aws SUMMIT

Optimizing Storage for Enterprise Workloads

David Stein

Storage Business Development , Amazon Web Services

Managing On-Premise Storage Arrays

Time

- Long 3-6-month acquisition
- 30-90 day implementation

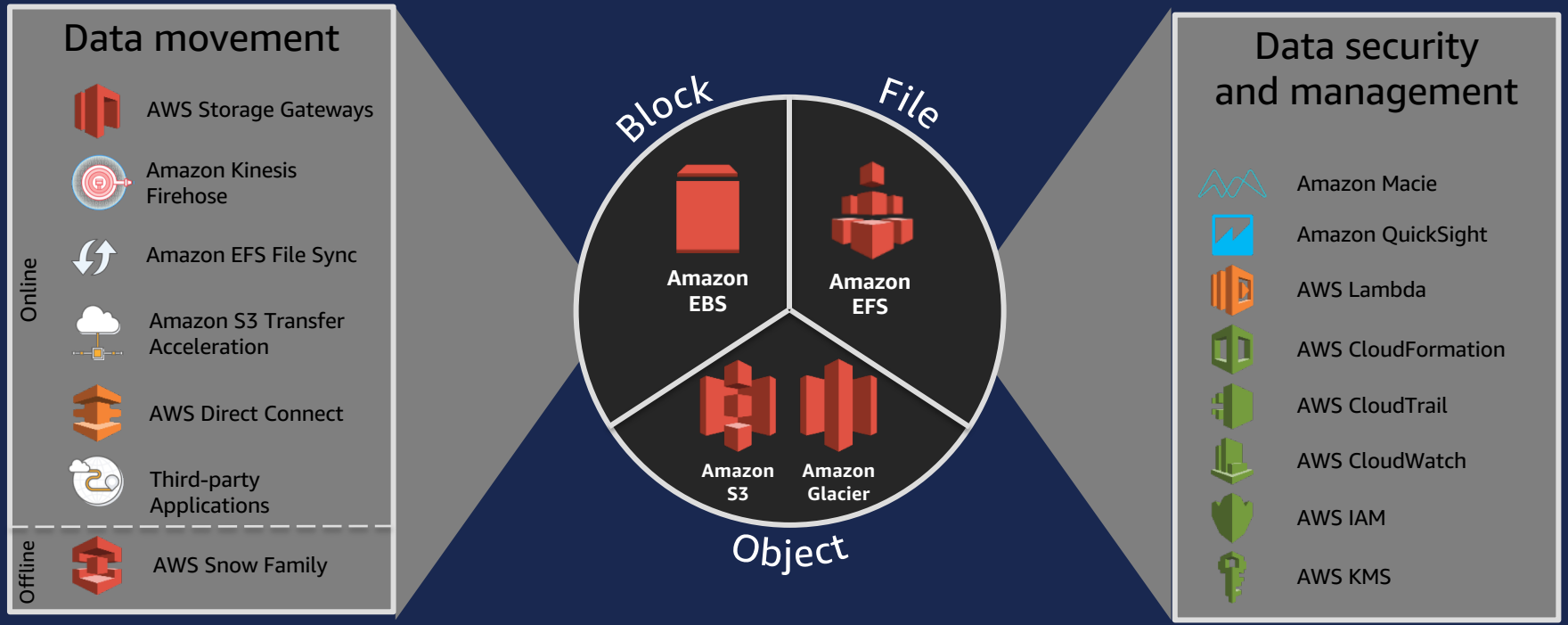
Costs

- Capex – up-front initial investment
- Guesswork on how much storage you will need, mix of SSD/HDD
- Expensive software for tiering and snapshotting

Operational

- Availability – need a second array and datacenter for HA
- Backup – secondary storage array plus offsite copy
- Scalability – difficult to scale: time/cost to order more drives, upgrade controllers, EOL of array

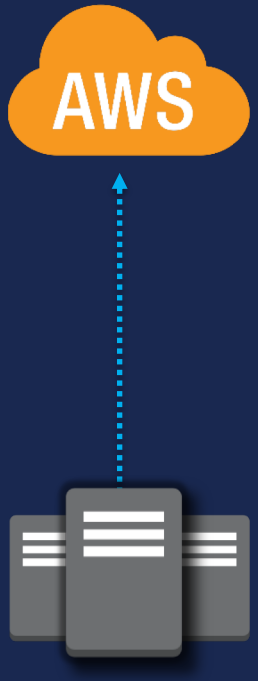
Complete set of storage building blocks



AWS storage customers

Data Migration

Moving to the Cloud



Common drivers of moving data to the cloud:

- Cost effectiveness
- Workforce productivity
- Business agility
- Operational resilience
- Access to innovative services

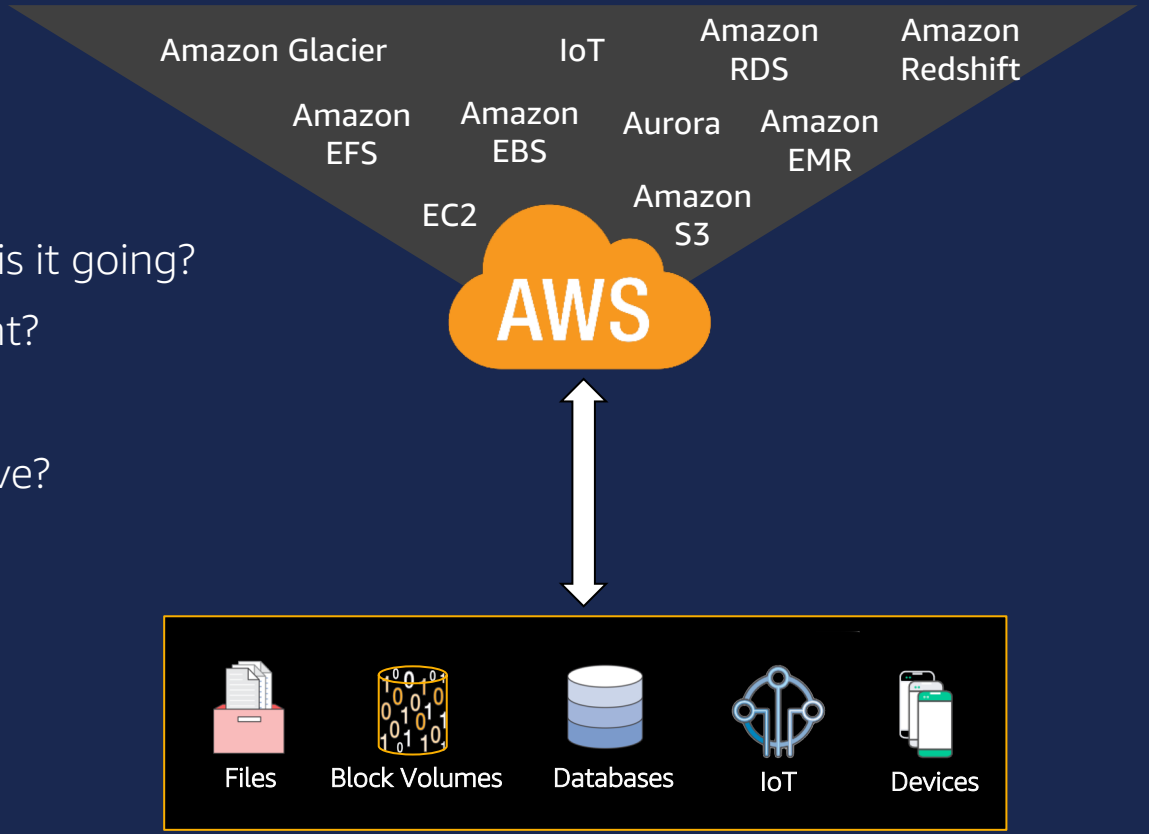
Pain points of data migration:

While every cloud adoption journey is unique, all require a data migration step, which can be hindered by blockers such as:

- Disruption
- Speed
- Costs
- Channels
- Cloud compatibility

Migrating Data: Five Key Questions

1) What kind of data is it, and where is it going?
2) One-time or continuous movement?
3) One-way or bidirectional access?
4) How much data & time do you have?
5) How might your WAN be a factor?



AWS Migration Capabilities

Discovery & Migration Tracking

AWS Migration Hub

AWS Application Discovery Service

Server & Database Migration

AWS Server Migration Service

AWS Database Migration Service

VMware on AWS

Data Migration Services

AWS Snow Family

(Snowball, Snowball Edge, Snowmobile)

Move terabytes to petabytes of data to AWS using appliances designed for secure, physical transport.

AWS Storage Gateway

(File, Volume, & Tape Gateways)

Connects on-premises applications to AWS. Uses common storage interfaces to migrate data.

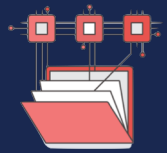
Streaming Data

Amazon Kinesis Firehose

Capture, process, & load streaming data into AWS to be used with business intelligence and analytics tools.

File

Amazon EFS File Sync



Sync files from a source file system into a destination Amazon EFS file system. The source file system can be on-premises or in the cloud

Partners

AWS Partner Network

Migration & Storage Competency Partners provide solutions and expertise to help businesses move to AWS.

Network Optimizations

AWS Direct Connect

Establishes private connectivity between AWS and your data center, office, or colocation environment.

Amazon S3 Transfer Acceleration

Makes Internet transfers to S3 faster. Ideal for recurring jobs such as uploads, backups, & local processing tasks

What are AWS DMS and AWS SCT?

AWS Database Migration Service (AWS DMS) easily and securely migrate and/or replicate your databases and data warehouses to AWS

AWS Schema Conversion Tool (AWS SCT) converts your commercial database and data warehouse schemas to open-source engines or AWS-native services, such as Amazon Aurora and Redshift

Over 64,000 databases migrated and counting …

When to use AWS DMS and AWS SCT?

Modernize

Modernize your database tier –

- Commercial to open-source
- Commercial to Amazon Aurora

Modernize your Data Warehouse –

- Commercial to Redshift

Migrate

- Migrate business-critical applications
- Migrate from Classic to VPC
- Migrate data warehouse to Redshift
- Upgrade to a minor version
- Consolidate shards into Aurora

Replicate

- Create cross-regions Read Replicas
- Run your analytics in the cloud
- Keep your dev/test and production environment sync

File Storage with Amazon Elastic File Storage (EFS)

What if you could…

IT administrators

- Eliminate file system management and maintenance, especially at scale

App owners and Developers

- Migrate existing code, applications, and tools into the AWS cloud
- Build new cloud-native applications on scalable file storage

Business Managers

- Predict pricing and eliminate up-front capital costs
- Increase agility
- Spend less time managing file storage and more time driving the business

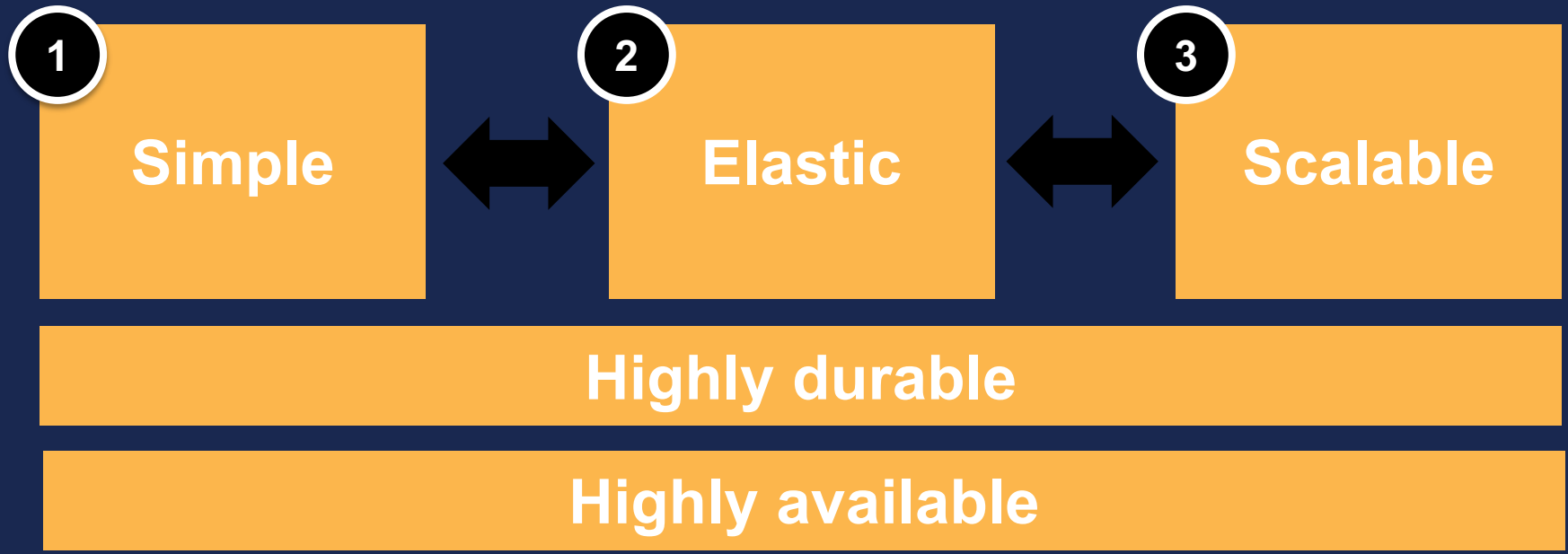
Amazon Elastic File System (EFS)

Provides simple, scalable, highly available & durable file storage in the cloud

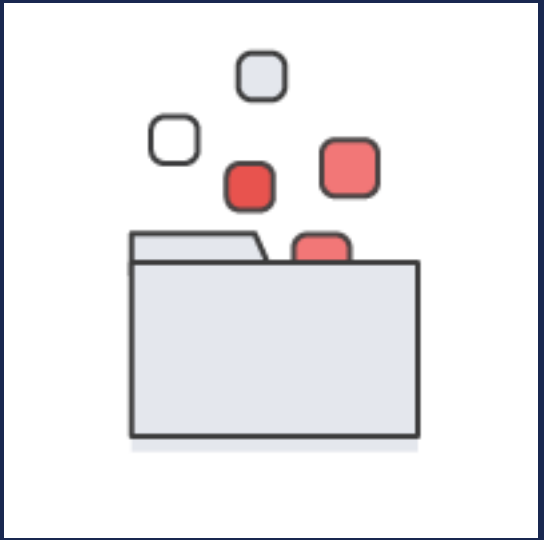
Petabyte scale file system distributed across an unconstrained number of storage servers in multiple Availability Zones (AZs)

Elastic capacity, automatically growing & shrinking as you add & remove files

We focused on changing the game



1 Amazon EFS is Simple



Fully managed

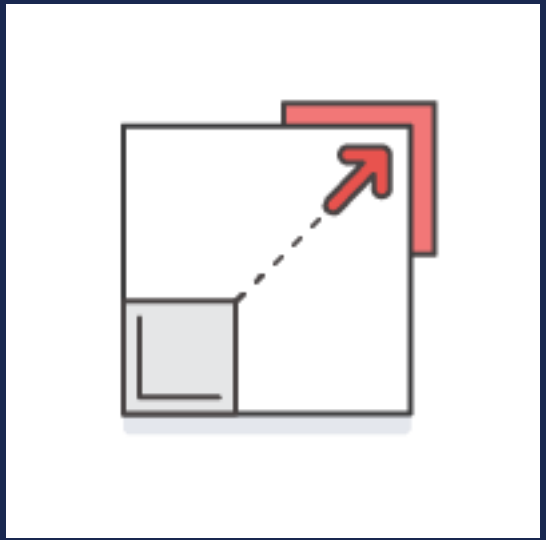
- No hardware, network, file layer
- Create a scalable file system in seconds!

