# 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.

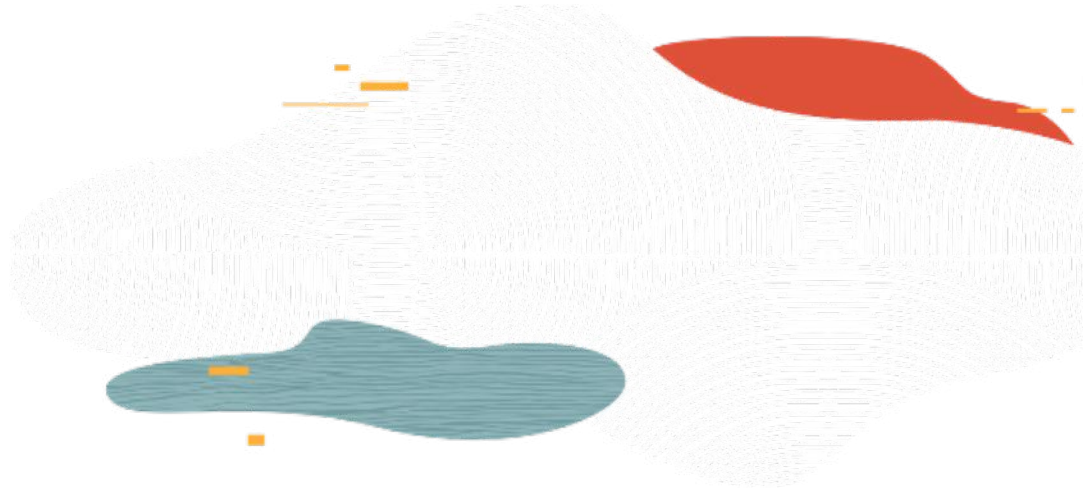
The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

ORACLE

# Storage

## L200

Flavio Pereira
Oracle Cloud Infrastructure
October, 2019



# Objectives

- Explain local NVMe SSD devices
- Block Storage Volume Groups and Performance
- File Storage Service Performance
- EBS Reference Architecture with FSS

ORACLE

# Local NVMe Devices

# Local NVMe Storage

**Local NVMe**

## Overview

- Direct access and full control of local storage using high performance NVMe drives.
- Best choice for workloads with extremely demanding IO such as Big Data and HPC processing
- DenseIO instance shapes include locally attached NVMe devices

