

ORACLE

Block Volume

L100

Rohit Rahi

Oracle Cloud Infrastructure

Oct 2019

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.

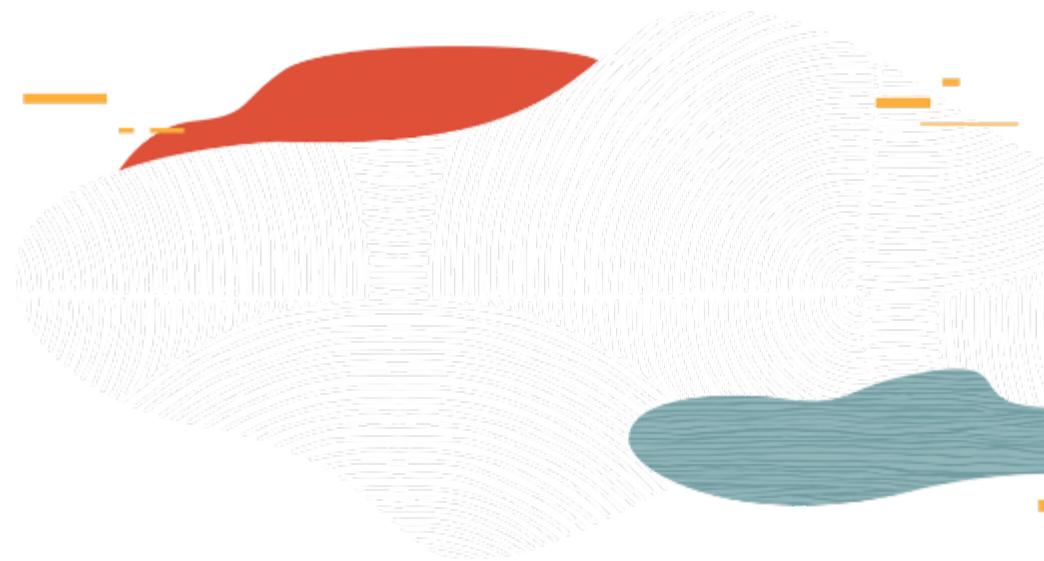


OCI Storage Services

	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 - 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)

ORACLE

Local NVMe Storage



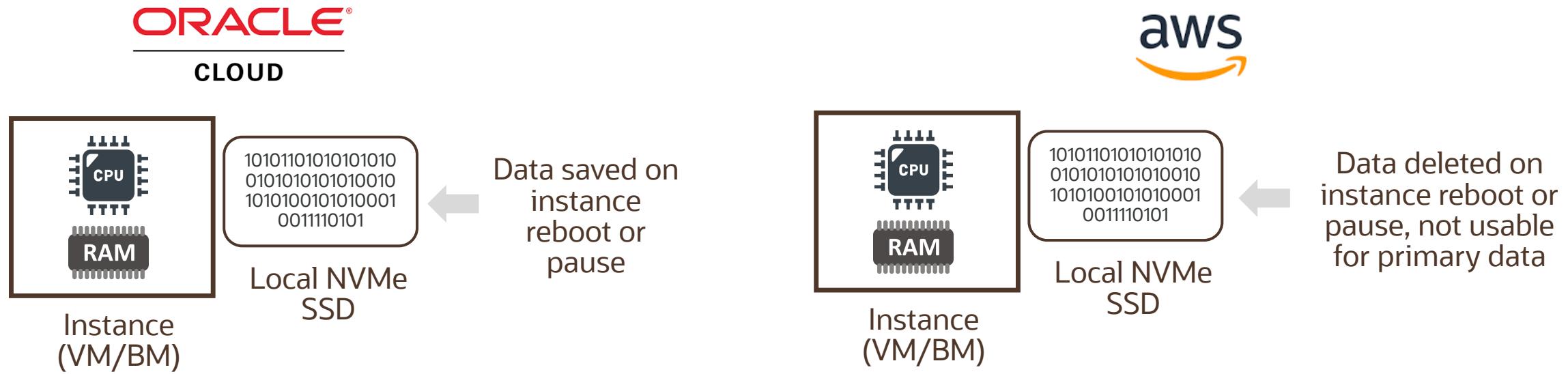
O

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	<pre>[opc@nvme ~]\$ lsblk</pre>
VM.DenseIO2.8	2 drive = 6.4 TB raw	<pre>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</pre>
VM.DenseIO2.16	4 drives = 12.8 TB raw	
VM.DenseIO2.24	8 drives = 25.6 TB raw	

