ORACLE
Cloud Infrastructure

Block Volume
Local NVMe Storage

Level 100

Rohit Rahi

November 2018

ORACLE

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Objectives

After completing this lesson, you should be able to:

- Understand use cases for local and block storage
- Describe OCI Block Volume service capabilities
- Create block volumes and walk through features

OCI Storage Services

	Local NVMe	Block Volume	File Storage	Object Storage	Archive Storage
Type	NVMe SSD based temporary storage	NVMe SSD based block storage	NFSv3 compatible file system	Highly durable Object storage	Long-term archival and backup
Durability	Non-persistent; survives reboots	Durable (multiple copies in an AD)	Durable (multiple copies in an AD)	Highly durable (multiple copies across ADs)	Highly durable (multiple copies across ADs)
Capacity	Terabytes+	Petabytes+	Exabytes+	Petabytes+	Petabytes+
Unit Size	51.2 TB for BM, 6.4-25.6 TB for VM	50 GB to 32 TB/vol 32 vols/instance	Up to 8 Exabyte	10 TB/object	10 TB/object
Use cases	Big Data, OLTP, high performance workloads	Apps that require SAN like features (Oracle DB, VMW, Exchange)	Apps that require shared file system (EBS, HPC)	Unstructured data incl. logs, images, videos	Long term archival and backups (Oracle DB backups)

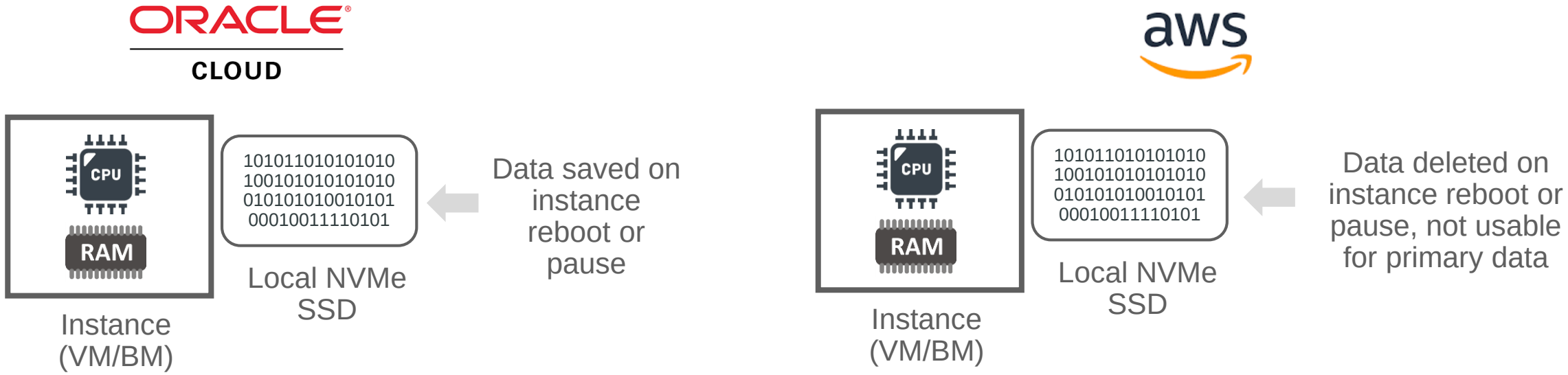
Local NVMe SSD Devices

- Some instance shapes in OCI include locally attached NVMe devices
- Local NVMe SSD can be used for workloads that have high storage performance requirements
- Locally attached SSDs are not protected and OCI provides no RAID, snapshots, backups capabilities for these devices
- Customers are responsible for the durability of data on the local SSDs

Instance type	NVMe SSD Devices
BM.DenseIO2.52	8 drives = 51.2 TB raw
VM.DenseIO2.8	2 drive = 6.4 TB raw
VM.DenseIO2.16	4 drives = 12.8 TB raw
VM.DenseIO2.24	8 drives = 25.6 TB raw

```
[opc@nvme ~]$ lsblk
NAME    MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
nvme0n1 259:0    0  5.8T  0 disk
nvme1n1 259:3    0  5.8T  0 disk
nvme2n1 259:1    0  5.8T  0 disk
nvme3n1 259:2    0  5.8T  0 disk
nvme4n1 259:5    0  5.8T  0 disk
nvme5n1 259:6    0  5.8T  0 disk
nvme6n1 259:4    0  5.8T  0 disk
nvme7n1 259:7    0  5.8T  0 disk
sda       8:0    0 46.6G  0 disk
├─sda2    8:2    0    8G  0 part [SWAP]
├─sda3    8:3    0 38.4G  0 part /
└─sda1    8:1    0  200M  0 part /boot/efi
```