## Key Capabilities

- Up to 51.8 TB of local storage
- 3 million IOPS
- <100 microseconds latency

| Instance type | NVMe 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 |

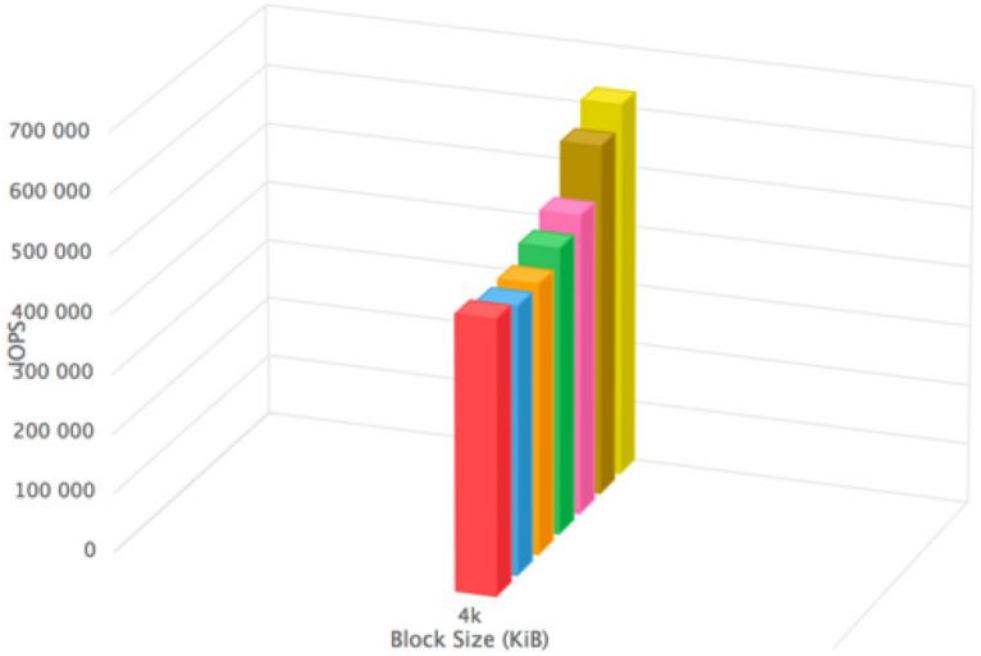
# What to do when an NVMe Device Fails

- If an NVMe device fails while the instance is in service, you should start another instance with the same amount of storage or more and then copy the data onto the new instance, replacing the old instance.
- There are multiple toolsets for copying large amounts of data, with rsync being the most popular. Since the connectivity between instances is a full 10Gb/sec, copying data should be quick.
- Remember that with a failed device, your array may no longer be protected, so you should copy the data off of the impacted instance as quickly as possible.
- Oracle Cloud Infrastructure does not take images, back up, or use RAID or any other methods to protect the data on NVMe devices. It is your responsibility to protect and manage the durability the data on these devices.

# NVMe Performance

Using a Bare Metal shape with 52 CPUs- BM.DenseIO2.52 the following command was executed following the cloud harmony test suite and we are getting **~500K IOPS** (50/50) on a Read and Write Mix test for **a single NVMe device.**

```
# run.sh --target=/dev/nvme1n1 --test=iops --nopurge --noprecondition --fio_direct=1 --fio_size=10g --
skip_blocksize 512b --skip_blocksize 1m --skip_blocksize 8k --skip_blocksize 16k --skip_blocksize 32k --
skip_blocksize 64k --skip_blocksize 128k
```

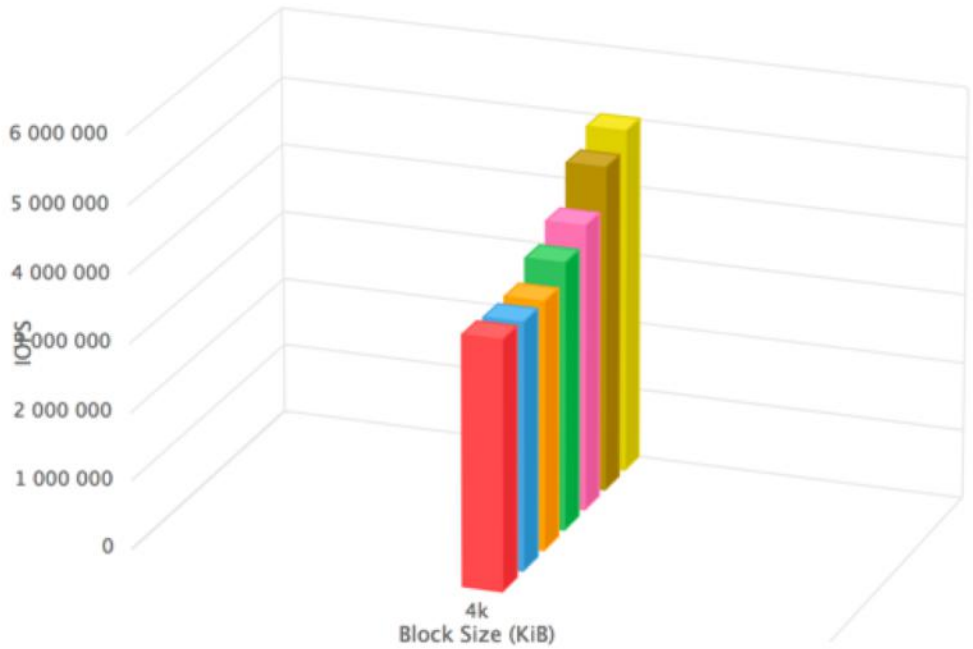



| Block Size (KiB) | Read / Write Mix % | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0/100 | 5/95 | 35/65 | 50/50 | 65/35 | 95/5 | 100/0 |
| 4k | 460985.8 | 449328.4 | 453960.6 | 480391.2 | 503803.8 | 588040.6 | 627179.6 |