NVMe SSD Persisted - Reboot/Pause



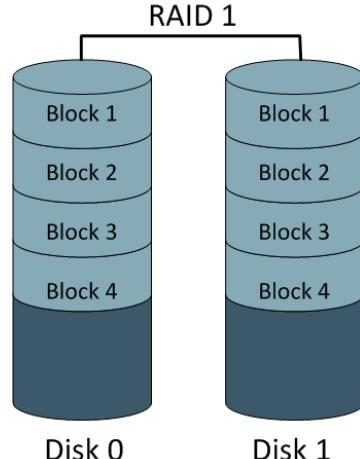
“With Oracle Cloud Infrastructure, companies can leverage NVMe for persistent storage to host databases and applications. However, other cloud providers typically do not offer such a capability. In cases where NVMe storage was an option with other vendors, it was not persistent. This meant that the multi-terabyte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

Accenture

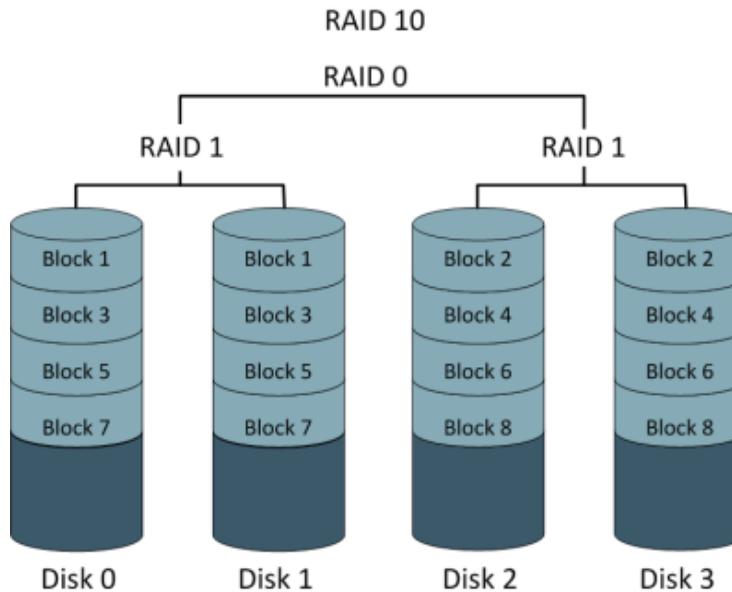


Protecting NVMe SSD Devices

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

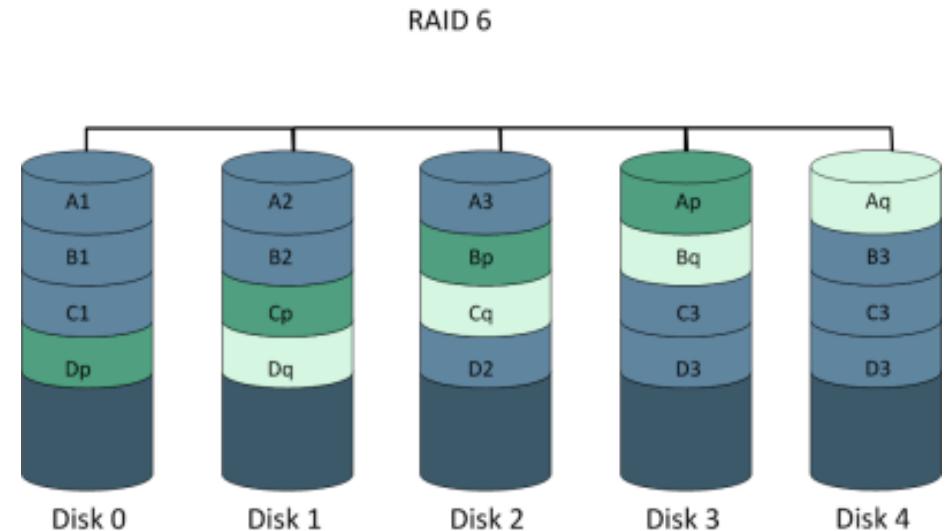


RAID 10: Stripes data across multiple mirrored pairs. As long as one disk in each mirrored pair is functional, data can be retrieved