Seamless integration with existing tools and apps

- NFS v4.1—widespread, open
- Standard file system access semantics
- Works with standard OS file system APIs

Simple pricing = simple forecasting

2 Amazon EFS is Elastic

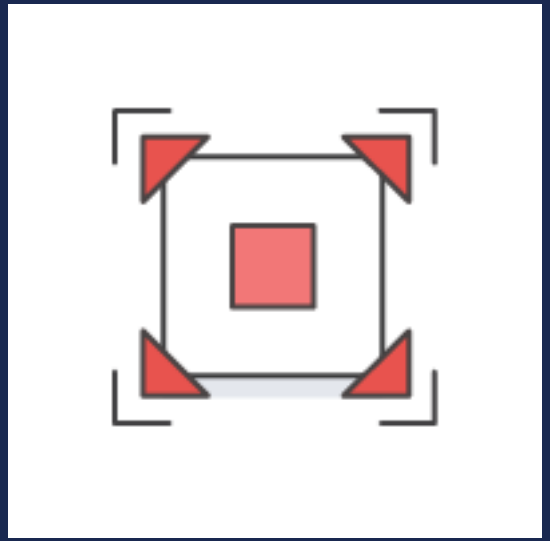


File systems grow and shrink automatically as you add and remove files

No need to provision storage capacity or performance

You pay only for the storage space you use, with no minimum fee

3 Amazon EFS is Scalable



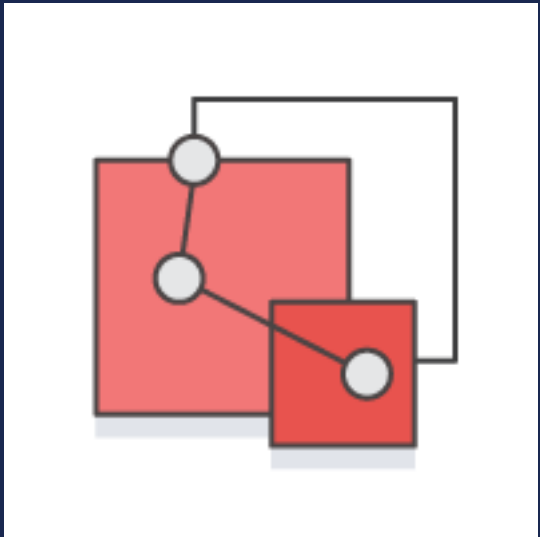
File systems can grow to petabyte scale

Throughput and IOPS scale automatically as file systems grow

Consistent low latencies regardless of file system size

Support for thousands of concurrent NFS connections

Highly Durable and Highly Available



Designed to sustain AZ offline conditions

Resources aggregated across multiple AZ's

Superior to traditional NAS availability models

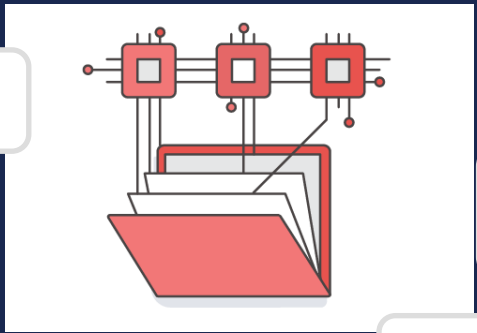
Appropriate for Production / Tier 0 applications

What customers are using EFS for today

Web serving

Content management

Database backups



Analytics

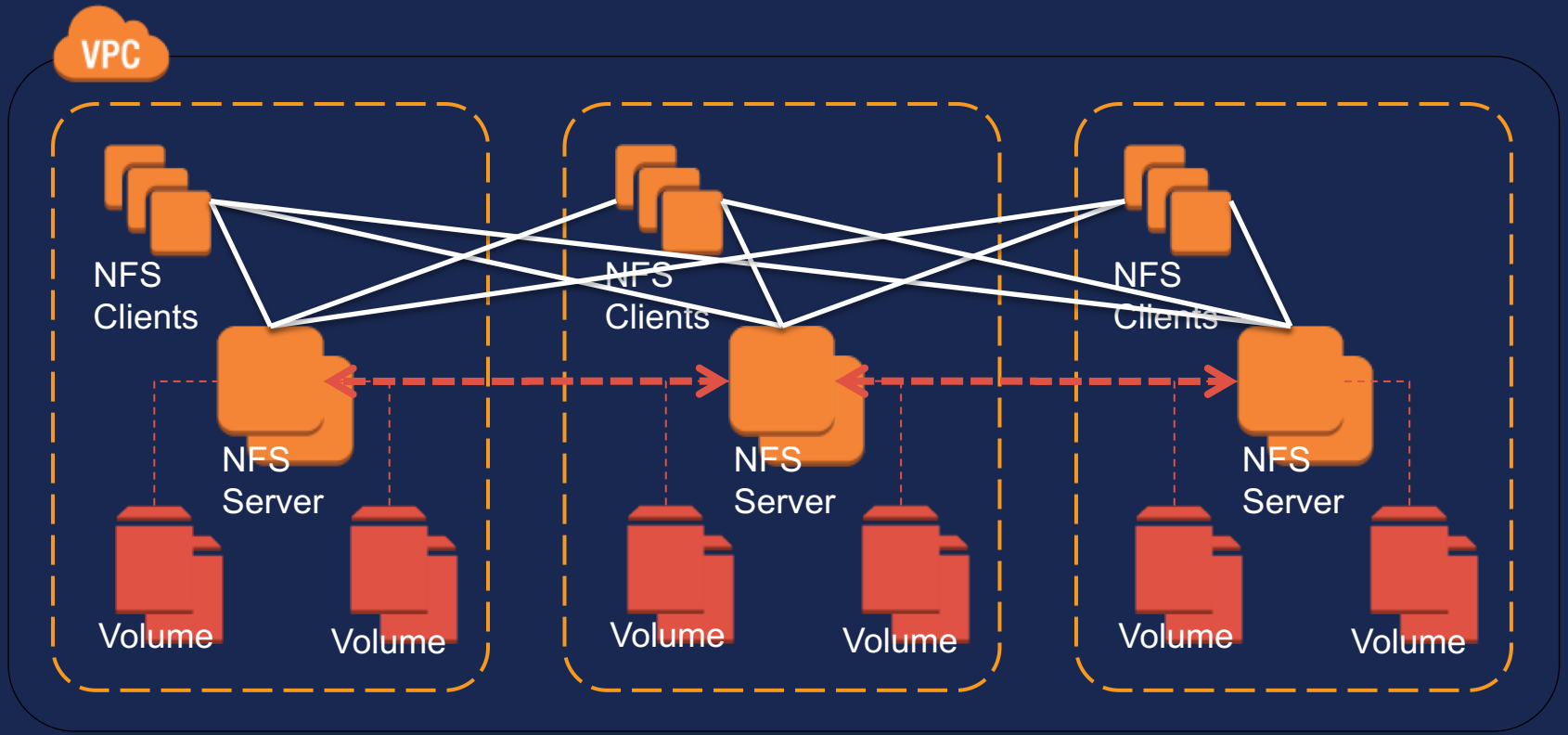
Media and Entertainment workflows

Container storage

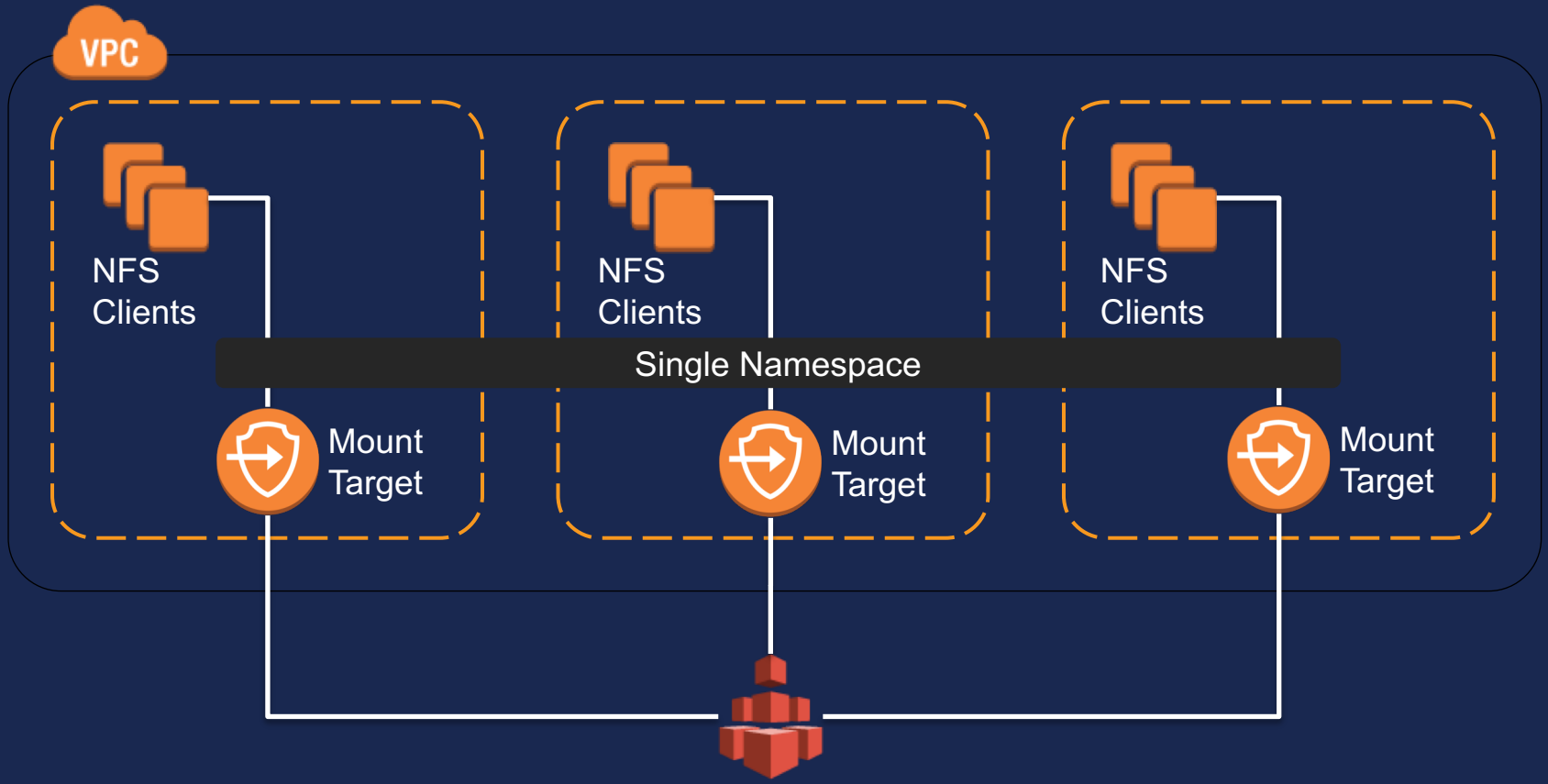
Workflow management

Home directories

Do It Yourself NFS Architecture



Amazon EFS Architecture



EFS TCO example

Let's say you need to store ~500 GB and require high availability and durability

Using a shared file layer on top of EBS, you might provision 600 GB (with ~85% utilization) and fully replicate the data to a second Availability Zone for availability/durability

Example comparative cost:

Storage (2x 600 GB EBS gp2 volumes):	**$120 per month**
Compute (2x m4.xlarge instances):	**$350 per month**
Inter-AZ data transfer costs (est.):	**$129 per month**
Total	**$599 per month**

EFS cost is (500GB * $0.30/GB-month) = **$150 per month**, with no additional charges

EFS File Sync

Sync data from existing file systems into Amazon EFS file systems

Simple
Set up and manage easily from the AWS Console

Fast
Up to 5x faster than standard Linux copy tools

Secure
Encrypted parallel data transfer to AWS

Use EFS File Sync to copy

- file systems from on-premises to EFS
- DIY in-cloud file systems to EFS
- EFS file systems between AWS Regions

