Oracle Database Best Practice with NeoSapphire Series

Table of Contents

1. Overview

Oracle Database is widely used for online transaction (OLTP) and real-time analytics applications. For these types of applications, I/O response time is most critical, as it impacts the overall productivity of an enterprise.

This document provides information on the reference architecture, implementation, and recommendations for deployment of Oracle Database with the AccelStor NeoSapphire all-flash array, using the NeoSapphire 3605 as an example. Employing the HammerDB benchmark tool, the TPM capability of the NeoSapphire all-flash array as compared with other existing storage arrays is clearly illustrated.

2. Goal

To provide an industry-standard benchmark for Oracle Database as greatly improved when used with the NeoSapphire all-flash array series.

3. Introduction: FlexiRemap Technology

FlexiRemap technology is flash-oriented software developed by AccelStor. This software features an architectural re-design tailored for the fundamentals and internal workings of flash memory. Unlike RAID algorithms, FlexiRemap remaps the data whenever beneficial before passing it to the underlying flash memory, avoiding unnecessary overhead and extending the lifespan of SSDs. Such remapping is especially crucial for random-write access patterns, which comprise most enterprise-data traffic and are a major cause of storage-access performance bottlenecks.

FlexiRemap features an exclusive data remapping algorithm based on the physical nature of NAND flash memory. Instead of writing the new data to the original location, FlexiRemap organizes the inbound writes to sequential logical addresses. 4KB random write performance is significantly improved, and total performance is dramatically enhanced for 1,020K IOPS as compared with RAID algorithm, which only generates around 5K IOPS to 78K IOPS (Figure 1).
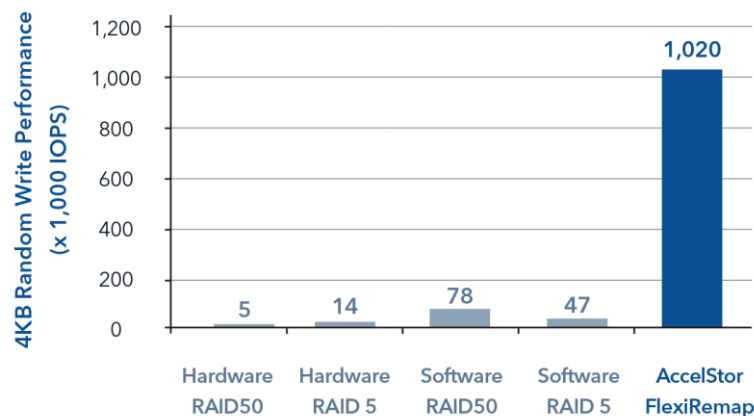

Figure 1: FlexiRemap vs. RAID – Performance in IOPS.

4. Reference Architecture

The test environment is listed below:

Server

- ☐ 2* Intel CPU E5-2670 v2 (25M Cache, 3.30 GHz, 8 cores, 16 Threads) http://ark.intel.com/products/75273/Intel-Xeon-Processor-E5-2667-v2-25M-Cache-3_30-GHz
- ☐ 128GB RAM
- ☐ Emulex Fibre Channel HBA (16G port x 2)
- ☐ Oracle Linux Server release 7.2 http://www.oracle.com/technetwork/server-storage/linux/downloads
- ☐ Oracle Database 12.1.0.2 64 bits
- ☐ HammerDB Release 2.20 for Linux 64 bit http://www.hammerdb.com/