Blocks mirrored and striped

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



Blocks striped with dual parity across drives

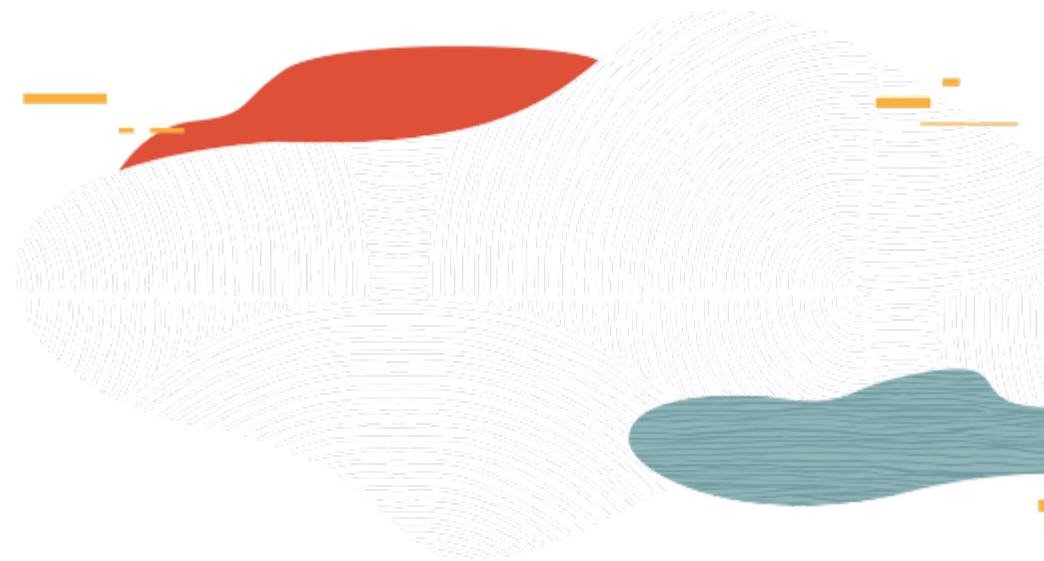
SLA for NVMe Performance

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/cloudbanana/block-storage>

ORACLE

Block Volume Intro



O

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	<ul style="list-style-type: none">• 32 attachments/instance, up to 1 PB (32 TB/volume x 32 volumes/instance)• Up to 620K or more IOPS, near line rate throughout.
Durability	Multiple replicas across multiple storage servers within the AD
Security	Encrypted at rest and transit

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

** 256 KB block size

Creating and Attaching a Block Volume

Create Block Volume

NAME

CREATE IN COMPARTMENT
intoraclerohit (root)

AVAILABILITY DOMAIN
dKYS:US-ASHBURN-AD-1

SIZE (IN GB)
1024
Size must be between 50 GB and 32,768 GB (32 TB). Volume performance

COMPARTMENT FOR BACKUP POLICIES
intoraclerohit (root)

BACKUP POLICY i
 Select a Backup Policy

ENCRYPTION
 ENCRYPT USING ORACLE-MANAGED KEYS
Leaves all encryption-related matters to Oracle.

ENCRYPT USING CUSTOMER-MANAGED KEYS
Requires you to have access to a valid Key Management key.

Attach Block Volume

Choose how you want to attach your block volume.

ISCSI
 PARAVIRTUALIZED

ACCESS

READ/WRITE - SINGLE INSTANCE
Configures the volume attachment as read/write, not shared with other instances.

READ/WRITE - MULTIPLE INSTANCES
Configures the volume attachment as read/write, shared with other instances.

READ-ONLY
Select to configure the volume attachment as read-only, enabling faster performance.

SELECT VOLUME ENTER VOLUME OCID

BLOCK VOLUME COMPARTMENT
intoraclerohit (root)

BLOCK VOLUME
backupBV

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

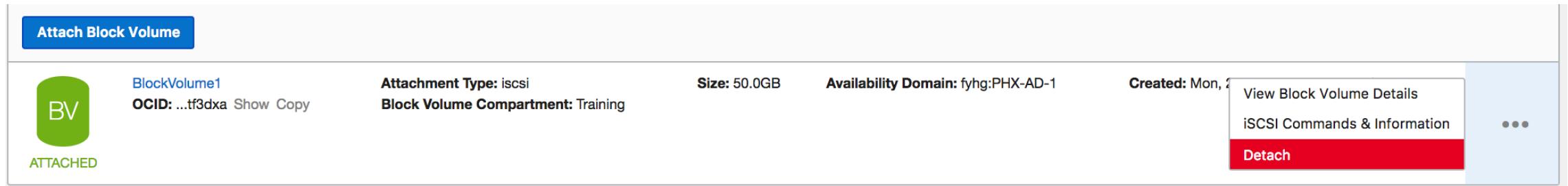
By default, all Block Volumes are Read/Write