NVMe SSD Persisted - Reboot/Pause



"With Oracle Cloud Infrastructure, companies can leverage NVMe for persistent storage to host databases and applications. However, other cloud providers typically do not offer such a capability. In cases where NVMe storage was an option with other vendors, it was not persistent. This meant that the multi-terabyte database that researchers loaded to this storage was lost when the server stopped.

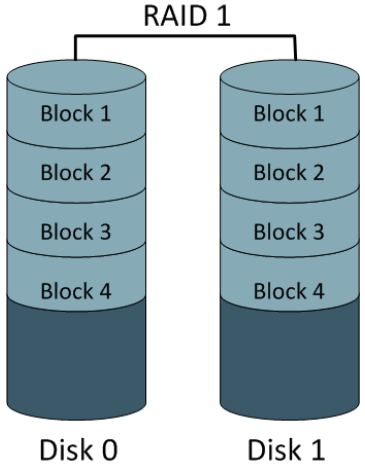
Accenture

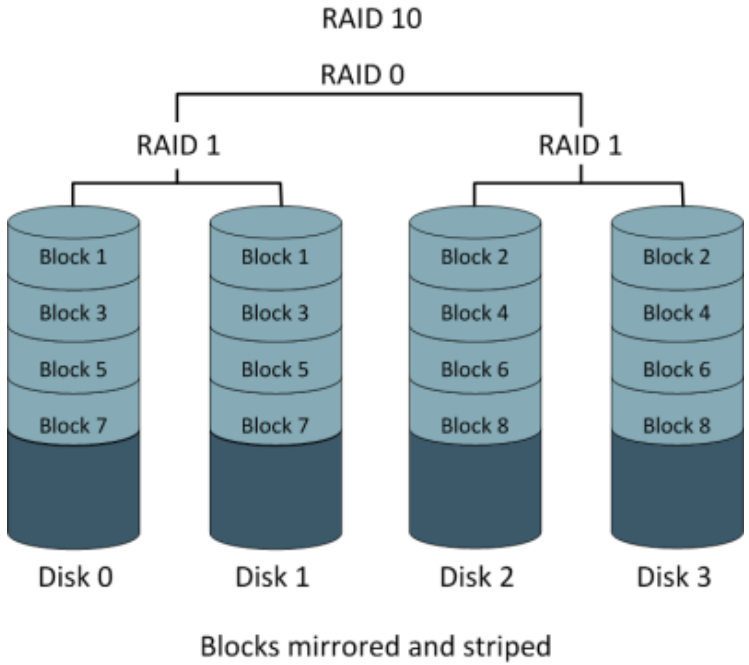
Protecting NVMe SSD Devices

RAID 1: An exact copy (or mirror) of a set of data on two or more disks

RAID 10: Stripes data across multiple mirrored pairs. As long as one disk in each mirrored pair is functional, data can be retrieved

RAID 6: Block-level striping with two parity blocks distributed across all member disks





Blocks mirrored and striped

Blocks striped with dual parity across drives

SLA for NVMe Performance

Shape	Minimum Supported IOPS
VM.DenseIO1.4	200k
VM.DenseIO1.8	250k
VM.DenseIO1.16	400k
BM.DenseIO1.36	2.5MM
VM.DenseIO2.8	250k
VM.DenseIO2.16	400k
VM.DenseIO2.24	800k
BM.DenseIO2.52	3.0MM

- OCI provides a service-level agreement (SLA) for NVMe performance
- Measured against 4k block sizes with 100% random write workload on Dense IO shapes where the drive is in a steady-state of operation
- Run test on Oracle Linux shapes with 3rd party Benchmark Suites, https://github.com/cloudharmony/block-storage