NeoSapphire all-flash array 3605

- ☐ 1U rackmount design with 10 hot-swappable SSD
- ☐ 360K sustained IOPS for 4KB random write
- ☐ 5TB usable capacity
- ☐ 2 x 16G Fibre Channel LC SFP+
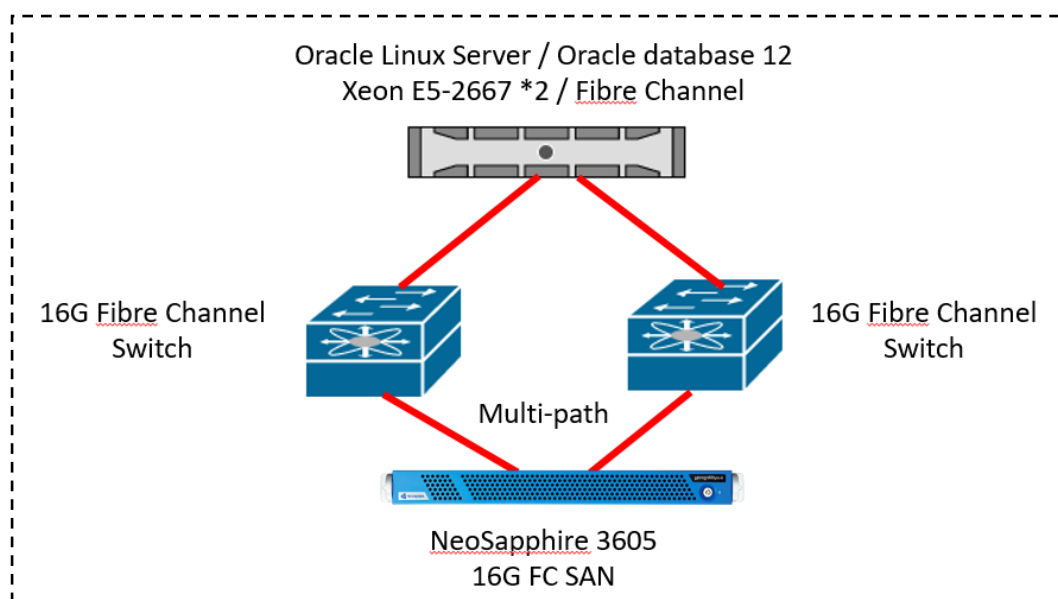- ☐ System image version: 1.1.1(Build 3532)

Figure 2: Topology

5. LUN Configuration

Please note that LUN size in a disk group should be the same.

Disk group name	Number of LUNs	Size	content
u01	1	50GB	Oracle software
DATA	2	500GB	Table data and indexes
REDO	1	200GB	Online Redo log
TEMP	1	100GB	Temporary Tablespace
FRA	2	500GB	Flash Recovery Area

6. Multipath Configuration

In SAN storage for Oracle Database, it is recommended that a multipath design be implemented to ensure fault tolerance and performance enhancement. In addition to better performance, the multipath connection also avoids a single point of failure. If one path is lost, the remaining paths still support I/O operation without interruption. Specify the LUN wwid in the file multipath.conf and assign an alias name for easy manageability. An example of the content of a multipath configuration is indicated below. Follow the above format and add all LUNs to the file multipath.conf.

Before setting up the DM-Multipath on your system, make sure the device-mapper-multipath package has been installed.

```
defaults {
 user_friendly_names yes
 find_multipaths yes
}
multipaths{
 multipath{
        wwid                 234074f14773b8cec   #device id
        alias                u01
        path_grouping_policy  multibus
        path_selector        "round-robin 0"
        failback             immediates
        rr_weight            priorities
        no_path_retry        5
     }
}
```

Parameters	Explanation
wwid	World Wide Identifier (WWID) is used to guarantee every device ID is unique.
alias	User-friendly name.
path_grouping_policy	Specifies the default path group policy. multibus = all valid paths in 1 priority group.
path_selector	Specifies the default algorithm to determine what path to use for the next I/O operation.
failback	Manages path group failback.
rr_weight	If set to priorities, the system will determine path weight by "path prio* rr_minio".
no_path_retry	Specifies the number of times the system should try to use a failure path before disabling it.

Once the multipath configuration is complete, restart the multipath service to sync the new change and make it effective.

```
#service multipathd restart
```

You can check the following commands for the new LUN names.

```
#multipath –ll | grep AS
temp (234074f147f99c9a4) dm-6 AStor   ,NeoSapphire 3600
u01 (234074f14773b8cec) dm-4 AStor   ,NeoSapphire 3600
redo (234074f14e6df8463) dm-3 AStor   ,NeoSapphire 3600
fra2 (234074f1442853c96) dm-10 AStor   ,NeoSapphire 3600
fra1 (234074f142e45943b) dm-8 AStor   ,NeoSapphire 3600
data2 (234074f14fd2f55b1) dm-7 AStor   ,NeoSapphire 3600
data1 (234074f14a5ecda3c) dm-5 AStor   ,NeoSapphire 3600
```

7. Create ASM Disk Group

For Oracle ASM using disks, the disks need to be identified with the correct ownership and permissions. ASMLib is used to manage these tasks.