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



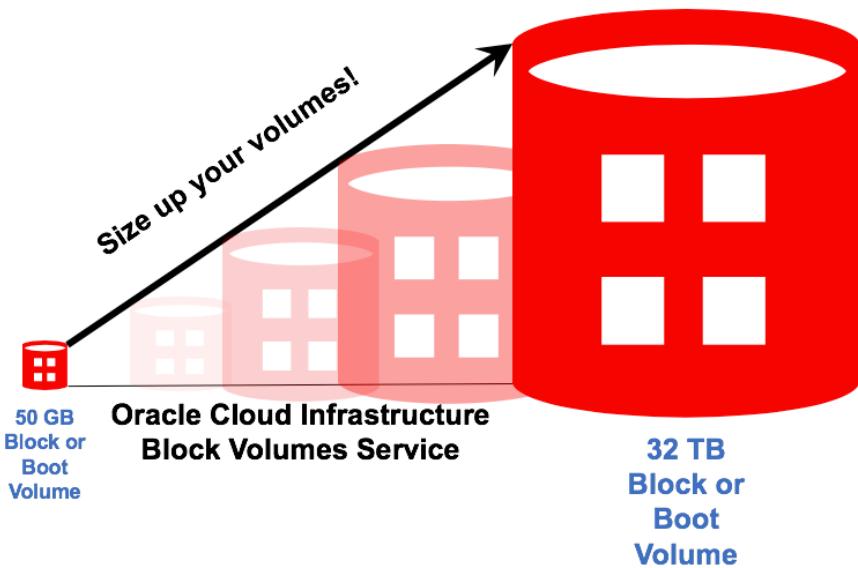
Detaching and Deleting Block Volumes

- 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



Block Volume Offline Resize

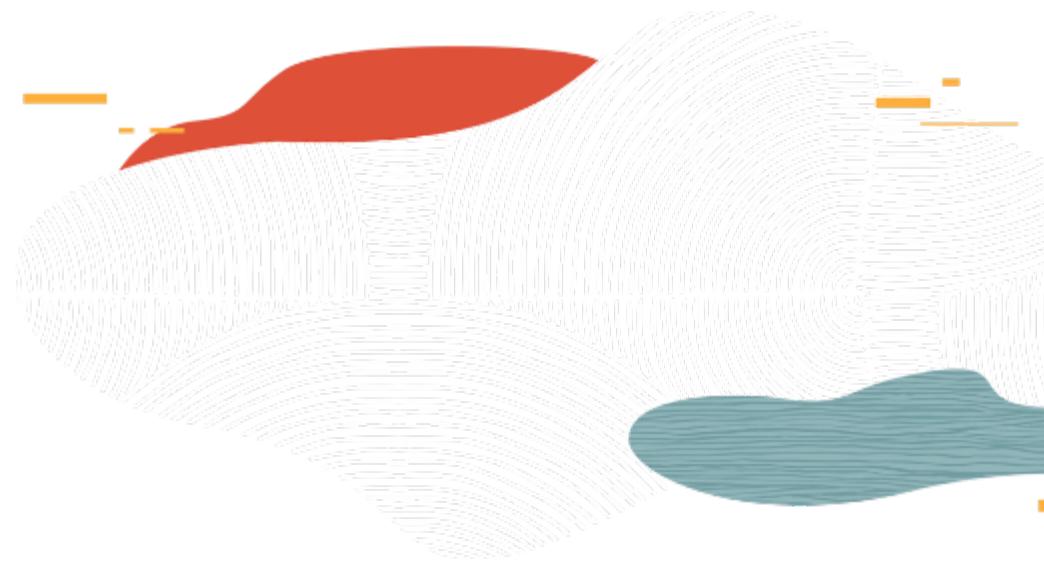
The Oracle Cloud Infrastructure Block Volume service lets you expand the size of block volumes and boot volumes. You have three options to increase the size of your volumes:



- Expand an existing volume in place with offline resizing (cannot resize an attached volume)
- Restore from a volume backup to a larger volume.
- Clone an existing volume to a new, larger volume.

