you
- May share hardware with other instances in same account
- No control over instance placement (can move hardware after Stop / Start)

EC2 - Dedicated Tenancy



EC2 - Price Comparison

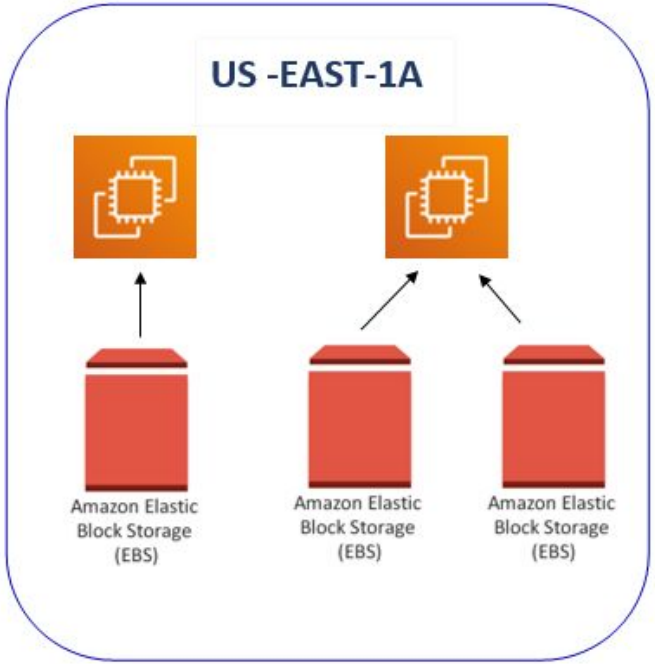
EC2 Type	Price Per Hour ($)
On-Demand	0.1
Spot Instance	0.032 - 0.045
Spot Block (1 to 6 hours)	0.032 - 0.045
Reserved Instance (12 months) - no upfront	0.062
Reserved Instance (12 months) - full payment	0.058
Reserved Convertible Instance (12 months) - no upfront	0.071
Reserved Scheduled Instance (12 months)	0.09 - 0.095

EBS Volume

- A network drive you can attach to your instances while they run
- It allows your instances to **persist data**, even after their termination
- They can only be mounted to one instance at a time (at the CCP level)
- They are bound to a **specific availability zone**

- Analogy:Think of them as a “network USB stick”
- Free tier: 30 GB of free EBS storage of type General Purpose (SSD) or Magnetic per month

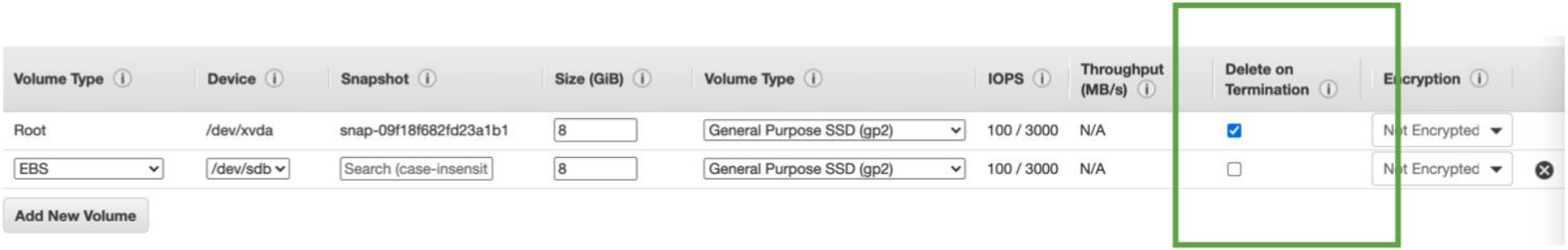
EBS Volume


US -EAST-1A
Amazon Elastic Block Storage (EBS)
Amazon Elastic Block Storage (EBS)
Amazon Elastic Block Storage (EBS)

EBS Delete On Termination

- Controls the EBS behaviour when an EC2 instance terminates
  - By default, the root EBS volume is deleted (attribute enabled)
  - By default, any other attached EBS volume is not deleted (attribute disabled)
- This can be controlled by the AWS console / AWS CLI
- Use case: preserve root volume when instance is terminated