# NVMe Performance

Using a Bare Metal shape with 52 CPUs- BM.DenseIO2.52 the following command was executed following the cloud harmony test suite and we are getting **~ 3.0MM** IOPS (50/50) on a Read and Write Mix test for **all NVMe devices combined**

```
# run.sh `ls /dev/nvme[0-9]n1 | sed -e 's/\//\--target=\//'` --test=iops --nopurge --noprecondition --
fio_direct=1 --fio_size=10g --skip_blocksize 512b --skip_blocksize 1m --skip_blocksize 8k --skip_blocksize
16k --skip_blocksize 32k --skip_blocksize 64k --skip_blocksize 128k
```





| Block Size (KiB) | Read / Write Mix % | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0/100 | 5/95 | 35/65 | 50/50 | 65/35 | 95/5 | 100/0 |
| 4k | 3648650 | 3611607.2 | 3631868.4 | 3914589.8 | 4178064.6 | 4744123.8 | 5016890.2 |

# 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

SLA: https://cloud.oracle.com/en_US/iaas/sla

ORACLE

# Demo: NVMe Devices

ORACLE

# Block Storage

# Block Storage

**Block Storage**

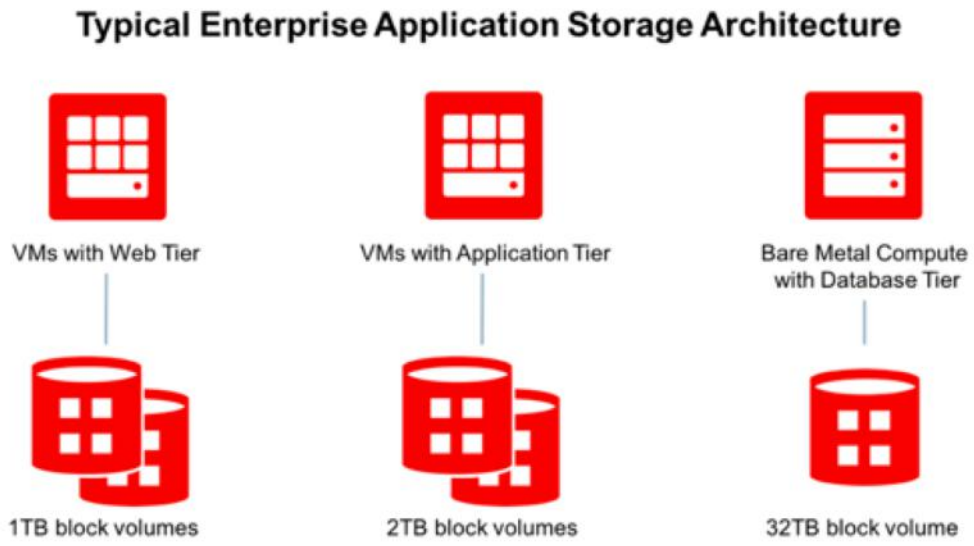
## Overview

- Volume sizes goes from **50GB to 32TB** in 1GB increments. You can attach up to **32 volumes on each instance**
- **Scheduled Backups** – set it and forget it
- **Cross-region Backup Copy** enables you to easily prepare for a disaster or migrate your workload across regions