EFS File Sync

Fast and simple way to copy file systems to Amazon EFS

- Sync Agent deployed as a VM
- No data is stored on the Sync Agent
- Optimized TCP stack for high-throughput
- Data is compressed and encrypted across the network

Block Storage

Enterprise Workloads on AWS

Mission Critical Databases

Amazon RDS

Oracle
SQL
MySQL
PostgreSQL

Amazon Aurora

MySQL
PostgreSQL

Data Warehouse

Amazon Redshift

Vertica
Teradata

Data Analytics

Amazon EMR

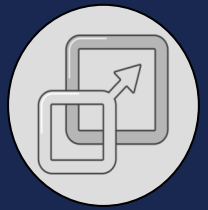
Cloudera
Hortonwork
sMapR

File

Amazon EFS

NFS

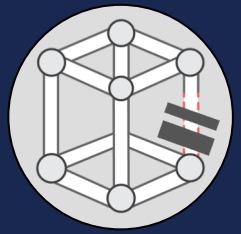
Amazon EBS – Let us do the undifferentiated heavy lifting



Transparent

Minimal re-architecting for applications

Adjust performance and price on the fly

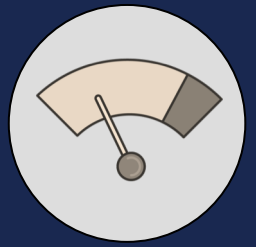


Reliable

Highly Available

Failure Tolerant

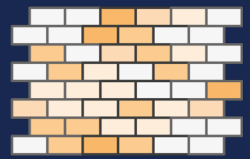
Foundational for enterprise applications



Performant

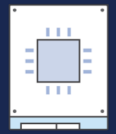
Consistently high IOPS and throughput

Optimized for low-latency workloads



AWS
block storage
offerings

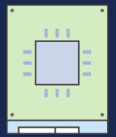
EC2
instance
store



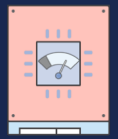
I2/I3

D2/H1

EBS
SSD-backed
volumes



gp2



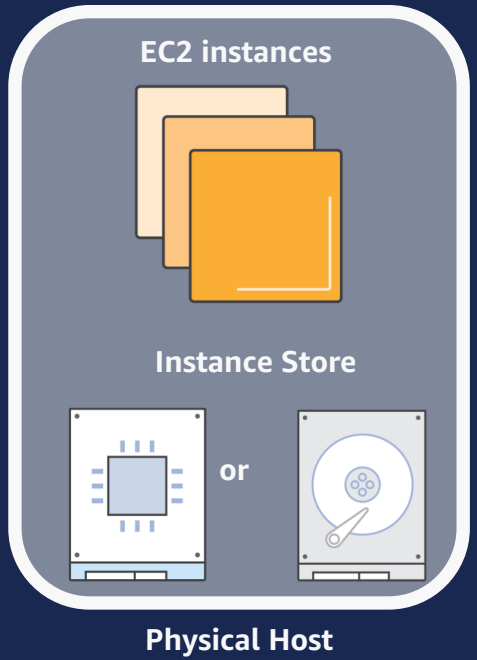
io1

EBS
HDD-backed
volumes

st1

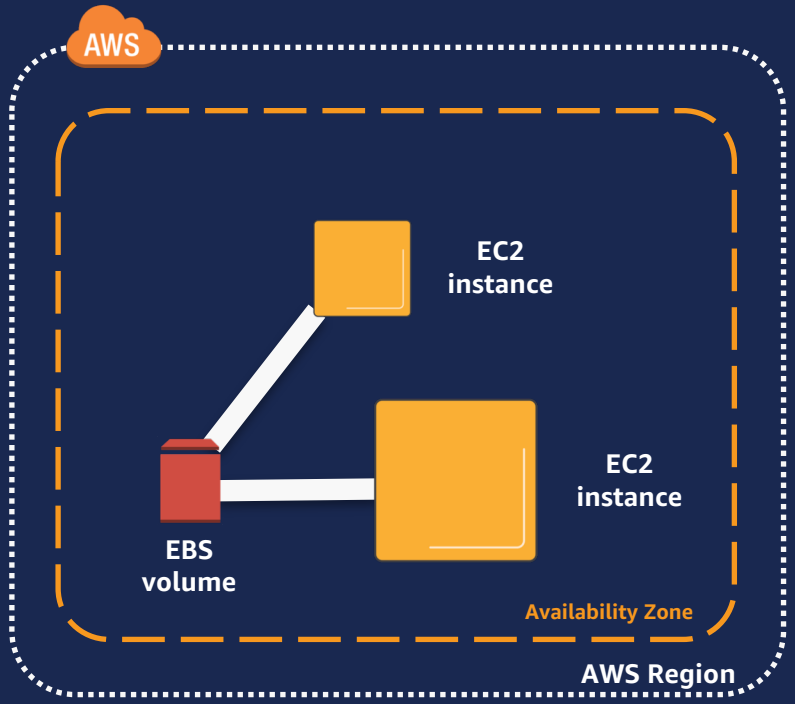
sc1

What is Amazon EC2 instance store?



- Local to instance
- Non-persistent data store
- Data not replicated (by default)
- No snapshot support
- SSD or HDD

What is EBS?

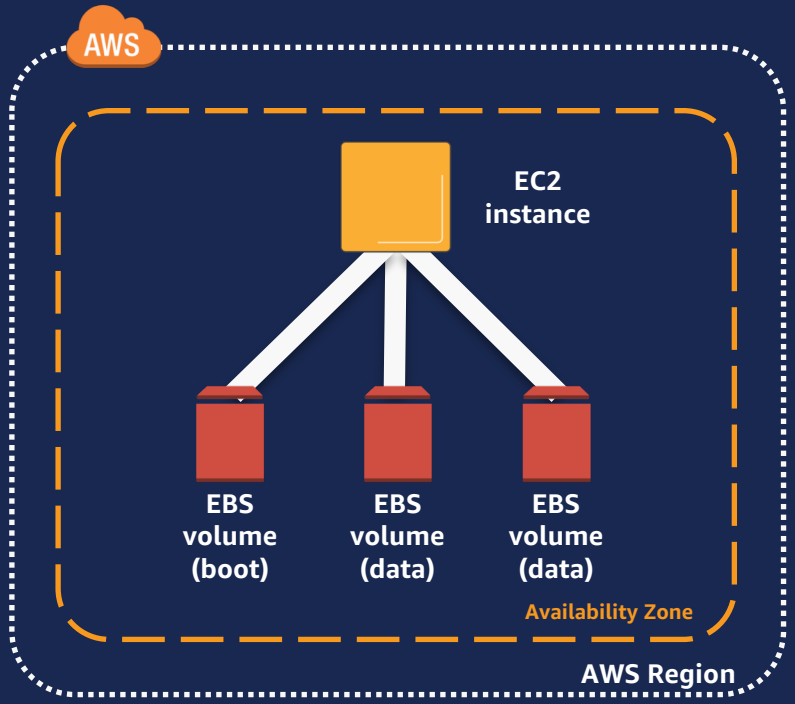


Volumes persist independent of EC2

Select storage and compute based on your workload

Detach and attach between instances within the same Availability Zone

What is EBS?

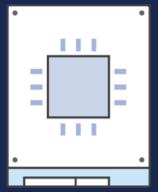


Volumes attach to one instance

Many volumes can attach to an instance

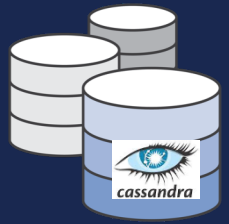
Separate boot and data volumes

Amazon Elastic Block Store (EBS) Use Cases



Solid State (SSD)

Hard Disk Drive (HDD)



Relational Databases

MySQL, PostgreSQL, SQL, Oracle, SAP

NonRelational/NoSQL

Cassandra, MongoDB, CouchDB

Big Data Analytics

Kafka, Splunk, Hadoop, Data Warehousing

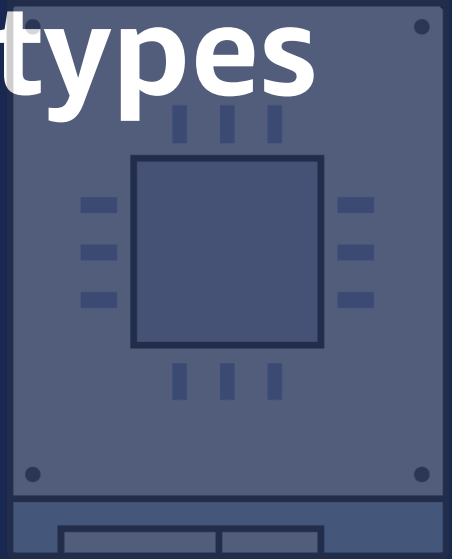
File / Media

CIFS/NFS, Transcoding, Encoding, Rendering

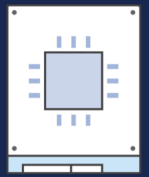
Benefits of EBS

- **Persistent** – delete instances without losing your data, detach and reattach volumes without having to replicate data between nodes
- **Elastic** –dynamically increase capacity, change volume types, and tune performance
- **Available** – designed for 99.999% availability
- **Durable** – designed for an annual failure rate (AFR) of between 0.1% - 0.2%. EBS Snapshots for backups.
- **Secure** – seamless support for data-at-rest and data-in-transit encryption between EC2 instances and EBS volumes

EBS volume types

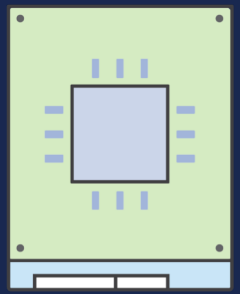


EBS volume types



SSD

HDD

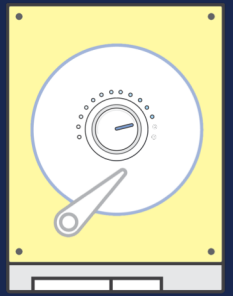


gp2

General Purpose SSD

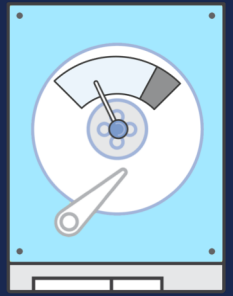
io1

Provisioned IOPS SSD



st1

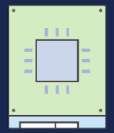
Throughput Optimized HDD



sc1

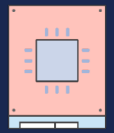
Cold HDD

Amazon EBS Offerings



GP2

IOPS performance for most workloads with predictable baseline and burst



IO1

Critical and IO intensive workloads needing consistent performance

SC1

Colder throughput oriented workloads (e.g. log processing)

ST1

Big data & analytics workloads needing consistent high throughput

Snapshots

Highly durable regional backups for EBS Volumes

SAN-like features at Cloud-scale – Performance and Data Services

Optimizing Value: Price to Performance

Simplifying customer experience

SSD: GP2 or IO1?

GP2

GP2 volumes are designed to deliver the provisioned performance 99% of the time.
IOPS performance for most workloads with predictable baseline and burst.

Provisioned IOPS (io1)

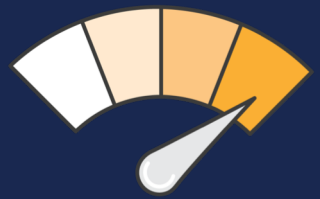
IO1 volumes deliver within 10 percent of the provisioned IOPS performance 99.9 percent of the time over a given year.
Critical and IO intensive workloads needing consistent performance.

Choosing an EBS volume type

What is more important to your workload:

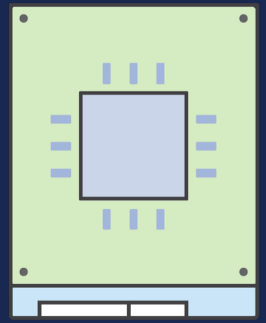
IOPS

or



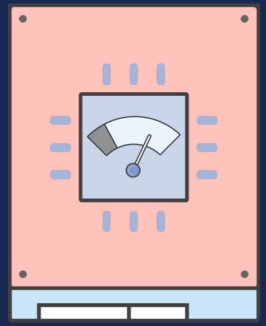
Throughput?