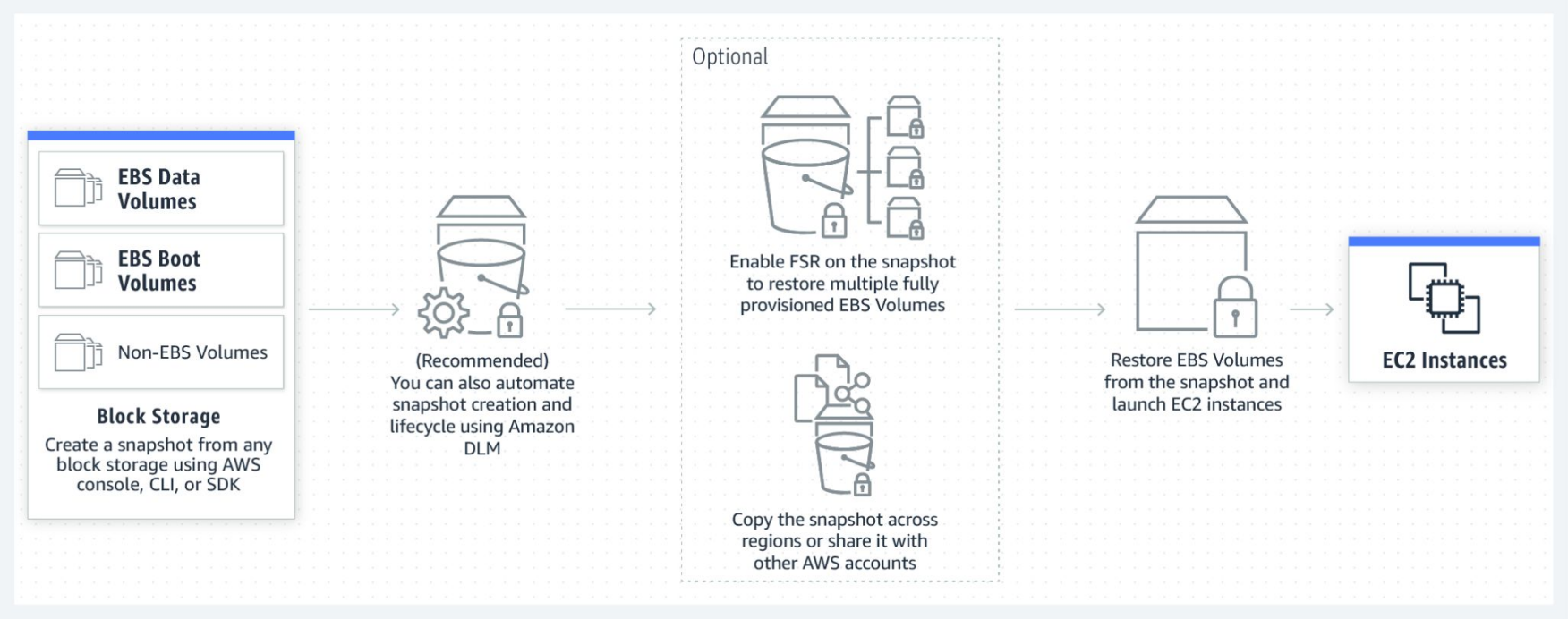
EBS Delete On Termination



EBS Copying & Snapshots

- Make a backup (snapshot) of your EBS volume at a point in time
- Not necessary to detach volume to do snapshot, but recommended
  - Just to make sure that no new data is added while snapshot is in progress
- Can copy snapshots across AZ or Region

EBS Copying & Snapshots



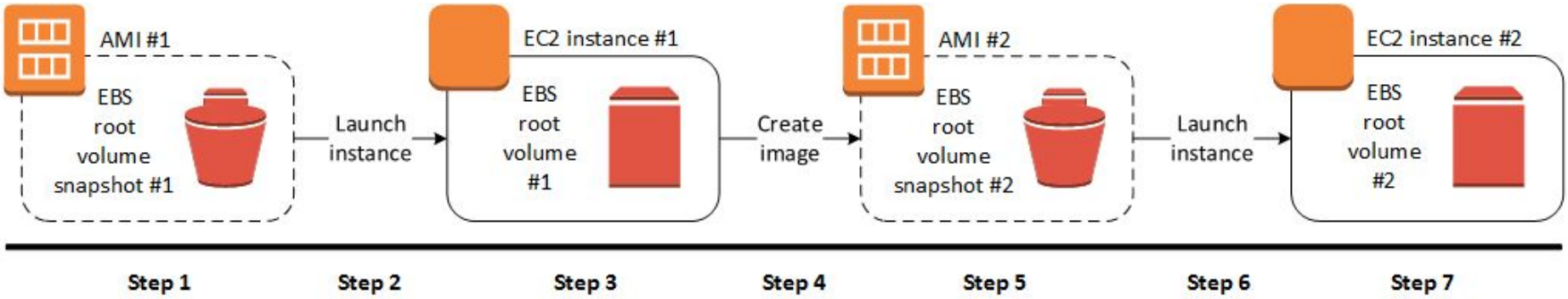
Amazon Machine Image (AMI)