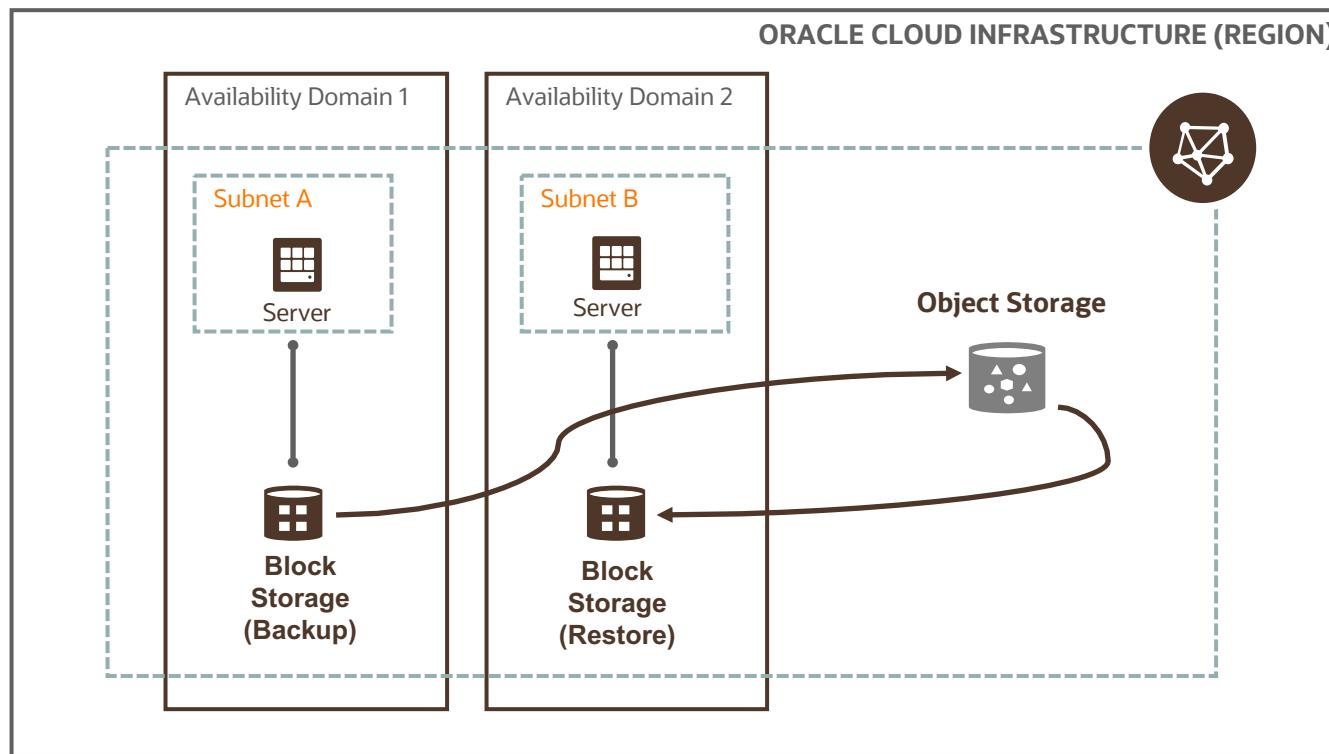
You can only increase the size of the volume, **you cannot decrease the size**

Backup and Restoration



Backup and Restoration

- Complete point-in-time 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 (for multi-AD regions)
- Can copy block volume backups from one-region to another



Backup and Restoration

- Backups are done using point-in-time snapshot; therefore, while the backup is being performed in the background asynchronously, your applications can continue to access your data without any interruption or performance impact
 - For a 2 TB volume being backed up for the first time, ~30 mins
 - For a 50 GB boot volume being backed up for the first time, ~ few mins
- On-demand, one-off block volume backups provide a choice of incremental versus full backup options