## Key Capabilities

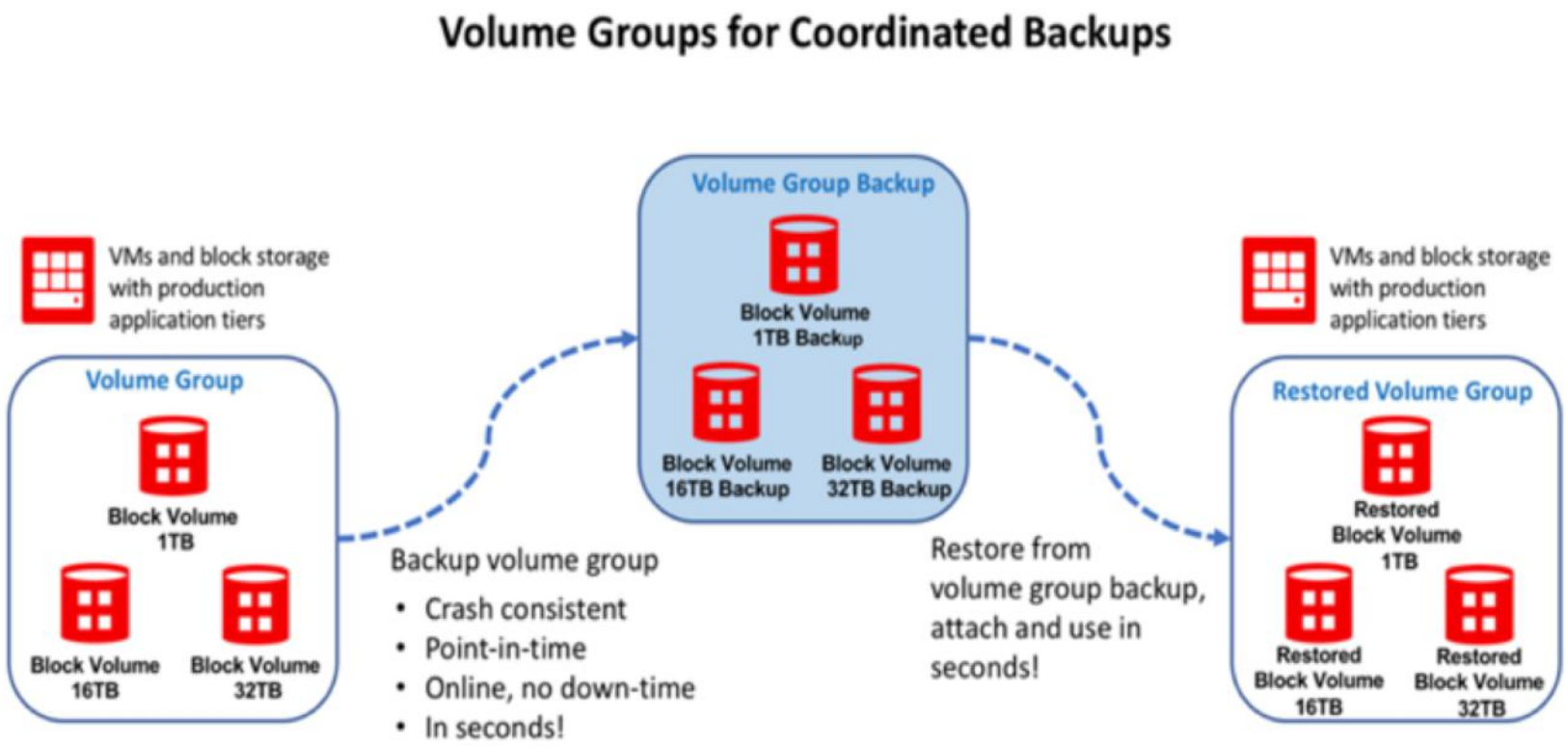
- Industry-leading price-performance
- SAN-like management capabilities
- Scalable to **1 PB and 650,000 IOPS** per Compute instance
- **SLA backed** performance guarantees