I/O Provisioned Volumes



gp2

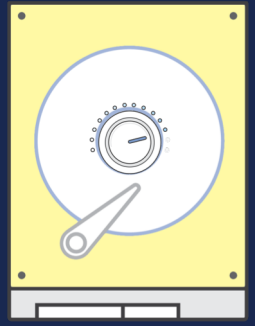
$0.119 per GB



io1

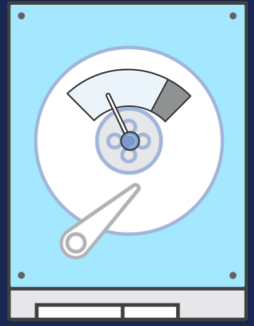
$0.149 per GB
$0.078 per PIOPS

Throughput Provisioned Volumes



st1

$0.054 per GB



sc1

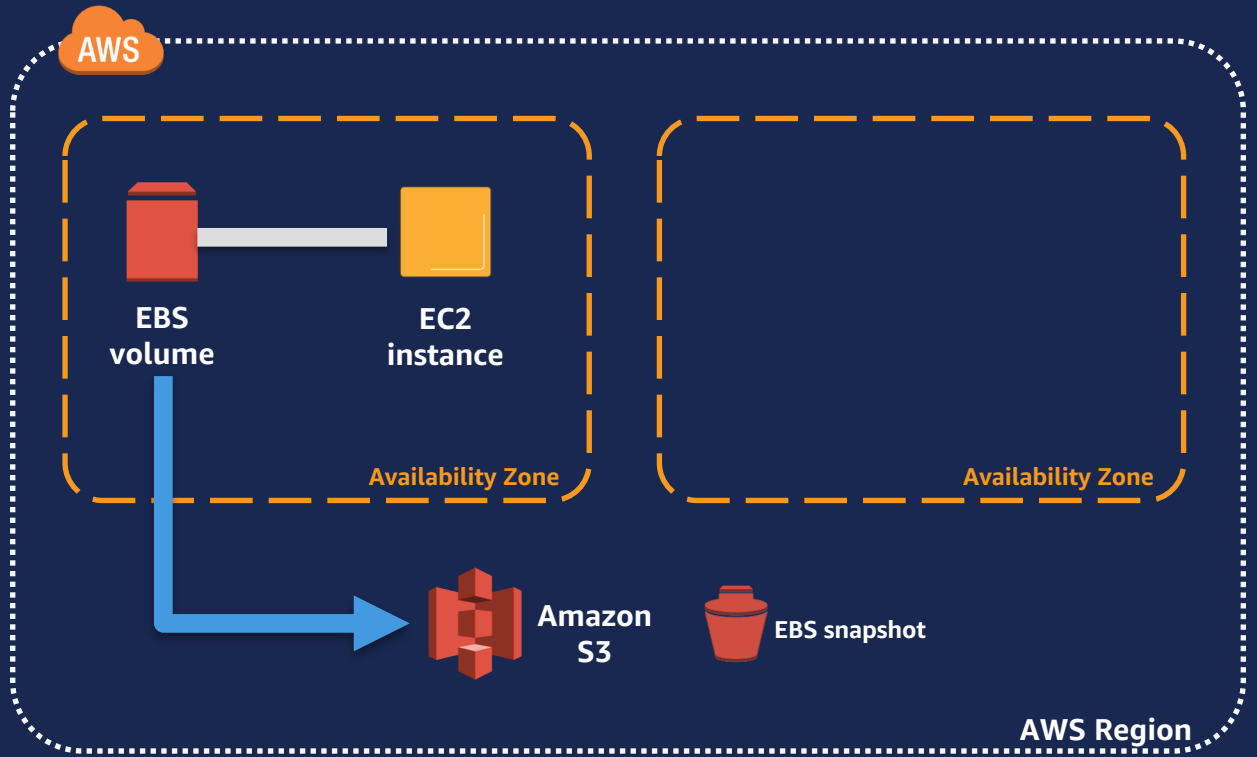
$0.03 per GB

Snapshot storage for all volume types is $0.054 per GB per month

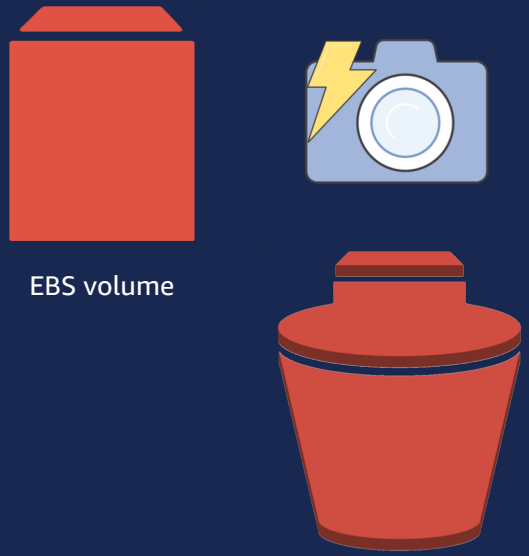
* All prices are per month, and from the eu-central-1 Region as of April 2018

EBS Snapshots

What is an EBS snapshot?

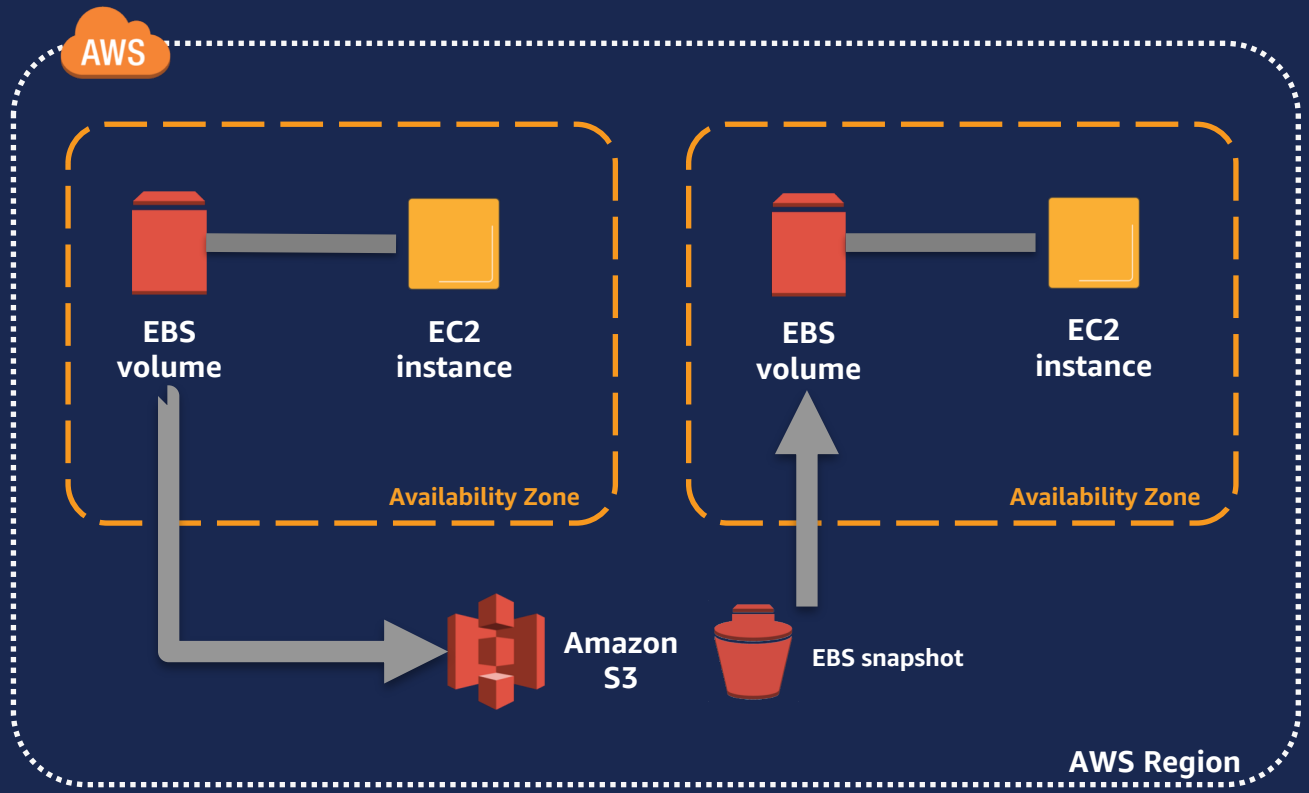


How does an EBS snapshot work?

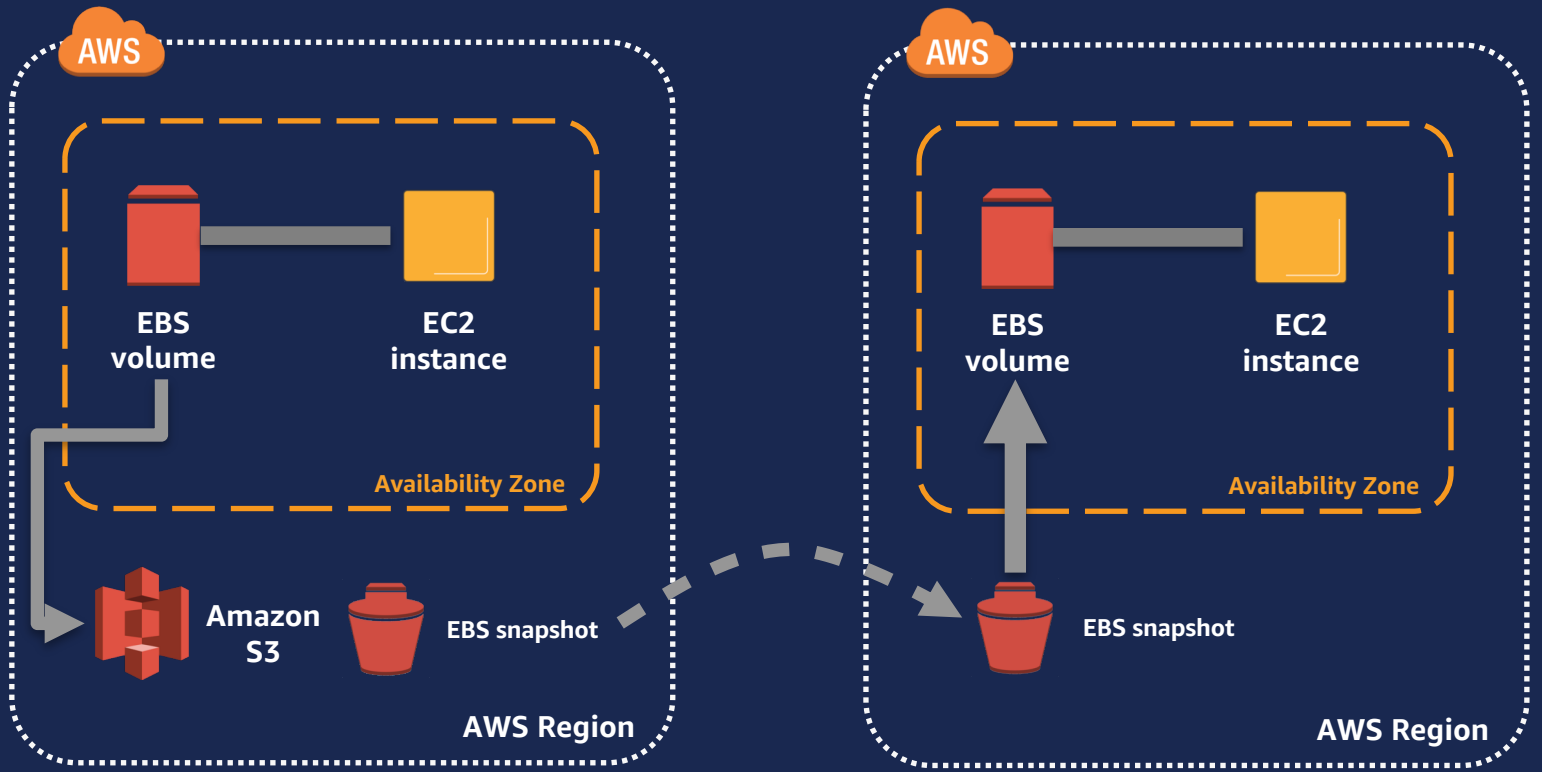


- Point-in-time backup of modified volume blocks
- Stored in S3, accessed via EBS APIs
- Subsequent snapshots are incremental
- Deleting snapshot will only remove data exclusive to that snapshot
- Crash consistent

What can you do with a snapshot?



What can you do with a snapshot?



EBS enables EC2 auto recovery

Amazon CloudWatch per-instance metric alarm:

StatusCheckFailed_System

When alarm triggers?

RECOVER instance

Instance retains:

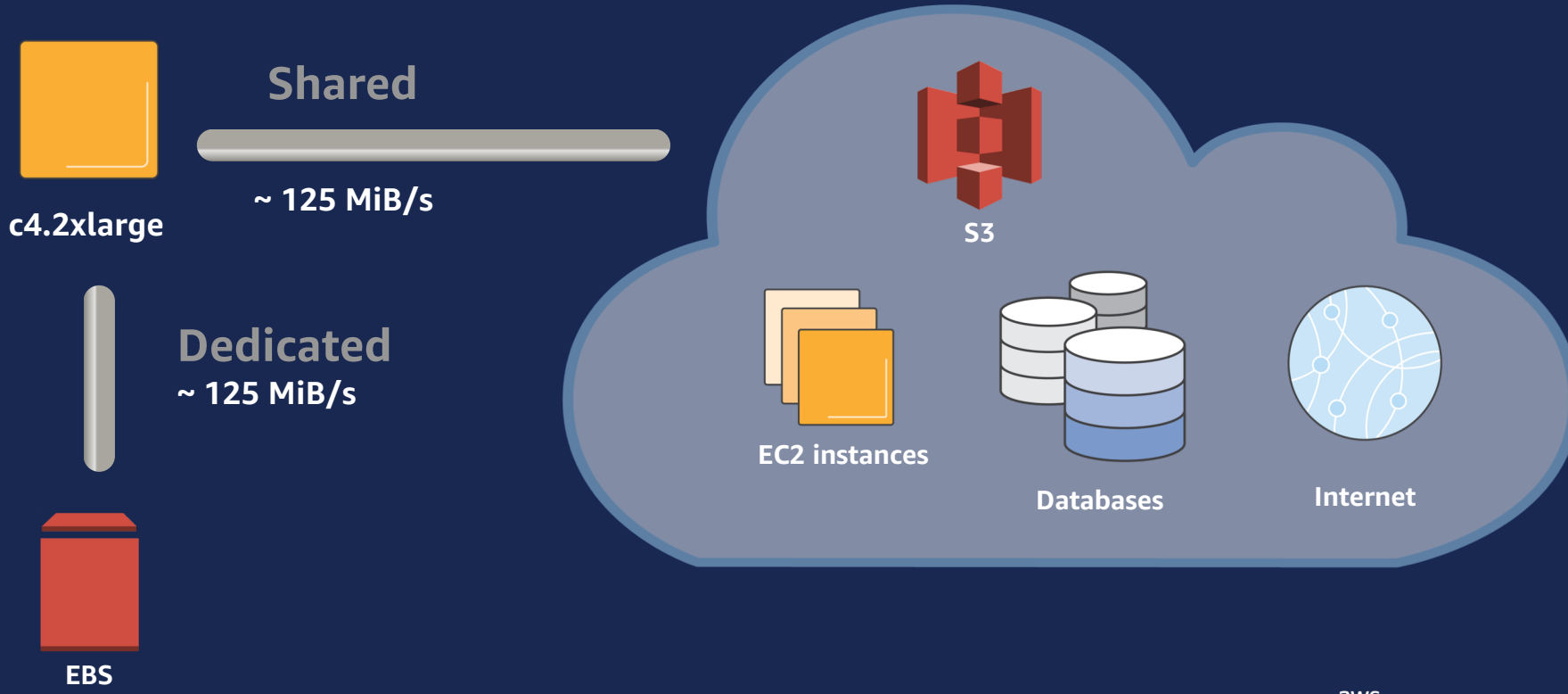
- Instance ID
- Instance metadata
- Private IP addresses
- Elastic IP addresses
- EBS volume attachments

** Supported on C3, C4, C5, M3, M4, P2, R3, R4, T2, and X1 instance types with EBS-only storage*

EBS Optimized

Choosing the right EC2 instance

What is an EBS-optimized instance?