Block Volume Service

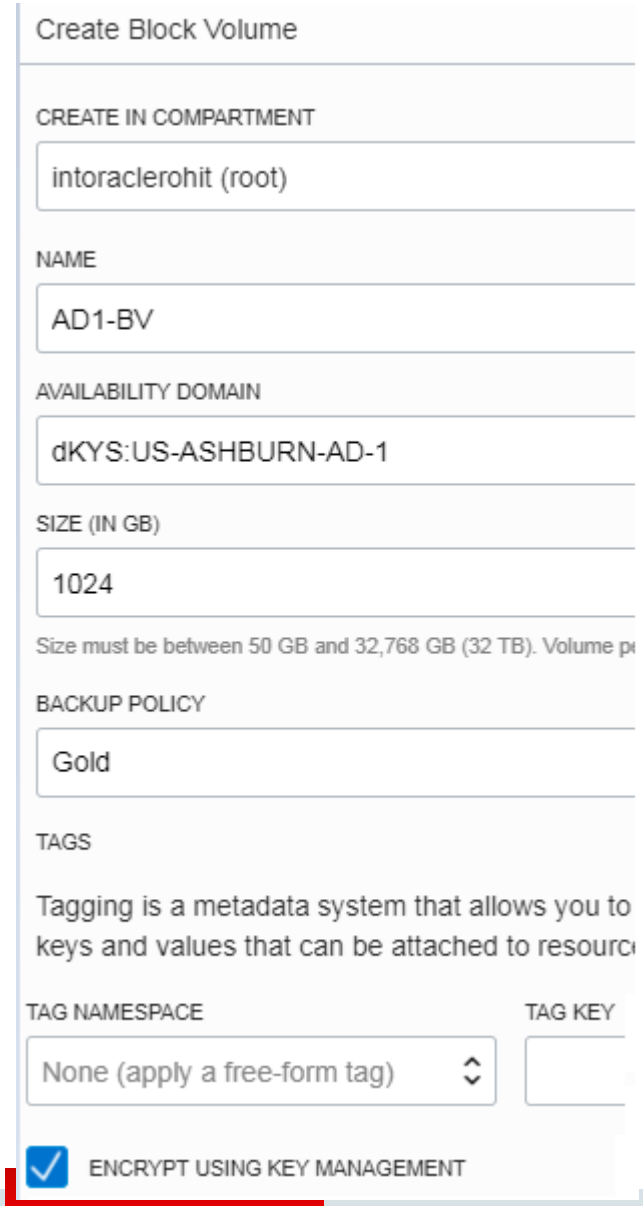
- Block Volume Service let you store data on block volumes independently and beyond the lifespan of compute instances
- Block volumes operates at the raw storage device level and manages data as a set of numbered, fixed-size blocks using a protocol such as iSCSI
- You can create, attach, connect, and move volumes, as needed, to meet your storage and application requirements
- Typical Scenarios
  - Persistent and Durable Storage
  - Expand an Instance's Storage
  - Instance Scaling

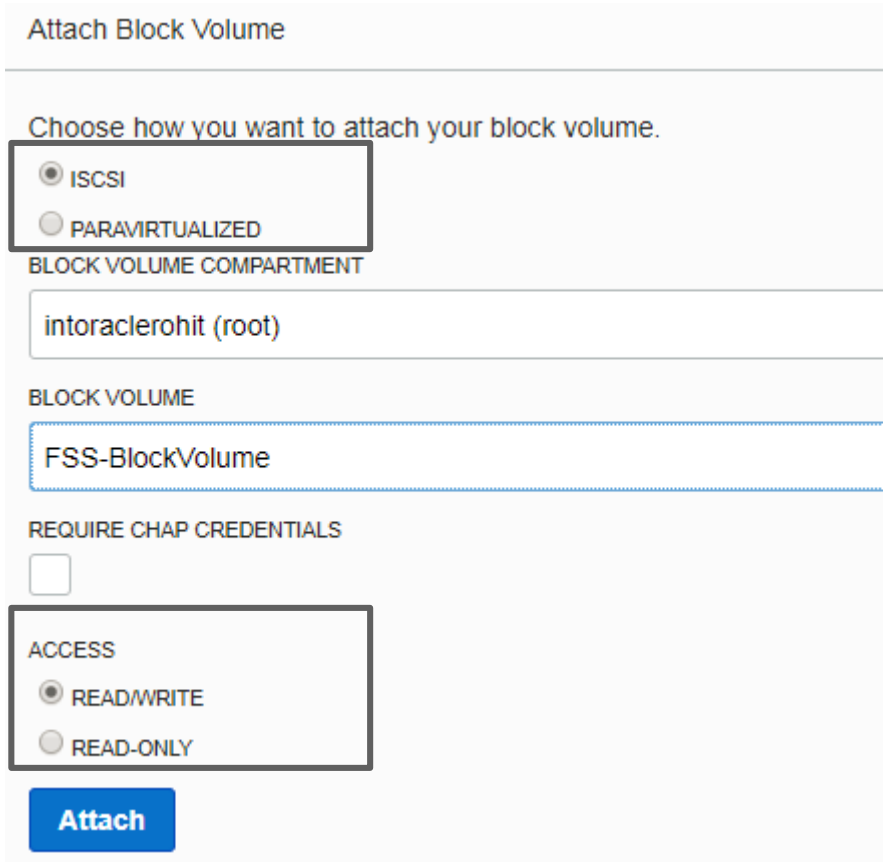
Block Volume Service (contd.)