# Volume Groups

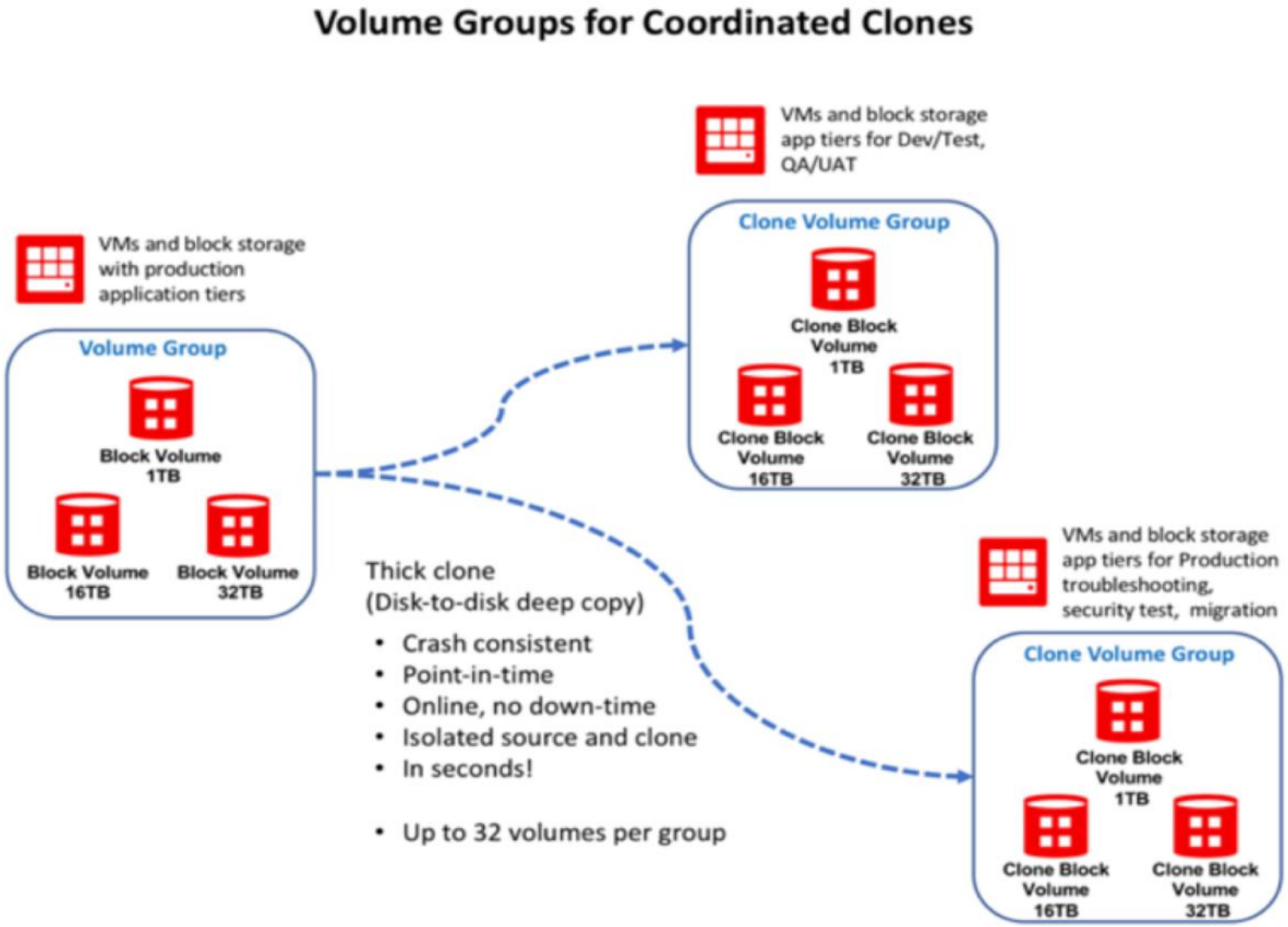



- Group together block and boot volumes from multiple compartments across multiple compute instances in a volume group.
- You can use volume groups to create volume group backups and clones that are point-in-time and crash-consistent.
- Manually trigger a full or incremental backup of all the volumes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



# Volume Groups for Coordinate Clones



# Block Volume Performance

- There are 3 aspects to the block volume performance:
  - IOPS
  - Throughput
  - Latency

The Block Volume service uses NVMe-based storage infrastructure, and is designed for consistency. You just need to provision the capacity needed and **performance scales linearly per GB volume size** up to the service maximums.

**Throughput performance on VM instances is dependent on the network bandwidth that is available to the instance**, and further limited by that bandwidth for the volume.

**IOPS performance is independent of the instance type or shape**, so is applicable to all bare metal and VM shapes, for iSCSI attached volumes.

# Block Volume Performance

The following table describes the performance characteristics of the service:

| Volume Size | Max Throughput (1 MB block size) | Max Throughput (8 KB block size) | Max IOPS (4 KB block size) |
|---|---|---|---|
| 50 GB | 24 MB/s | 24 MB/s | 3,000 |
| 100 GB | 48 MB/s | 48 MB/s | 6,000 |
| 200 GB | 96 MB/s | 96 MB/s | 12,000 |
| 300 GB | 144 MB/s | 144 MB/s | 18,000 |
| 400 GB | 192 MB/s | 192 MB/s | 24,000 |
| 500 GB | 240 MB/s | 200 MB/s | 25,000 |
| 700 GB – 32 TB | 320 MB/s | 200 MB/s | 25,000 |

- The IOPS performance characteristics described in this topic are for volumes with iSCSI attachments.
- Block Volume performance SLA for IOPS per volume and IOPS per instance applies to iSCSI volume attachments only, not to paravirtualized attachments.