What is an EBS-optimized instance?

- Dedicated network bandwidth for EBS I/O
- Up to 80,000 16k IOPS and 1.75 GiB/s throughput
- Enabled by default on most current generation instances
- Can be enabled at instance launch or on a running instance
- Not an option on some 10 Gbps instance types
  - (c3.8xlarge, r3.8xlarge, i2.8xlarge)

More details:
https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSOptimized.html

Performance: EBS-optimized

c4.large

Dedicated to EBS

500 Mbps ~ 62.5 MiB/s
4,000 16K IOPS

2 TiB GP2 volume:
6,000 IOPS
160 MiB/s max throughput

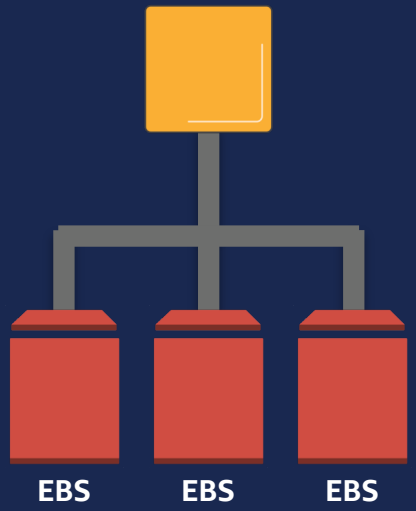
c4.2xlarge

Dedicated to EBS

1 Gbps ~ 125 MiB/s
8,000 16K IOPS

2 TiB GP2 volume:
6,000 IOPS
160 MiB/s max throughput

Best practice: RAID



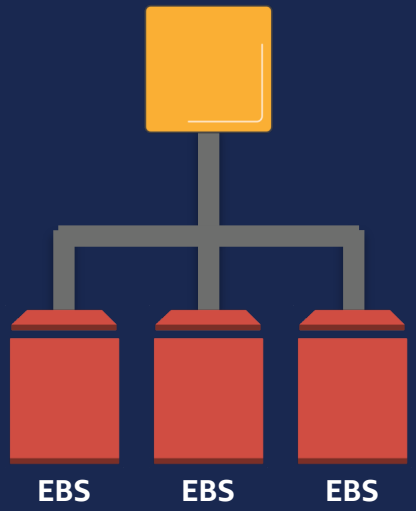
When to RAID?

- Storage requirement > 16 TiB

- Throughput requirement > 500 MiB/s
- Largest instance can support up to 1,750 MiB/s

- IOPS requirement > 32,000 @ 16K
- Largest instance can support up to 75,000 @ 16k

Best practice: RAID



Avoid RAID for redundancy

- EBS data is already replicated
- RAID1 halves available EBS bandwidth
- RAID5/6 loses 20% to 30% of usable I/O to parity

How to think about EFS perf relative to EBS

		Amazon EFS	Amazon EBS
Performance	Per-operation latency	Low, consistent	Lowest, consistent
	Throughput scale	Multiple GBs per second	Single GB per second
Characteristics	Data availability / durability	Stored redundantly across multiple AZs	Stored redundantly in a single AZ
	Access	1 to 1000s of EC2 instances, from multiple AZs, concurrently	Single EC2 instance in a single AZ
	Use cases	Big Data and analytics, media processing workflows, content management, web serving, home directories	Boot volumes, transactional and NoSQL databases, dataanalytics & warehousing, ETL

Oracle on EBS

Hosting Oracle Databases : Various Options

1

- App optimization
- Scaling
- High availabili
- Database backu
- DB s/w patche
- DB s/w install
- OS patches
- OS installatio
- Server maintena
- Rack & stack
- Power, HVAC,

Your databases in Amazon EC2

2

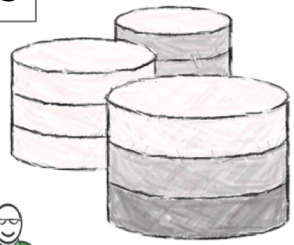
- App optimization
- Scaling
- High availability
- Database backups
- DB s/w patches
- DB s/w installs
- OS patches

- OS installation
- Server maintenanc
- Rack & stack
- Power, HVAC, net

amazon web services

Your Databases on Amazon RDS

3



App optimization

- Scaling
- High availability
- Database backups
- DB s/w patches
- DB s/w installs
- OS patches
- OS installation
- Server maintenance
- Rack & stack
- Power, HVAC, net

amazon web services

Amazon RDS

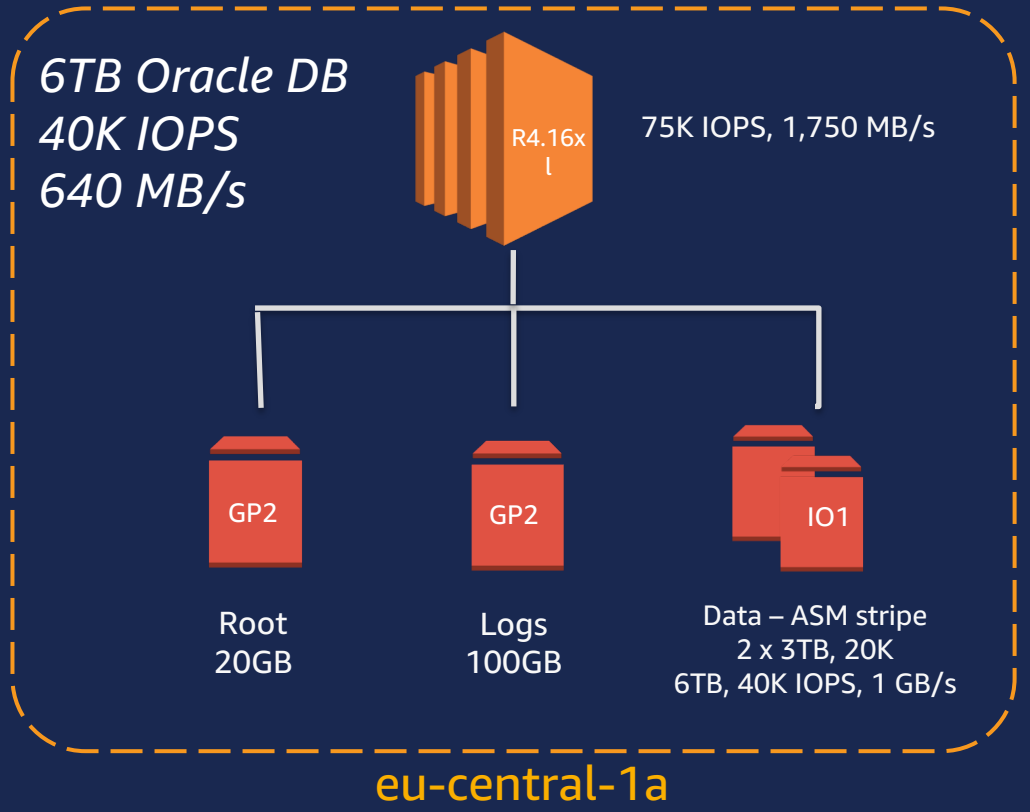
High — Maintenance — Low

Best Practice for running Oracle on AWS

If the workload is I/O bound use EBS Optimized Instance

Use IO1 to make sure Oracle Database meets IOPS requirement as per the AWR report (Automatic Workload Repository)

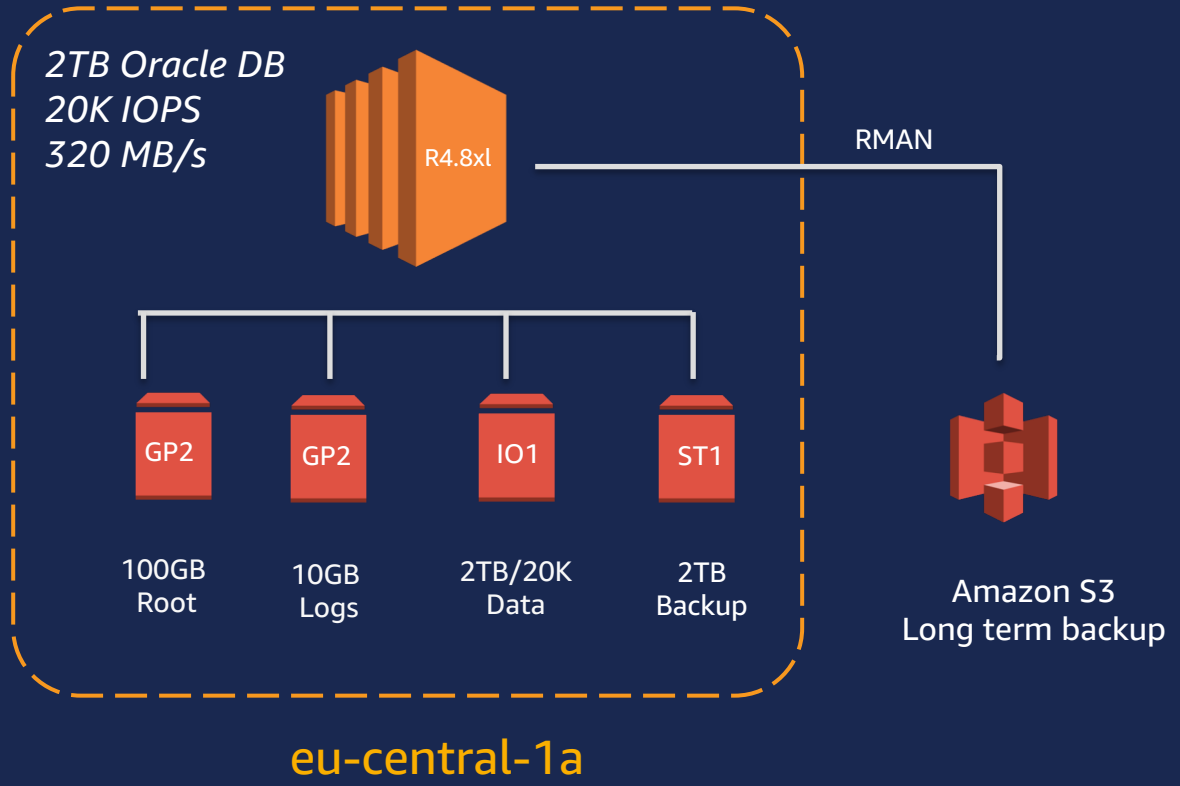
Use multiple EBS volumes and install ASM on it in order to keep the oracle files



Best Practice for running Oracle on AWS - Backup

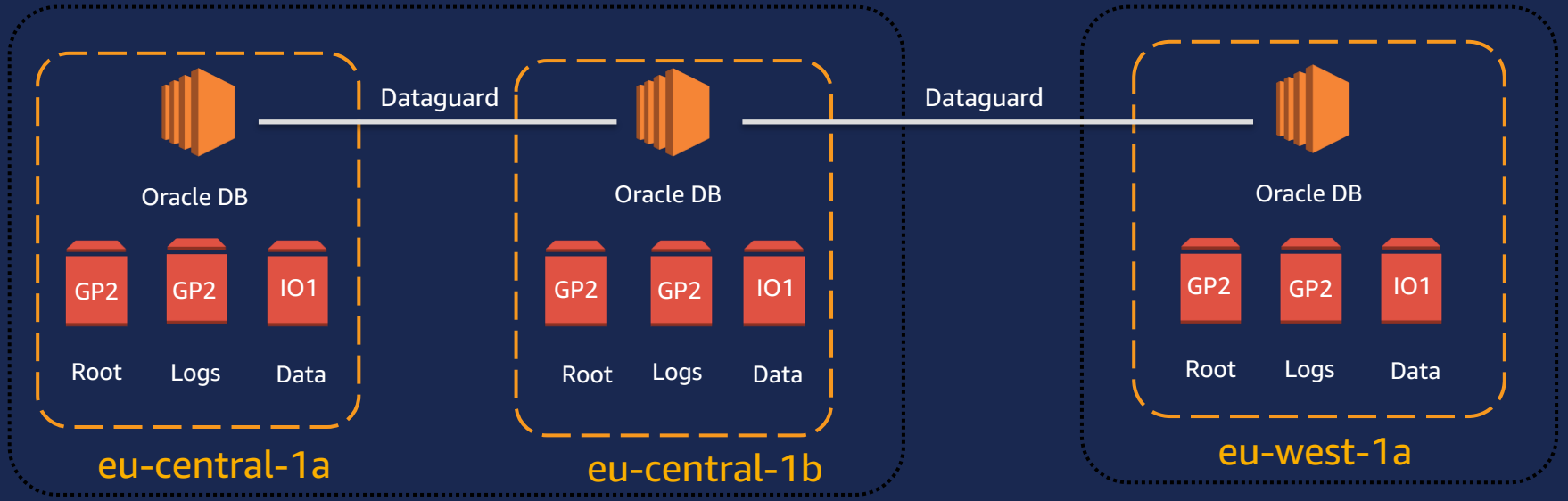
Create a Fast Recovery Area (FRA) in EBS ST1 for point-in-time recovery (0-7 days)

Use RMAN to backup the Oracle database files from EBS ST1 to S3 for long term retention.