Capacity	Configurable: 50 GB to 32 TB (1GB increments)
Perf: disk type	NVMe SSD based
Perf: IOPS	60 IOPS/GB - up to 25K IOPS*
Perf: Throughput/Vol	480 KBPS/GB - up to 320 MBPS**
Perf: Latency (P95)	Sub-millisecond latencies
Perf: Per-instance Limits	• 32 attachments/instance, up to 1 PB (32 TB/volume x 32 volumes/instance) • Up to 400K IOPS, near line rate throughput
Durability	Multiple replicas across multiple storage servers within the AD
Security	Encrypted at rest and transit
Restore from Backups (RTO)	<1 minute, regardless of size
Backup Performance (RPO)	~30 minutes (for 2TB), via snapshot

* For Bare Metal or 8-core+ VM compute instance, using 4KB blocks. VM perf is limited by VM network bandwidth.
** At 256 KB block size

Creating and Attaching a Block Volume





Paravirtualization is a light virtualization technique where a VM utilizes hypervisor APIs to access remote storage directly as if it were a local device

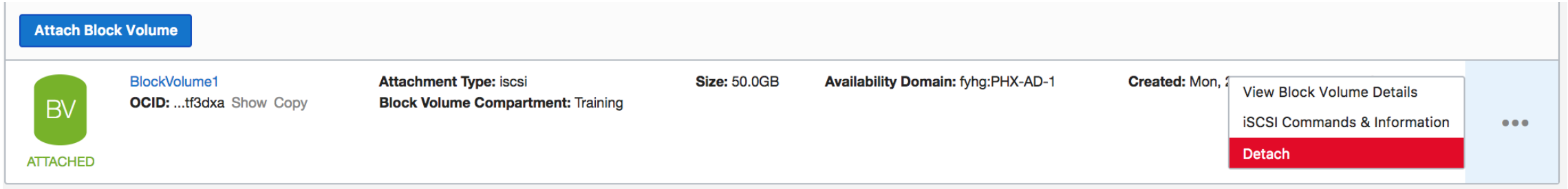
iSCSI block storage attachment utilizes the internal storage stack in the guest OS and network hardware virtualization to access block volumes. Hypervisor is not involved in the iSCSI attachment process