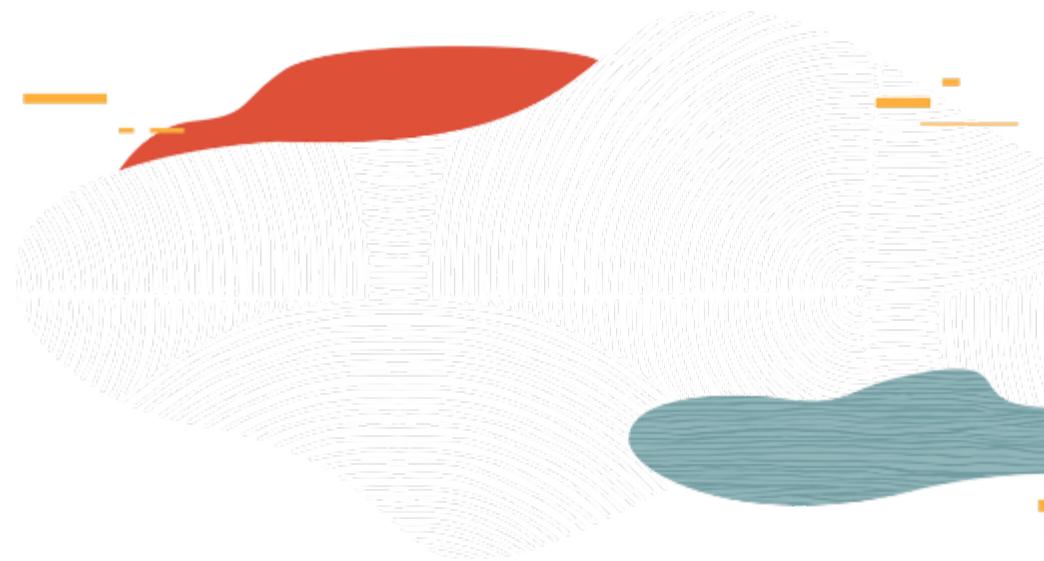
Backup and Restoration

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)
- Customized backup policy not available today

ORACLE

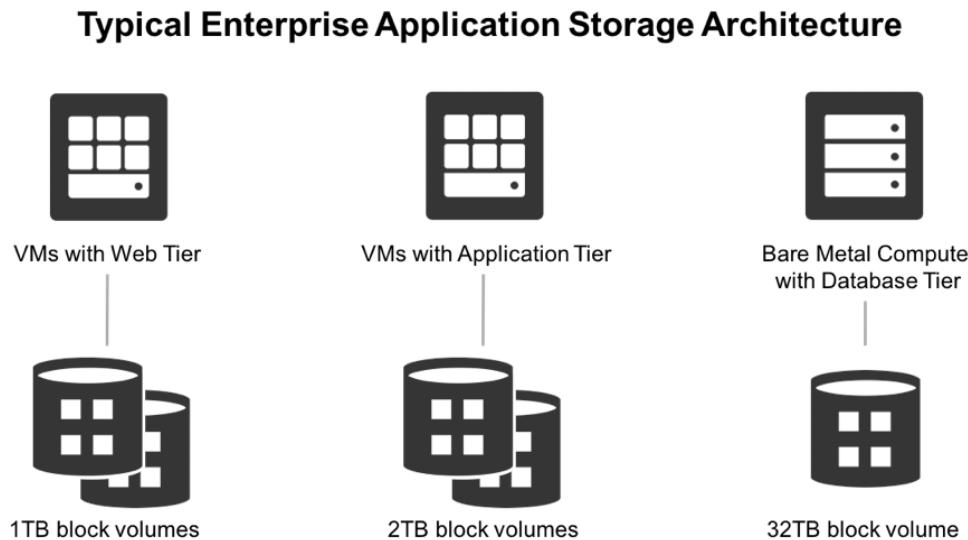
Clone and Volume Groups



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 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
- Clones cannot be copied to another region
- A clone can be attached and used as regular volume when its lifecycle state changes from "PROVISIONING" to "AVAILABLE", usually within seconds
- Clone and backup operations are mutually exclusive
- Number of clones created simultaneously
 - If the source volume is attached: You can create one clone at a time
 - If the source volume is detached: You can create up to 10 clones from the same source volume simultaneously