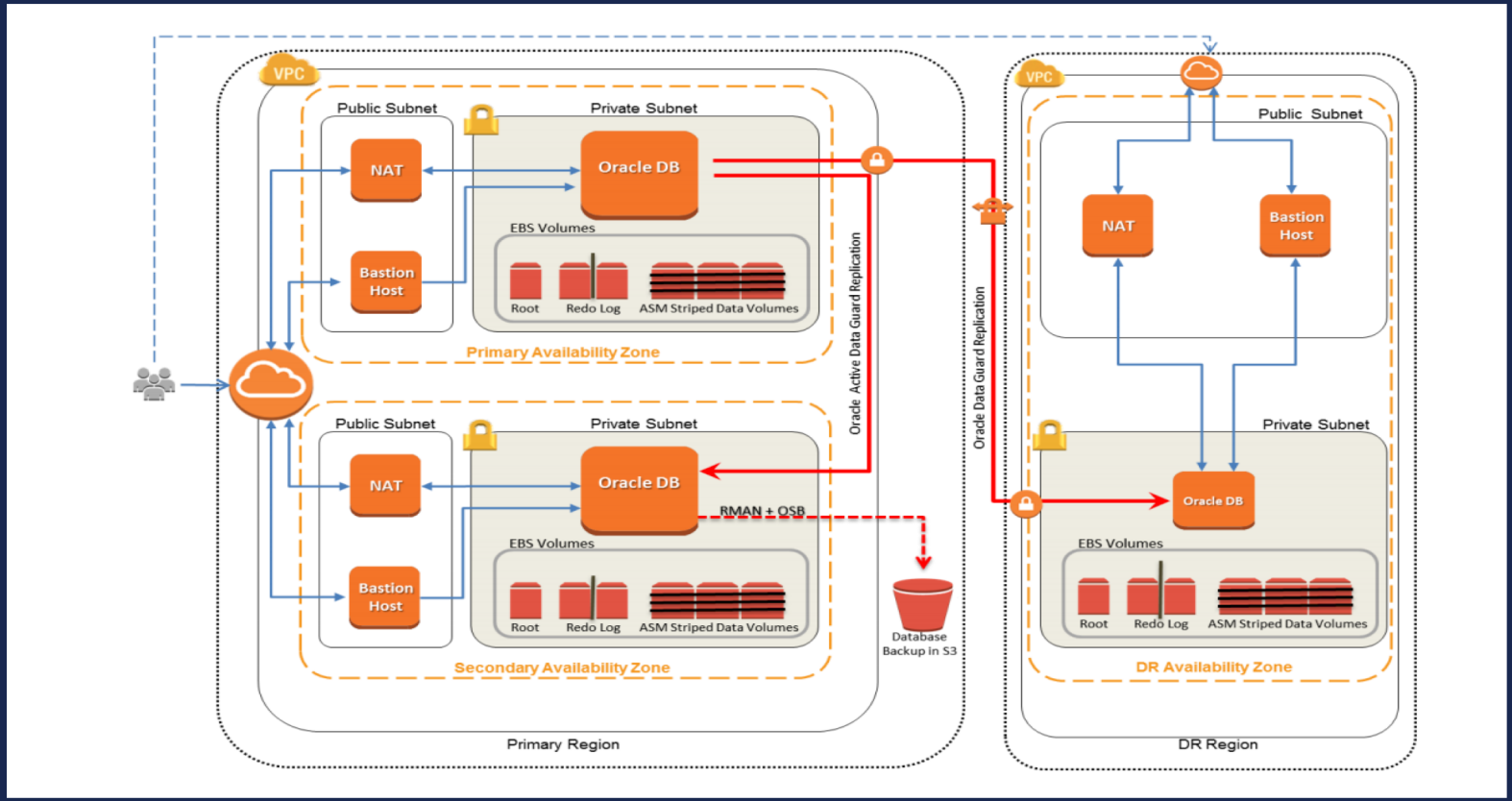


Best Practice for running Oracle on AWS - HA and DR

- Use Dataguard or Active Dataguard in a different AZ for HA
- Use Dataguard in a different region for DR
- Use Multi-AZ in RDS



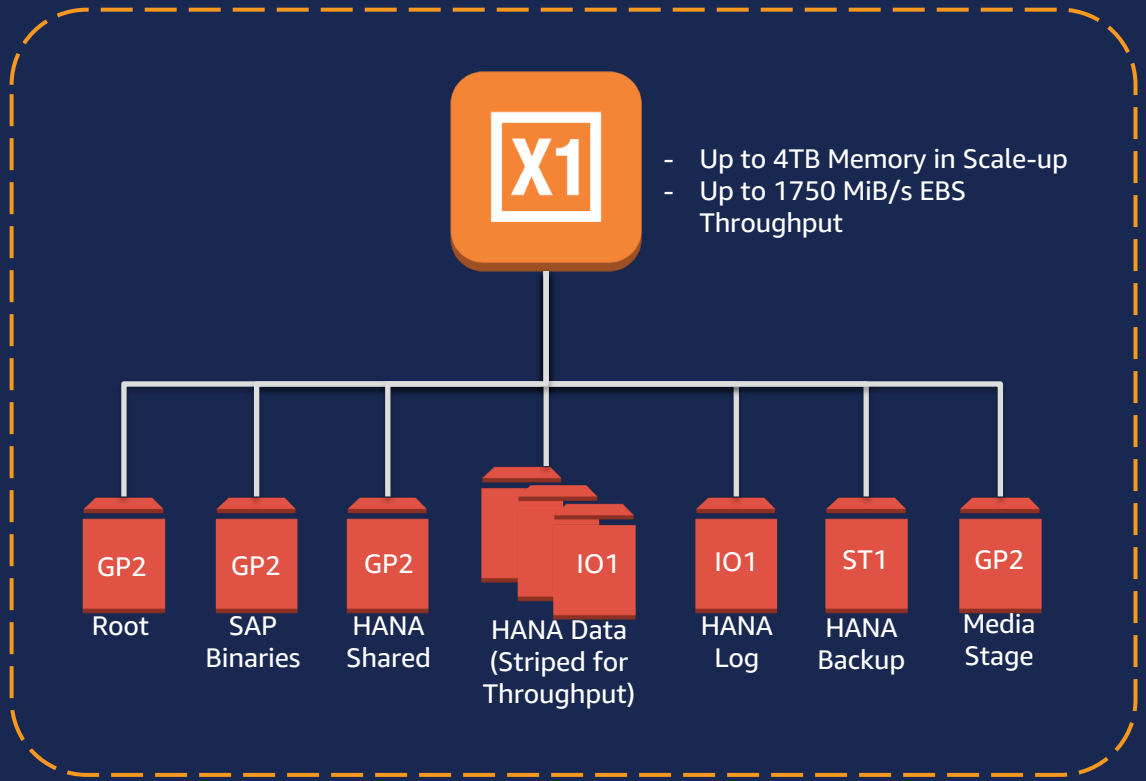
Oracle HA and DR Deployment Architecture



SAP on EBS

SAP HANA on AWS – Performance

- SAP mandates specific performance KPIs for SAP HANA Data and Log volumes
- Minimum 400 MB/s for SAP HANA Data
- Minimum 250 MB/s for SAP HANA Log
- Low latency for SAP HANA Log
- Both IO1 and GP2 Volumes are certified for SAP HANA Data and Log



SAP HANA on AWS – Performance

- x1e.32xlarge provides 4TB Memory
- 1750 MiB/s EBS Throughput
- Maximum data held in memory ~2TB

HANA Data Volume Type	HANA Data Volume Config	Maximum throughput to HANA Data	Time to load ~2TB Data into Memory
IO1	3 x 1600 GiB	1500 MiB/s	~23 Minutes
GP2	3 x 1600 GiB	480 MiB/s	~1 Hr 13 Minutes

Higher throughput to HANA Data

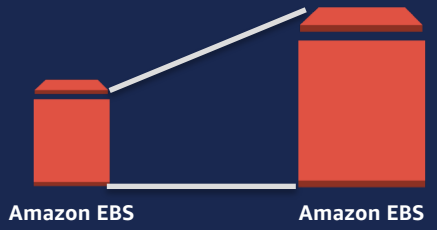
Lower time to load data during startup

✓ Use IO1 for business critical systems to reduce time to load data into memory

SAP HANA on AWS - Elasticity

- Scale-up from 244GiB memory to 4TB with EC2 elasticity (Stop Instance, Change Instance type and Start Instance).
- Increase storage size to match instance memory with EBS Elastic Volumes with simple API calls.

Just few minutes of downtime

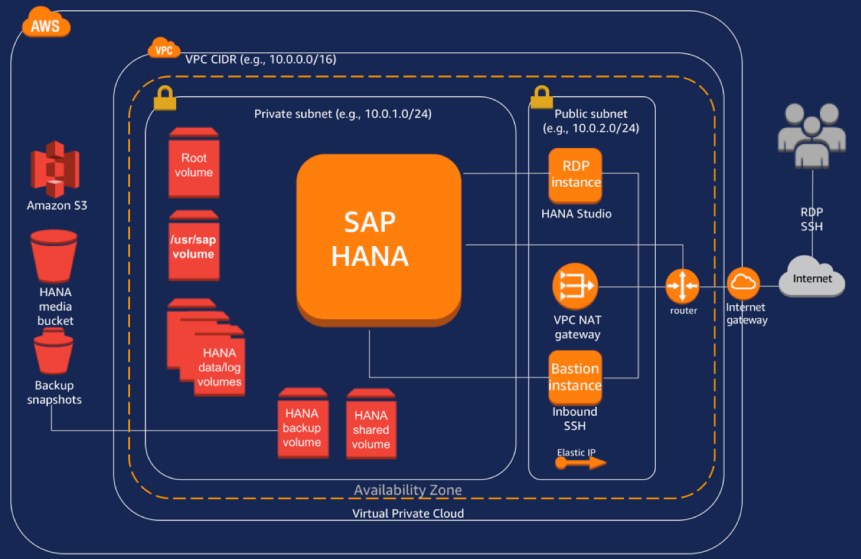


No downtime

SAP HANA on AWS – Deployment

SAP HANA QuickStart

- Uses AWS CloudFormation templates and scripts
- Provision SAP HANA system in < 1 hour
- Provides options to choose between GP2 and IO1 for HANA Data and Log
- Automatically provisions and configures EBS volumes based on instance type
- Can automatically encrypt the volumes during the launch



More Information - https://docs.aws.amazon.com/quickstart/latest/sap-hana/planning.html

SAP Workloads on AWS – Customer Momentum

Windows on EBS

Options for Deploying SQL Server on AWS

Amazon RDS for SQL Server

- **Consider RDS first**
- Focus on business value tasks
- High-level tuning
- Schema optimization
- No in-house database expertise
- Automatic Host Replacement

Scaling
High Availability
Database Backups
DBMS Patching
DBMS Install/Maintenance
OS Patching
OS Install/Maintenance
Power, HVAC, net

SQL Server on Amazon EC2

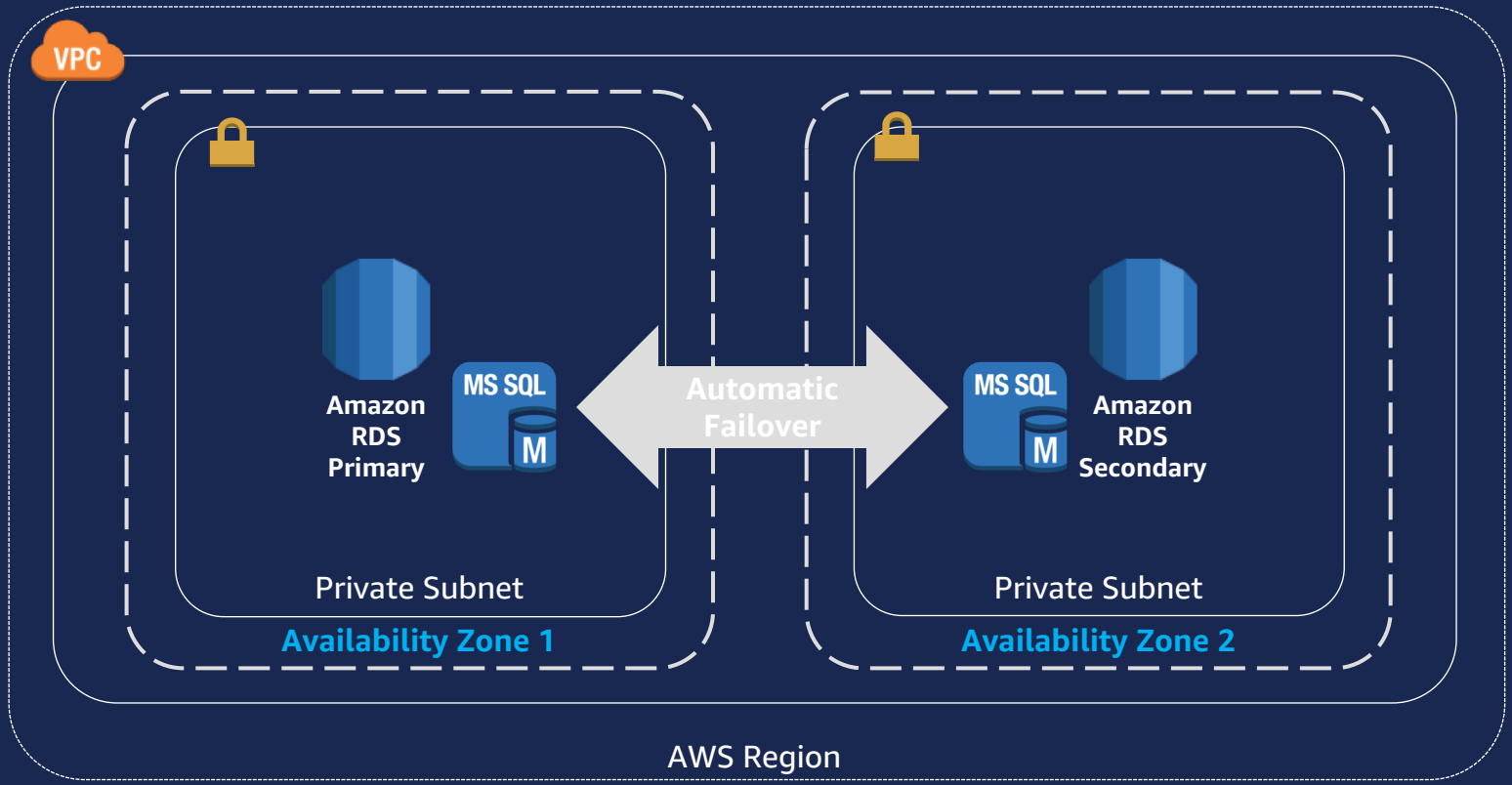
- Need full control over DB instance
- Replication
- Clustering
- Options that are not available in RDS