By default, all Block Volumes are Read/Write

Block Volume can also be read-only to prevent against accidental modification

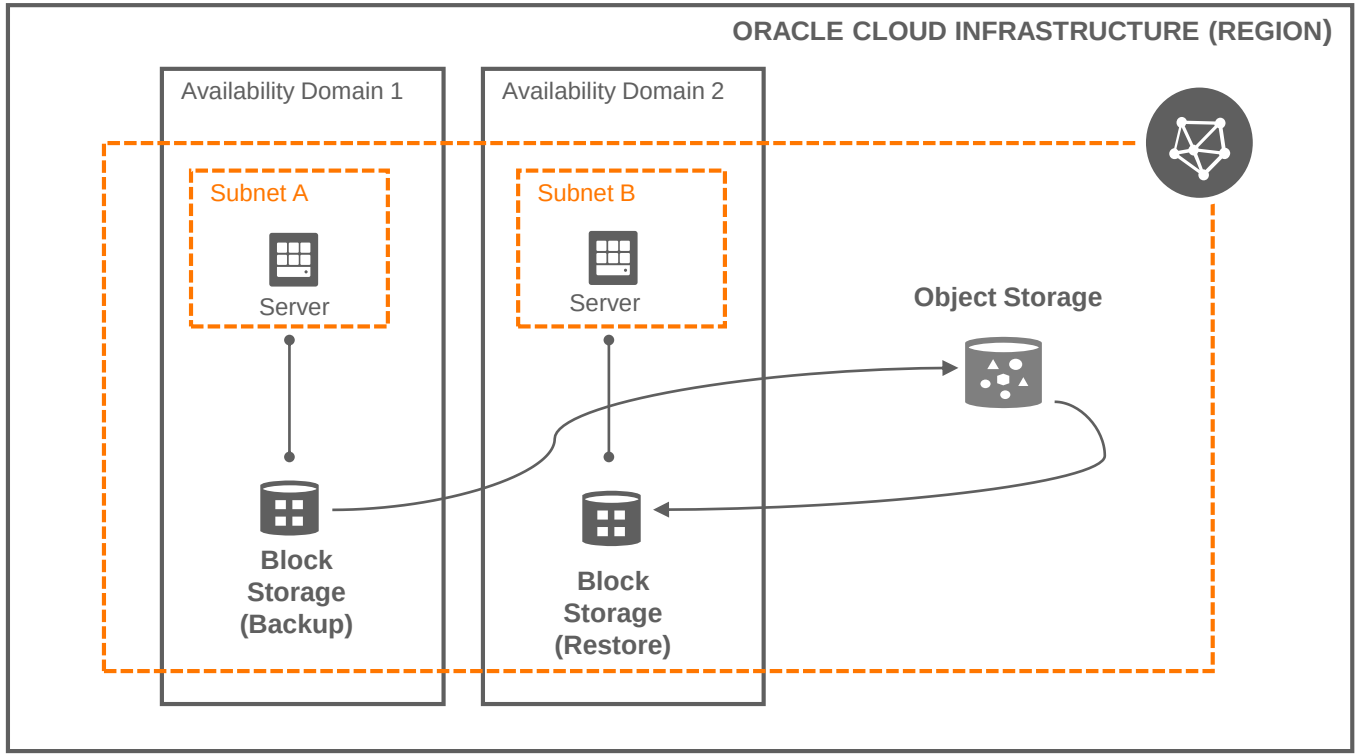
Detaching and Deleting Block Volumes

- When an instance no longer requires a block volume, you can disconnect and then detach it from the instance without any loss of data
- When you attach the same volume to another instance or to the same instance, DO NOT FORMAT the disk volume. Otherwise, you will lose all the data on the volume
- When the volume itself is no longer needed, you can delete the block volume
- You cannot undo a delete operation. Any data on a volume will be permanently deleted once the volume is deleted



Backup and Restoration

- Complete point-in-time complete snapshot copy of your block volumes
- Encrypted and stored in the Object Storage Service, and can be restored as new volumes to any Availability Domain within the same region



Backup and Restoration

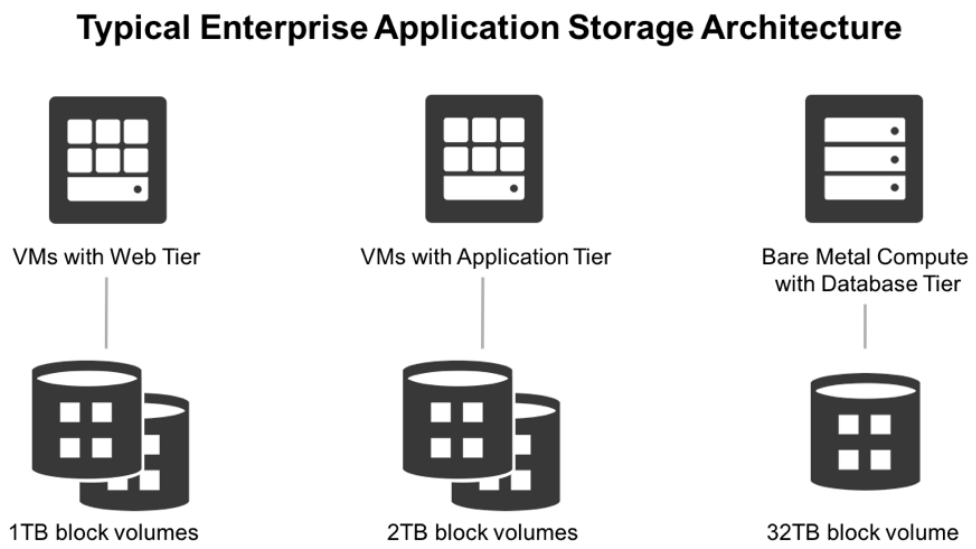
- Backup options:
  - On-demand, one-off: point-in-time snapshot
  - Automated policy-based: backups automatically on a schedule and retain them based on the selected backup policy. Three backup policies:
    - Bronze: monthly incremental backups, retained for twelve months (+full yearly backup, retained for 5 years)
    - Silver: weekly incremental backups, retained for four weeks (+ Bronze)
    - Gold: daily incremental backups, retained for seven days (+Silver, + Bronze)