# Block Volume Performance

Here is a sample FIO commands we use for throughput measurement on a single volume:

## Read-only:

```
# sudo fio --direct=1 --ioengine=libaio --size=10g --bs=4k --runtime=60 --numjobs=8 --iodepth=64 --time_based --rw=randread --group_reporting --filename=/dev/sdb --name=iops-test
```

## Write-only:

```
# sudo fio --direct=1 --ioengine=libaio --size=10g --bs=4k --runtime=60 --numjobs=8 --iodepth=64 --time_based --rw=randwrite  --group_reporting --filename=/dev/sdb  --name=iops-test
```

## Read/write Mix:

```
# sudo fio --direct=1 --ioengine=libaio --size=10g --bs=4k --runtime=60 --numjobs=8 --iodepth=64 --time_based --rw=randrw --group_reporting --filename=/dev/sdb --name=iops-test
```

**Note:** In read/write case, you need to add the read result and write result for duplex traffic.

Also, please note that all volumes attached to an instance share the same network bandwidth with the instance. If there is heavy network traffic or other volumes are under I/O pressure, the apparent performance of single volume may look degraded

# Block Volume Performance

Our block volume service performance measurement methodology and characteristics are covered in detail here:

https://blogs.oracle.com/cloud-infrastructure/block-volume-performance-analysis

https://docs.cloud.oracle.com/iaas/Content/Block/Concepts/blockvolumeperformance.htm

http://www.storagereview.com/oracle_cloud_infrastructure_compute_bare_metal_instances_review

## WARNING:

- Before running any tests, protect your data by making a backup of your data and operating system environment to prevent any data loss.
- Do not run FIO tests directly against a device that is already in use, such as /dev/sdX. If it is in use as a formatted disk and there is data on it, running FIO with a write workload (readwrite, randrw, write, trimwrite) will overwrite the data on the disk, and cause data corruption.
- Run FIO only on unformatted raw devices that are not in use.

ORACLE

# Demo: Block Storage

ORACLE

# File Storage

# File Storage

**File Storage**

## Overview

- Enterprise-grade **shared file system** for business applications
- Provides network-attached storage (NAS) in the Cloud that is management-free

## Key Capabilities

- Exabyte scale
- No need to provision, pay and scale as you go
- Easy snapshotting
- NFSv3 Support with Linux and Windows Compatibility

# File Storage Performance

| File System | Read Bandwidth (1MB Blocks) | Write Bandwidth (1MB Blocks) | Read IOPS (8K Blocks) |
|---|---|---|---|
| 1 TB | 100 MB/s | 50 MB/s | 2,500 |
| 10 TB | 1 GB/s | 500 MB/s | 25,000 |
| 100 TB | 10 GB/s | 5GB/s | 250,000 |

This table describes the level of performance expected for different size file systems. Although we do not guarantee these numbers, we expect that customers should be able to achieve at least this level of performance in the current File Storage service.

More details you can find on File Storage Performance Guide:
https://docs.cloud.oracle.com/iaas/Content/Resources/Assets/whitepapers/file-storage-performance-guide.pdf