Volume Groups

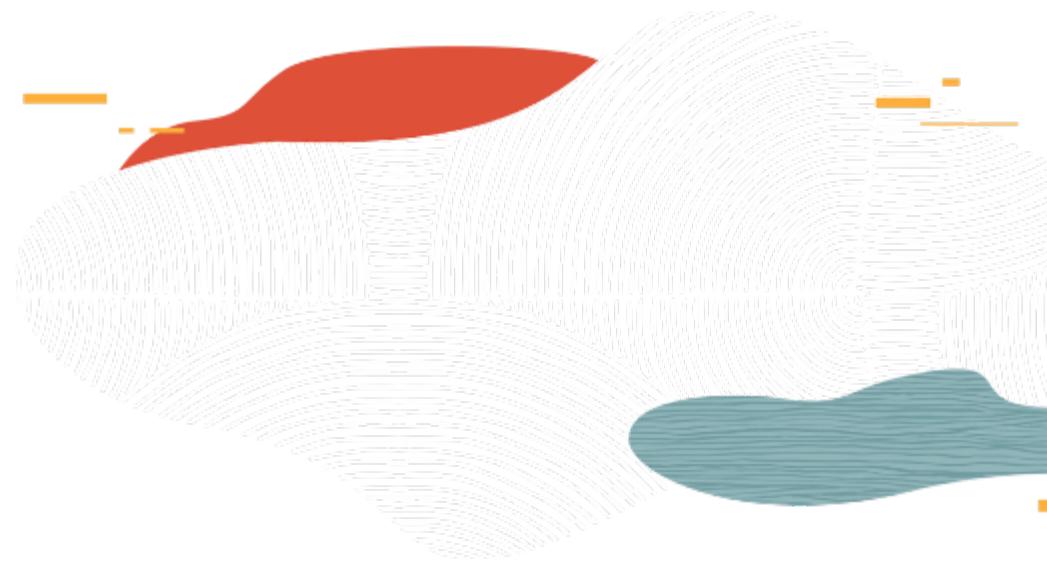


- Group together block and boot volumes from multiple compartments across multiple compute instances in a volume group.
- You can use volume groups to create volume group backups and clones that are point-in-time and crash-consistent.
- Manually trigger a full or incremental backup of all the volumes in a volume group leveraging a coordinated snapshot across all the volumes.
- This is ideal for the protection and lifecycle management of enterprise applications, which typically require multiple volumes across multiple compute instances to function effectively
- Volume Group feature is available with no additional charge



ORACLE

Boot Volumes



O

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
- Launch another instance with a boot volume
 - First create a custom image of your boot volume and then using the custom image launch the instance
 - Alternately, you can launch a new instance directly from an unattached boot volume if you don't wish to create a custom image
- Delete boot volume
 - You can delete an unattached boot volume.
 - You can optionally chose to automatically delete the boot volume when terminating an instance by selecting the checkbox in the delete confirmation dialog.
 - OCI does not allow you to delete the boot volume currently attached to an instance.
- Possible to take a manual backup, assign backup policy or create clone of boot volumes

Boot Volumes

- Attach a Boot Volume to an instance as a block volume for troubleshooting
 - You can attach any boot volume to an instance as block storage in order to debug issues. You will first need to detach a boot volume from its associated compute instance in order to attach it to a different instance.
 - You can follow the steps below to debug your boot volume:
 - 'Stop' the instance you want to debug and click on 'Boot Volume' filter, and then select the 'Detach Boot Volume' button. Alternately, you can terminate your instance which persists your boot volume by default.
 - Navigate to a new running instance you want to use to debug your boot volume, and click the 'Attach Block Volume' button.

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

BOOT VOLUME SIZE (IN GB)

Selected image's default boot volume size: 46.6 GB



CUSTOM BOOT VOLUME SIZE

100

Volume performance varies with volume size. Size must be an integer between selected image's default boot volume size. ([Learn more](#))

Linux default size is 46.6 GB

BOOT VOLUME SIZE (IN GB)

Selected image's default boot volume size: 256.0 GB



CUSTOM BOOT VOLUME SIZE

500

Volume performance varies with volume size. Size must be an integer between selected image's default boot volume size. ([Learn more](#))

Windows default size is 256GB



Summary

- OCI offers local NVMe SSD storage with SLAs for high-performance workloads
- OCI 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, Policy based) and restoration
- Cloning and Policy based 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 always free tier:
oracle.com/cloud/free/

OCI training and certification:
<https://www.oracle.com/cloud/iaas/training/>
<https://www.oracle.com/cloud/iaas/training/certification.html>
education.oracle.com/oracle-certification-path/pFamily_647

OCI hands-on labs:
ocitraining.qloudable.com/provider/oracle

Oracle learning library videos on YouTube:
youtube.com/user/OracleLearning

Thank you