- On-demand, one-off block volume backups provide a choice of incremental versus full backup options
- Can restore a volume in less than a minute regardless of the volume size

Clone

- Cloning allows copying an entire existing block volume to a new volume without needing to go through a backup and restore process
- Clone is a point-in-time direct disk-to-disk deep copy an of entire volume
- The clone operation is immediate, but actual copying of data happens in the background and can take up to 15 minutes for 1 TB volume
- A clone can only be created in the same AD with no need of detaching the source volume before cloning it
- A clone can be attached and used as regular volume when its lifecycle state changes from ”PROVISIONING” to "AVAILABLE", usually within seconds (At this time, the data is being copied in the background)

Volume Groups



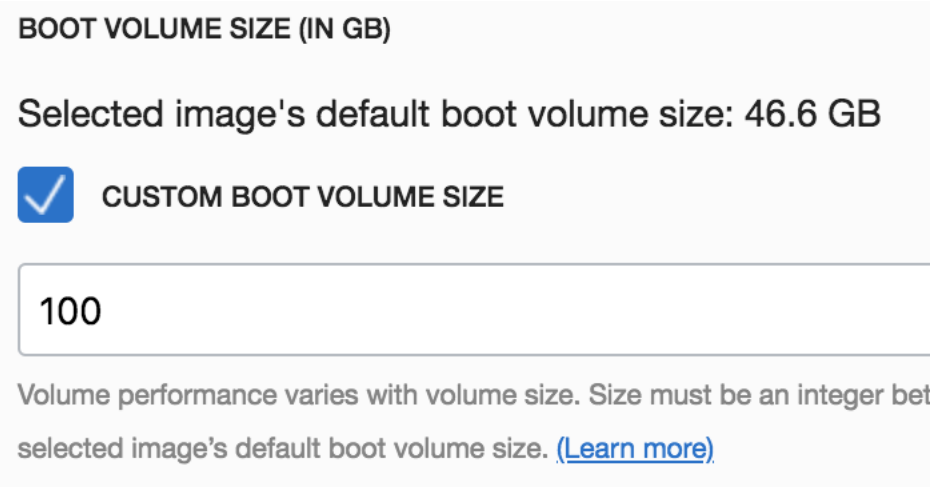
- Group together **block** and **boot volumes** from multiple compartments across multiple compute instances in a volume group
- You can use volume groups to create volume group **backups** and **clones** that are point-in-time and crash-consistent
- Manually trigger a **full or incremental backup** of all the volumes in a volume group leveraging a coordinated snapshot across all the volumes

Boot Volumes

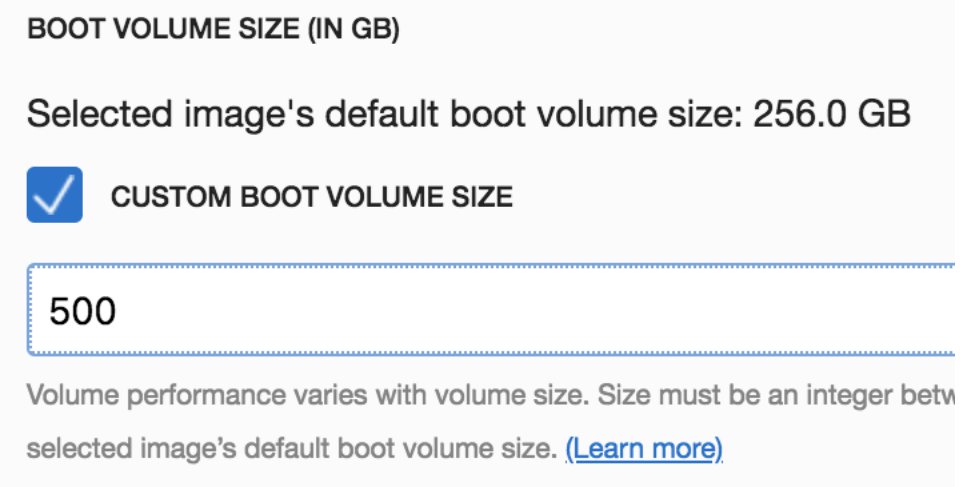
- A compute instance is launched using OS image stored on a remote boot volume
- Boot volume is created automated and associated with an instance until you terminate the instance
- Boot volumes are encrypted, have faster performance, lower launch times, and higher durability for BM and VM instances
- Compute instance can be scaled to a larger shape by using boot volumes
- You can preserve the boot volume when you terminate a compute instance
- Boot volumes are only terminated when you manually delete them
- Boot volumes cannot be detached from a running instance
- Possible to take a manual backup, assign backup policy or create clone of boot volumes