Scaling
High Availability
Database Backups
DBMS Patching
DBMS Install/Maintenance
OS Patching
OS Install/Maintenance
Power, HVAC, net

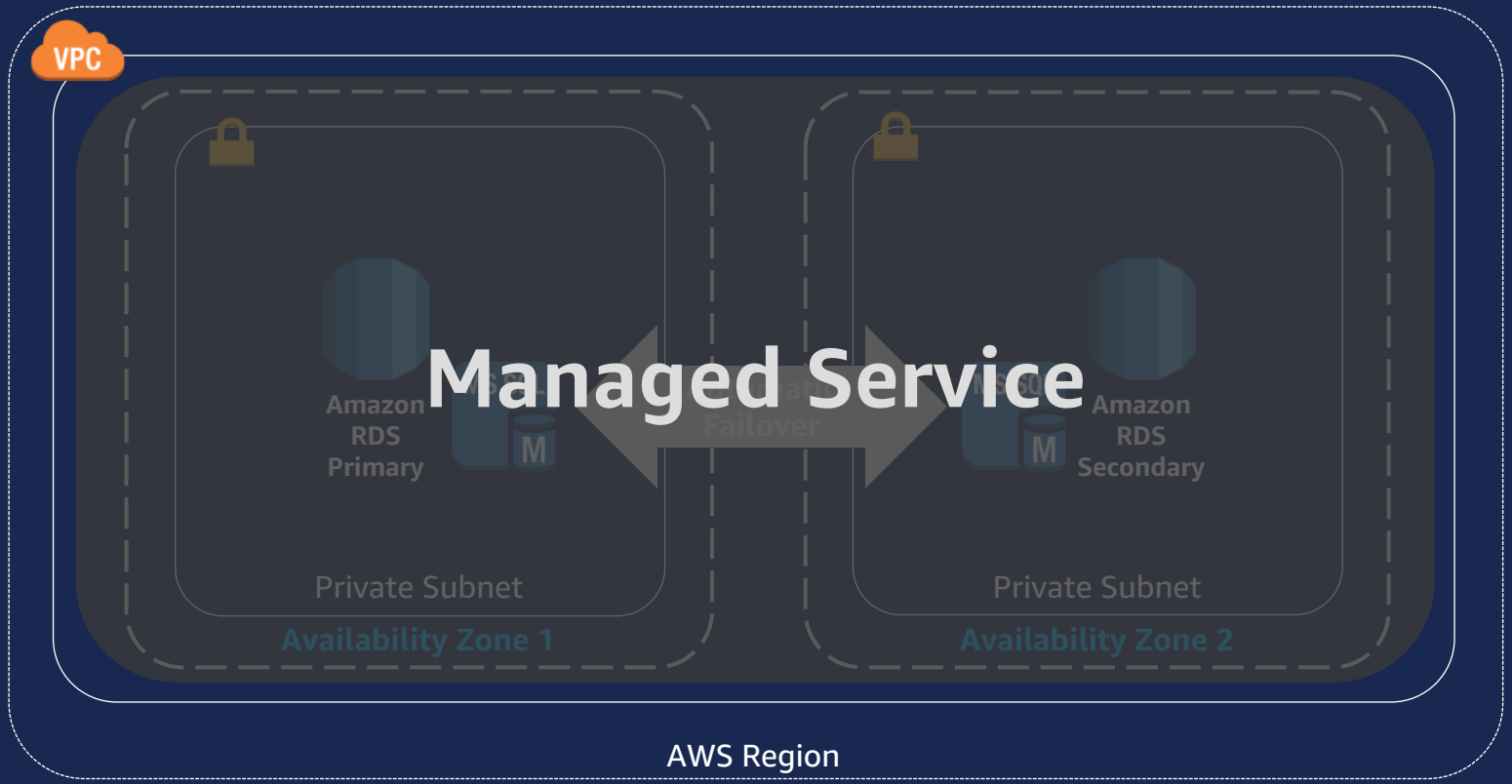
AWS managed

Customer managed

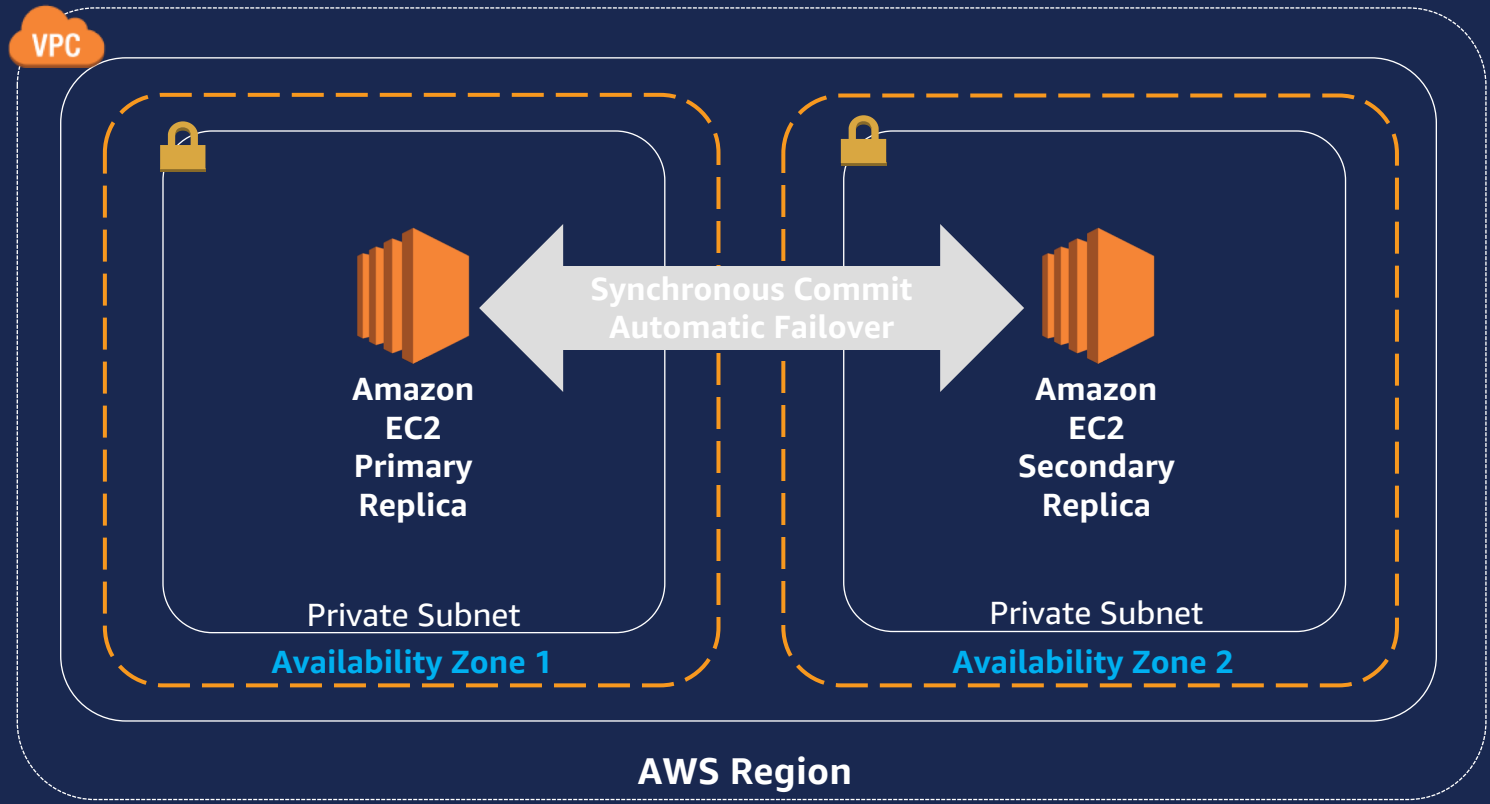
Multi-AZ SQL Server on Amazon RDS



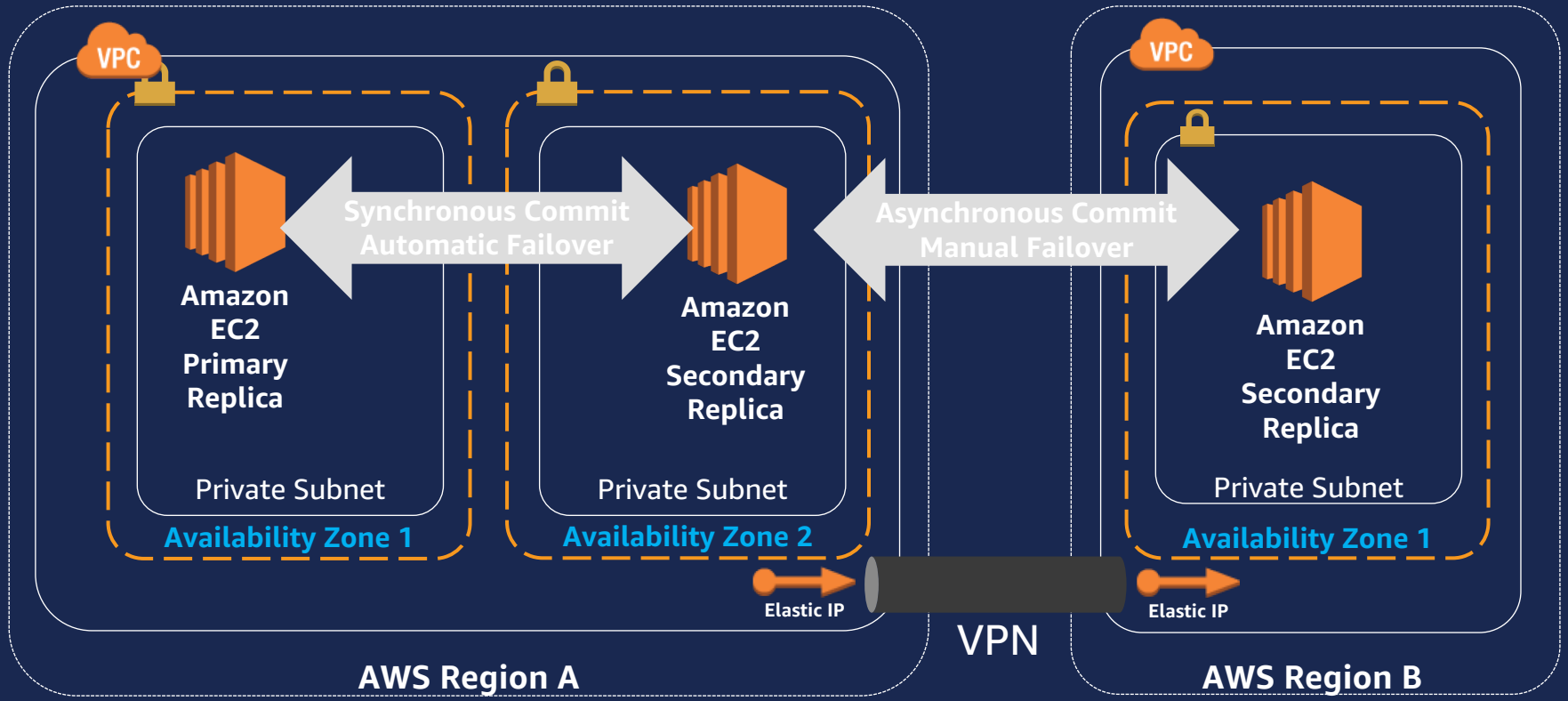
Multi-AZ SQL Server on Amazon RDS



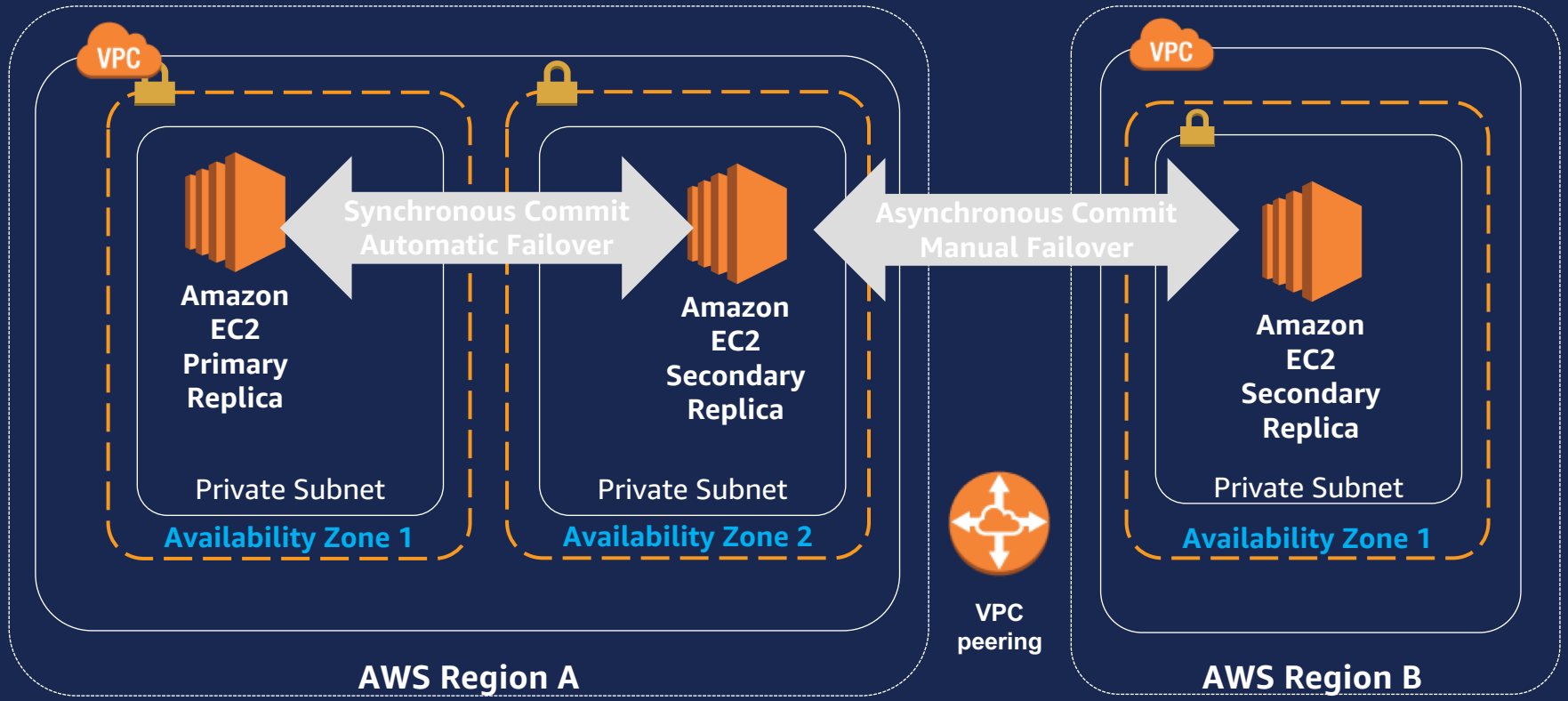
Multi-AZ AlwaysOn Availability Group



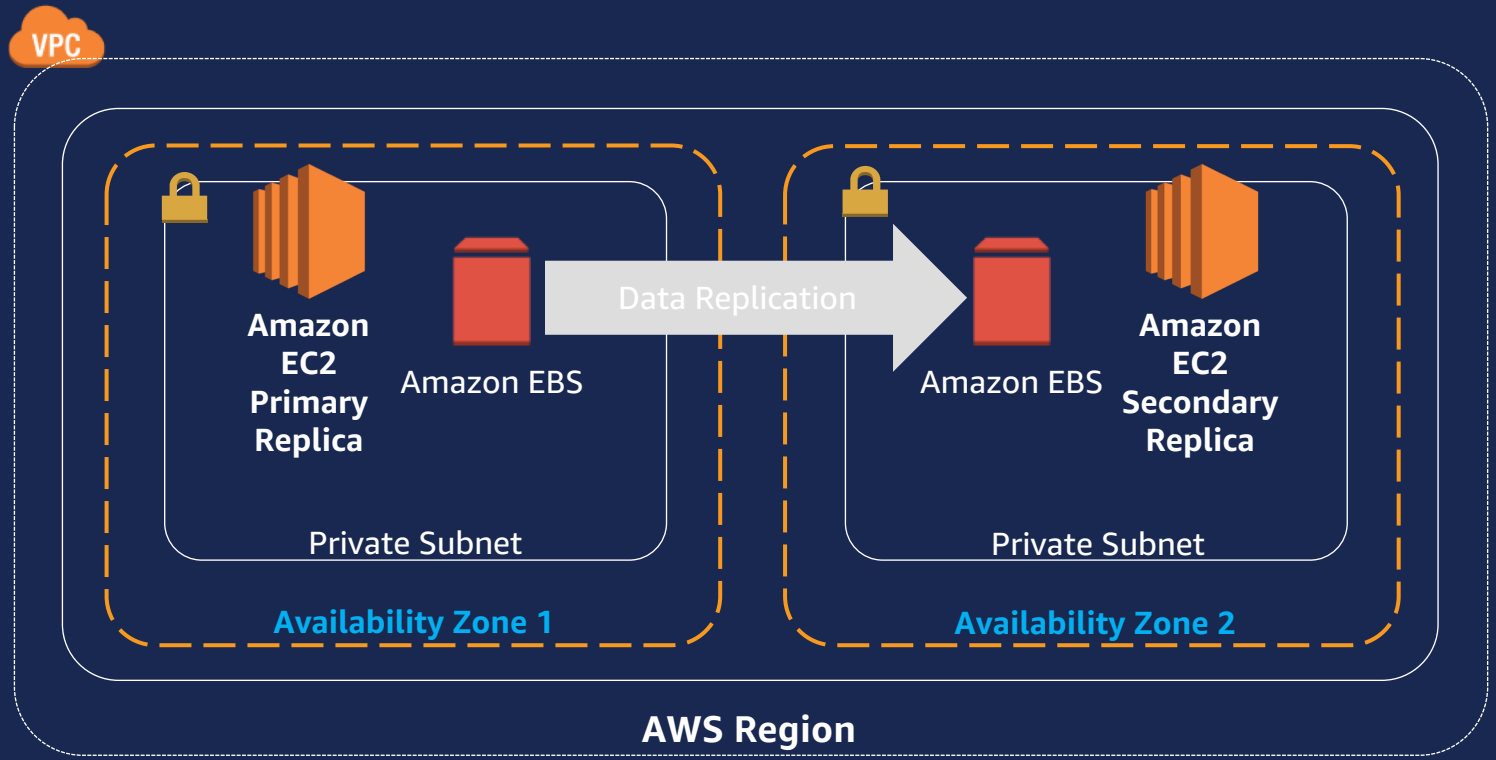
Multi-region AlwaysOn Availability Group



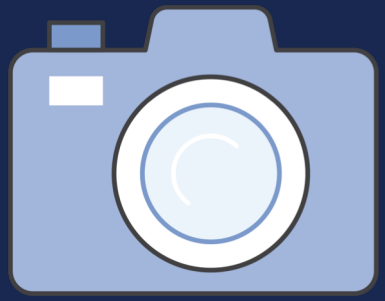
Multi-region AlwaysOn Availability Group



Multi-AZ Failover Cluster Instance



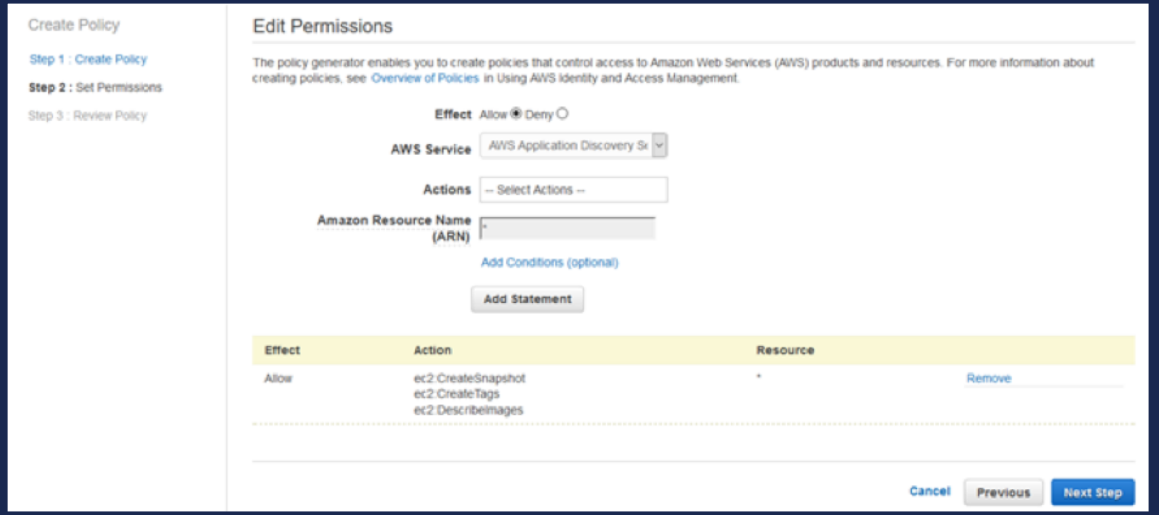
Best practice: Taking snapshots from Windows



1. sync equivalent available