- AMI are a customization of an EC2 instance
  - You add your own software, configuration, operating system, monitoring...
  - Faster boot / configuration time because all your software is pre-packaged
- AMI are built for a **specific region** (and can be copied across regions)
- You can launch EC2 instances from:
  - **A Public AMI:** AWS provided
  - **Your own AMI:** you make and maintain them yourself
  - **An AWS Marketplace AMI:** an AMI someone else made (and potentially sells)

Creating AMI from EC2

- Start an EC2 instance and customize it
- Stop the instance (for data integrity)
- Build an AMI – this will also create EBS snapshots
- Launch instances from other AMIs

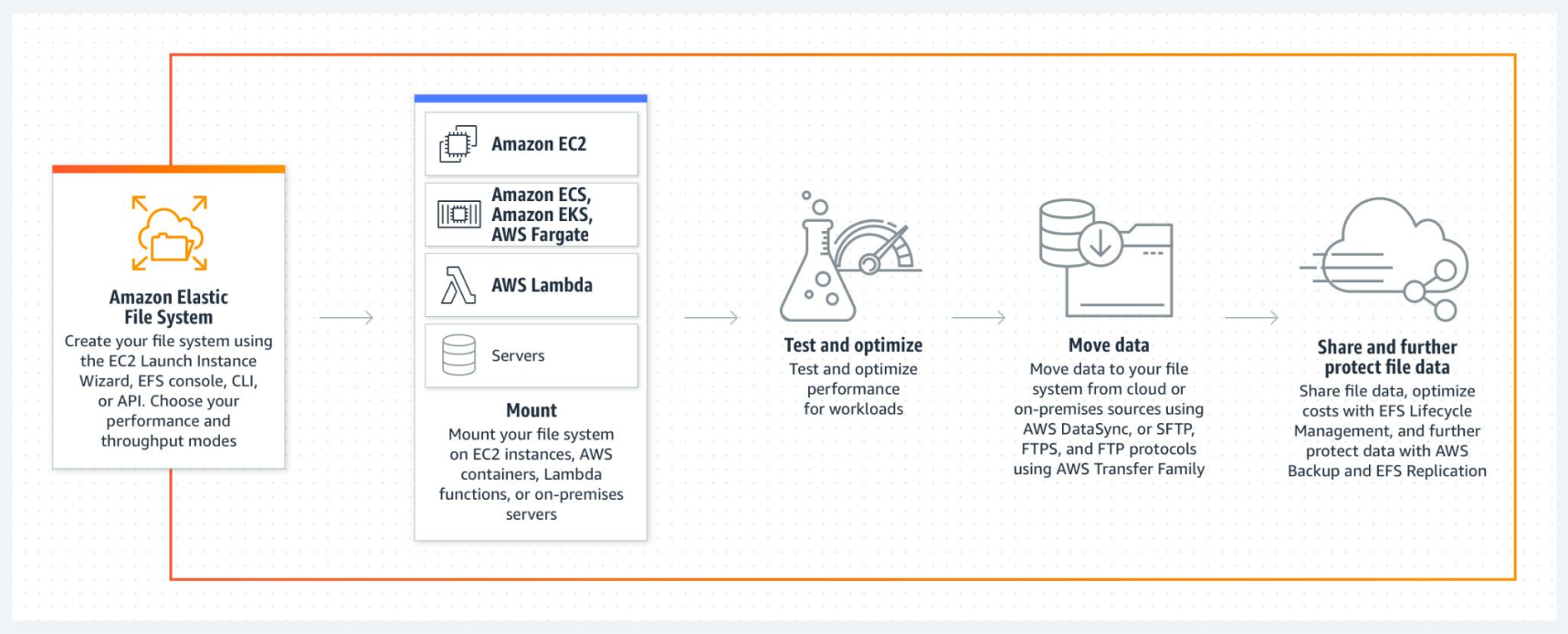
Creating AMI from EC2



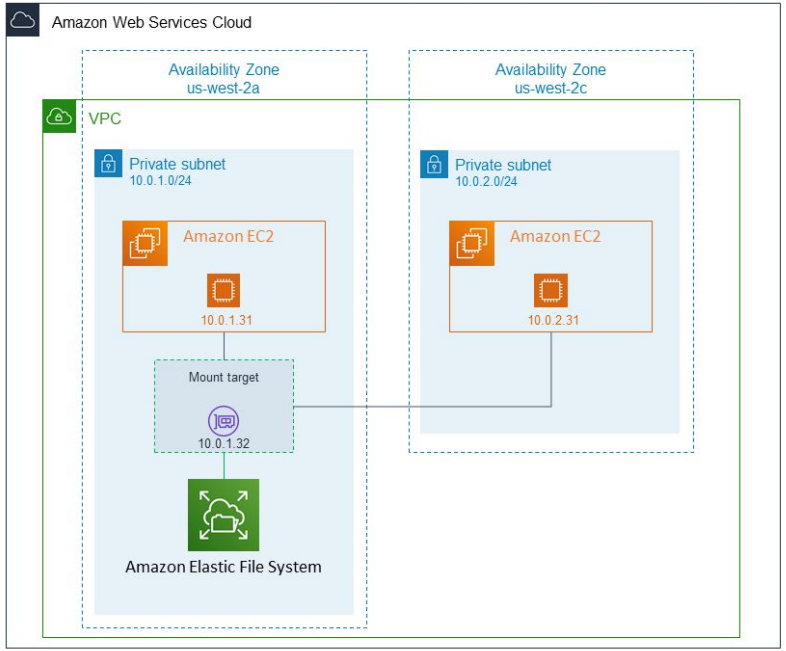
Elastic File System (EFS)

- Managed NFS (network file system) that can be mounted on 100s of EC2
- EFS works with Linux EC2 instances in multi-AZ
- Highly available, scalable, expensive (3x gp2), pay per use, no capacity planning

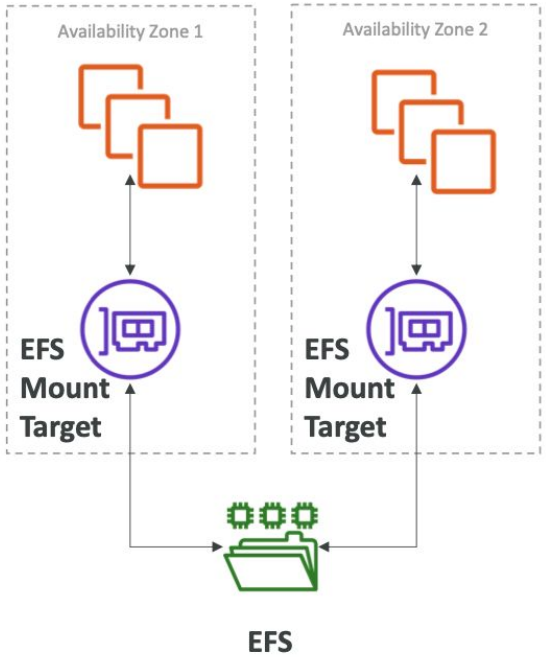
Elastic File System (EFS)

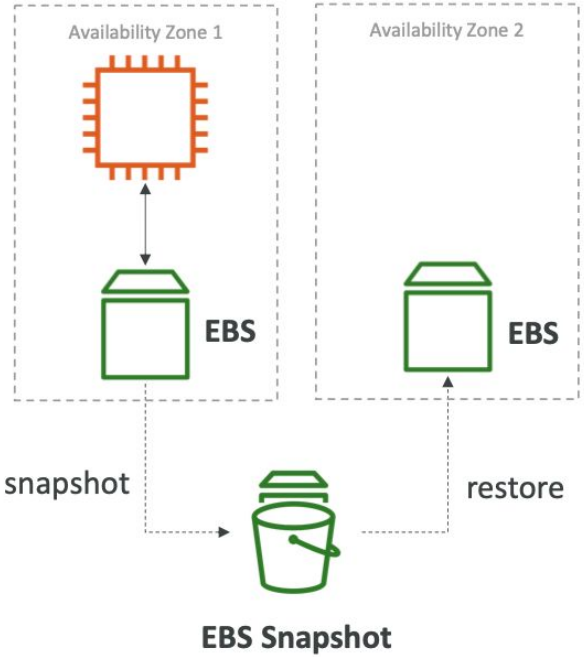


Elastic File System (EFS)



EFS vs EBS


Availability Zone 1
Availability Zone 2
EFS
Mount
Target
EFS
Mount
Target
EFS


Availability Zone 1
Availability Zone 2
EBS
EBS
snapshot
restore
EBS Snapshot

What’s Next?