Custom Boot Volumes

- You have the option of specifying a custom boot volume size
- In order to take advantage of the larger size, you must first extend the root (Linux-based images) or system (Windows-based images) partition



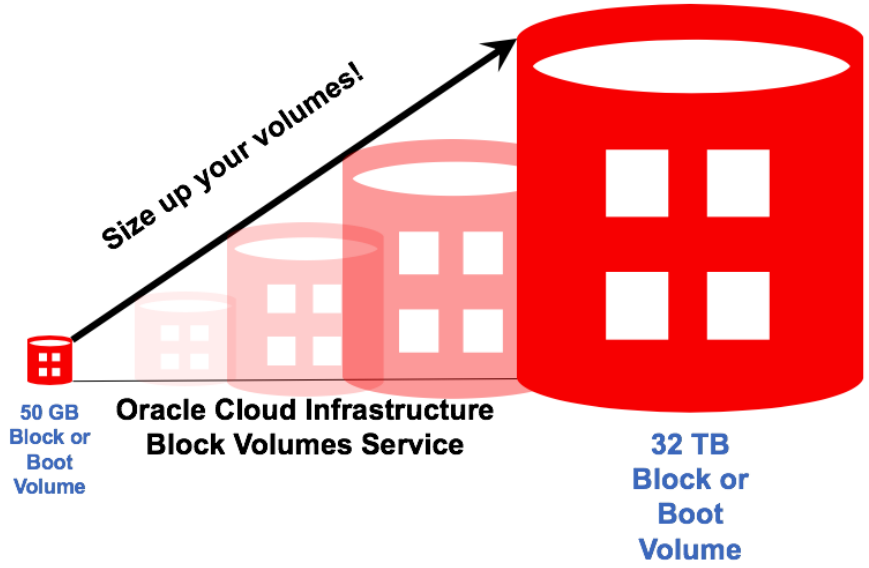
Linux default size is 46.6 GB



Windows default size is 256GB

Block Volume Offline Resize

The Oracle Cloud Infrastructure Block Volume service lets you expand the size of block volumes and boot volumes. You have three options to increase the size of your volumes:



- Expand an existing volume in place with offline resizing.
- Restore from a volume backup to a larger volume.
- Clone an existing volume to a new, larger volume.

You can only increase the size of the volume, **you cannot decrease the size**.

Block Volume Demo

Pricing

Block volumes are metered based on provisioned GB volume size

	Metric	Pay as You Go	Monthly Flex
Block volumes provisioned storage	GB/month	$0.0425	$0.0425

- Typical Enterprise Workload - 400GB Database
- OCI Block Volume: $17 per month w/ 25,000 IOPS (400 x $.0425)
- 25% better performance; 60 IOPS/GB vs 50 IOPS/GB for Amazon EBS
- AWS: You need to pay $0.125/GB-month for storage ($50) and $0.065 per IOPS provisioned ($1,300 for 20,000 IOPS) = $1,350 per month
- 79X cheaper than Amazon EBS

Summary

- Offers local NVMe SSD storage with SLAs for high-performance workloads
- Block Volume service - persistent, durable, high-performance block service with industry leading price/performance
- Create, attach, connect, and move volumes, as needed, to meet your storage and application requirements
- Block volume service supports backups (on-demand, automated) and restoration and
- Cloning and automated backups offered only by OCI Block Volume service
- Another unique feature, Volume Groups simplifies backups of running enterprise applications that span multiple storage volumes across multiple instances

ORACLE®
Cloud Infrastructure

cloud.oracle.com/iaas

cloud.oracle.com/tryit