2. Use Volume Shadow Copy Service (VSS) - aware utilities for backups
3. EBS: backups on dedicated volume for snapshots

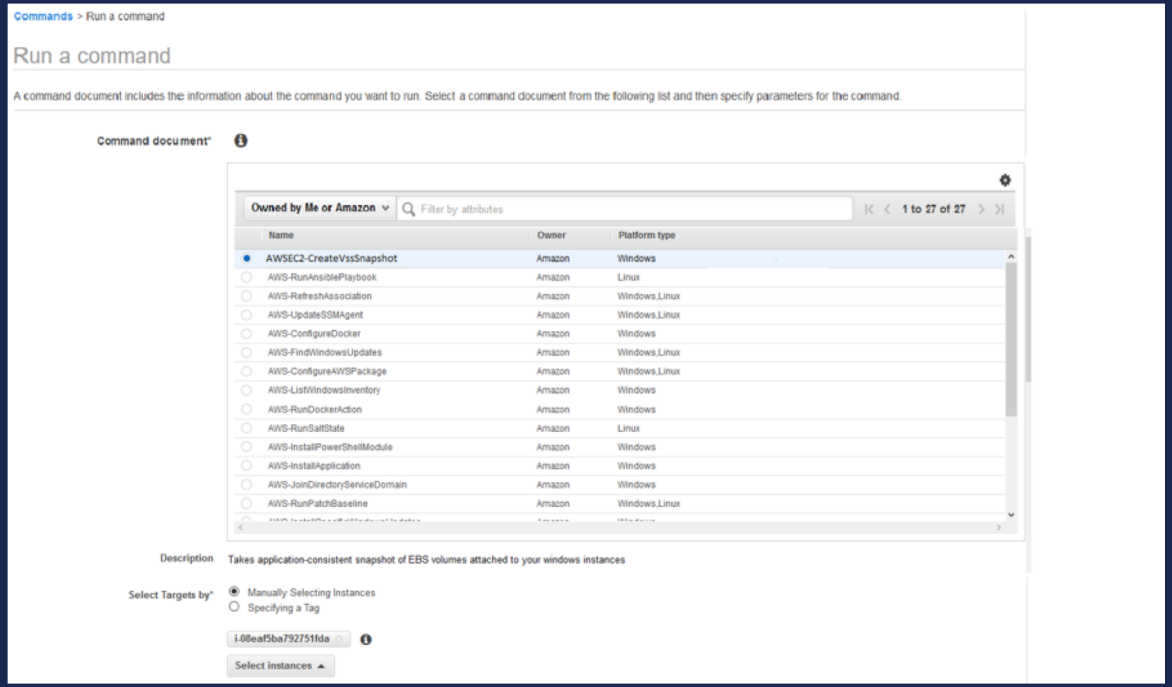
New: VSS support via EC2 SSM

- Use Policy Generator to create IAM policy for AWS Service “Amazon Simple Service Manager”
- *Actions: DescribeImages, CreateTags, and CreateSnapshot*
- *Create Amazon EC2 type IAM role and attach to Windows instances*



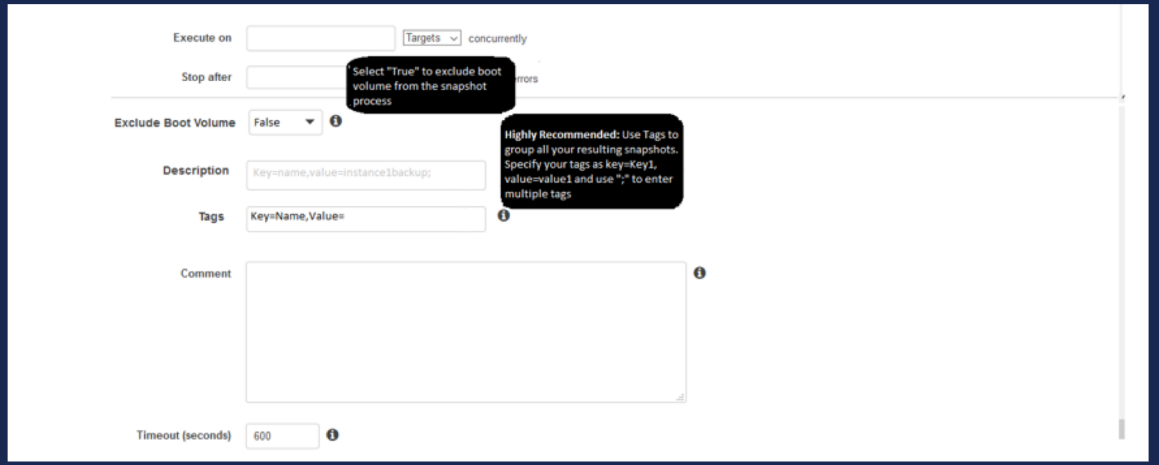
New: VSS support via EC2 SSM

- Call the Run Command *AWSEC2-CreateVssSnapshot*



New: VSS support via EC2 SSM

- Select the instance
- Add description, tags
- Can exclude boot volume
- Run Command makes the VSS agent flush I/O, freeze



- SSM VSS included in Microsoft Windows Server AMI version 2017.11.21 & up

Danke

Go Build!!