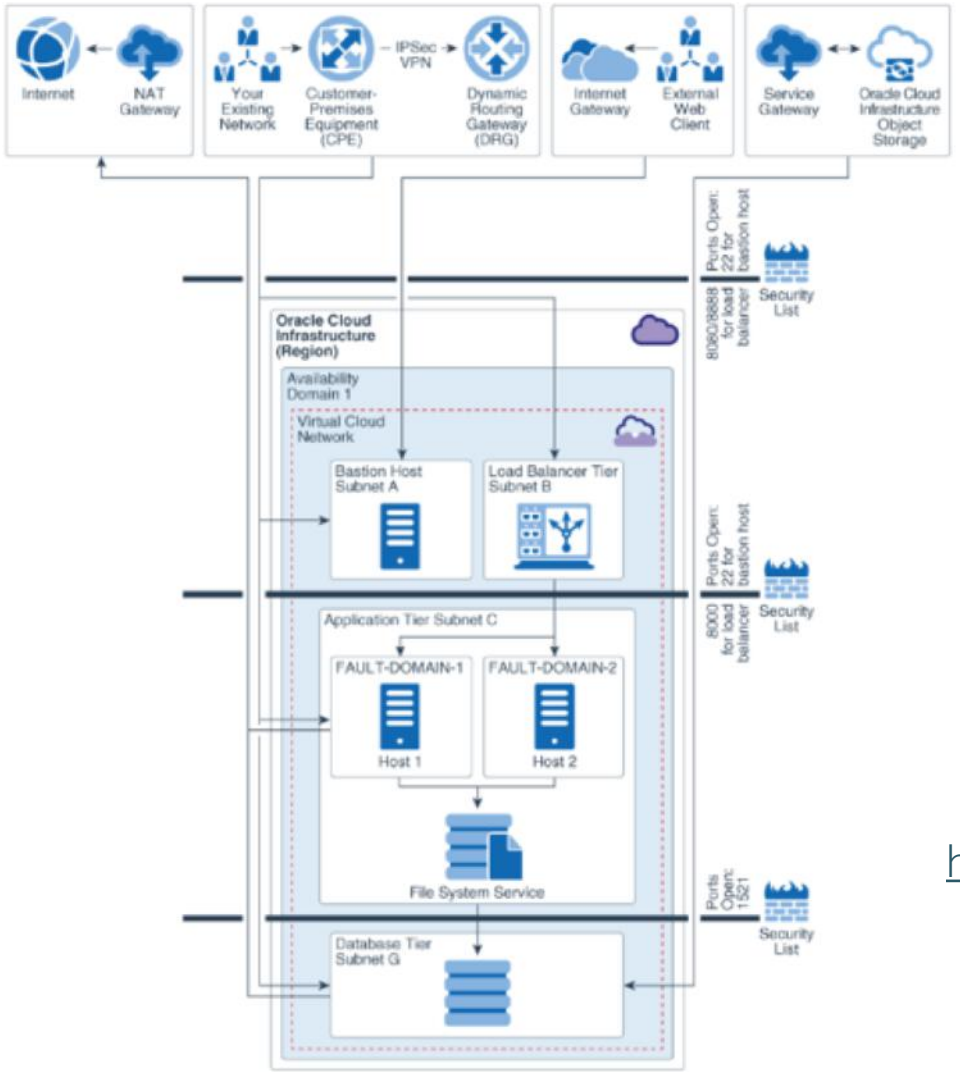
# File Storage Performance

To optimize the performance of File Storage, consider the following guidelines:

- File Storage scales based on consumed capacity, with larger file systems receiving more available bandwidth.
- File Storage performance increases with parallelism. Increase concurrency by using multiple threads, multiple clients, and multiple mount targets.
- To minimize latency, clients and file systems should be in the same availability domain
- For best performance, don't set any mount options such as rsize or wsize when mounting the file system. The system automatically negotiates optimal window sizes

Parallel File Tools for File Storage
https://blogs.oracle.com/cloud-infrastructure/announcing-parallel-file-tools-for-file-storage

# EBS Reference Architecture – File Storage Service




- When deploying an Oracle E-Business Suite application tier with multiple nodes, you can have either a shared or non-shared application tier file system.
- The Oracle Cloud Infrastructure File Storage Service (FSS) can be used to create the shared disk resource required for an Oracle E-Business Suite application tier file system.

https://www.oracle.com/webfolder/technetwork/tutorials/obe/cloud/compute-iaas/sharing_app_tier_file_system_ebs_fss/24fss.html

ORACLE

# Demo: File Storage

# Summary

- Describe and validate storage performance
- Use volume groups to manage snapshot and cloning activities for logical volumes spanning multiple block volumes
- Describe File Storage Service Performance options
- Understand how to use FSS with EBS application

ORACLE

**Oracle Cloud always free tier**:
oracle.com/cloud/free/

**OCI training and certification**:
oracle.com/cloud/iaas/training
oracle.com/cloud/iaas/training/certification
education.oracle.com/oracle-certification-path

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

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