In Linux, device manager "udev" is often used to perform such tasks.

In the file /dev/udev/rules.d/12-dm-permissions.rules, add the new disks into this file and change their owner and group to "grid, asmadmin" as in the example below.

```
bryan — root@astor:/etc/udev/rules.d — ssh root@10.144.1.171 — 121×28
[root@astor rules.d]# cat 12-dm-permissions.rules
ENV{DM_NAME}=="u01", OWNER:="grid", GROUP:="asmadmin", MODE:="660", SYMLINK+="iscsi/oraasm-$env{DM_NAME}"
ENV{DM_NAME}=="temp", OWNER:="grid", GROUP:="asmadmin", MODE:="660", SYMLINK+="iscsi/oraasm-$env{DM_NAME}"
ENV{DM_NAME}=="redo", OWNER:="grid", GROUP:="asmadmin", MODE:="660", SYMLINK+="iscsi/oraasm-$env{DM_NAME}"
ENV{DM_NAME}=="data1", OWNER:="grid", GROUP:="asmadmin", MODE:="660", SYMLINK+="iscsi/oraasm-$env{DM_NAME}"
ENV{DM_NAME}=="data2", OWNER:="grid", GROUP:="asmadmin", MODE:="660", SYMLINK+="iscsi/oraasm-$env{DM_NAME}"
ENV{DM_NAME}=="fra1", OWNER:="grid", GROUP:="asmadmin", MODE:="660", SYMLINK+="iscsi/oraasm-$env{DM_NAME}"
ENV{DM_NAME}=="fra2", OWNER:="grid", GROUP:="asmadmin", MODE:="660", SYMLINK+="iscsi/oraasm-$env{DM_NAME}"
```

For more detailed information about configuring users, groups, and environments, please refer to the Oracle Database Installation Guide.

Reboot to make the change effective. Run the command below to check.

```
[root@astor ~]# ll /dev | grep grid
brw-rw----. 1 grid asmadmin 252,  10 Oct 21 13:20 dm-10
brw-rw----. 1 grid asmadmin 252,   3 Oct 21 13:30 dm-3
brw-rw----. 1 grid asmadmin 252,   4 Oct 21 13:21 dm-4
brw-rw----. 1 grid asmadmin 252,   5 Oct 21 13:30 dm-5
brw-rw----. 1 grid asmadmin 252,   6 Oct 21 13:30 dm-6
brw-rw----. 1 grid asmadmin 252,   7 Oct 21 13:20 dm-7
brw-rw----. 1 grid asmadmin 252,   8 Oct 21 13:30 dm-8
```

● 4KB alignment

On the host side, a non-aligned partition can have a major impact on performance. This can be caused by a partition tool that creates a disk partition not aligned with storage block size. Before creating Oracle ASM disks, all of the new disks need to be partitioned and checked for alignment.

```
#parted /dev/mapper/data1 mklabel gpt mkpart primary 2048s 100% align-check optimal 1
#parted /dev/mapper/data2 mklabel gpt mkpart primary 2048s 100% align-check optimal 1
#parted /dev/mapper/fra1 mklabel gpt mkpart primary 2048s 100% align-check optimal 1
#parted /dev/mapper/fra2 mklabel gpt mkpart primary 2048s 100% align-check optimal 1
#parted /dev/mapper/redo mklabel gpt mkpart primary 2048s 100% align-check optimal 1
#parted /dev/mapper/temp mklabel gpt mkpart primary 2048s 100% align-check optimal 1
```

● Create Oracle ASM disks

Follow the steps below to create all new disks.

```
#oracleasm createdisk data1 /dev/mapper/data1
Writing disk header: done
Instantiating disk : done
```
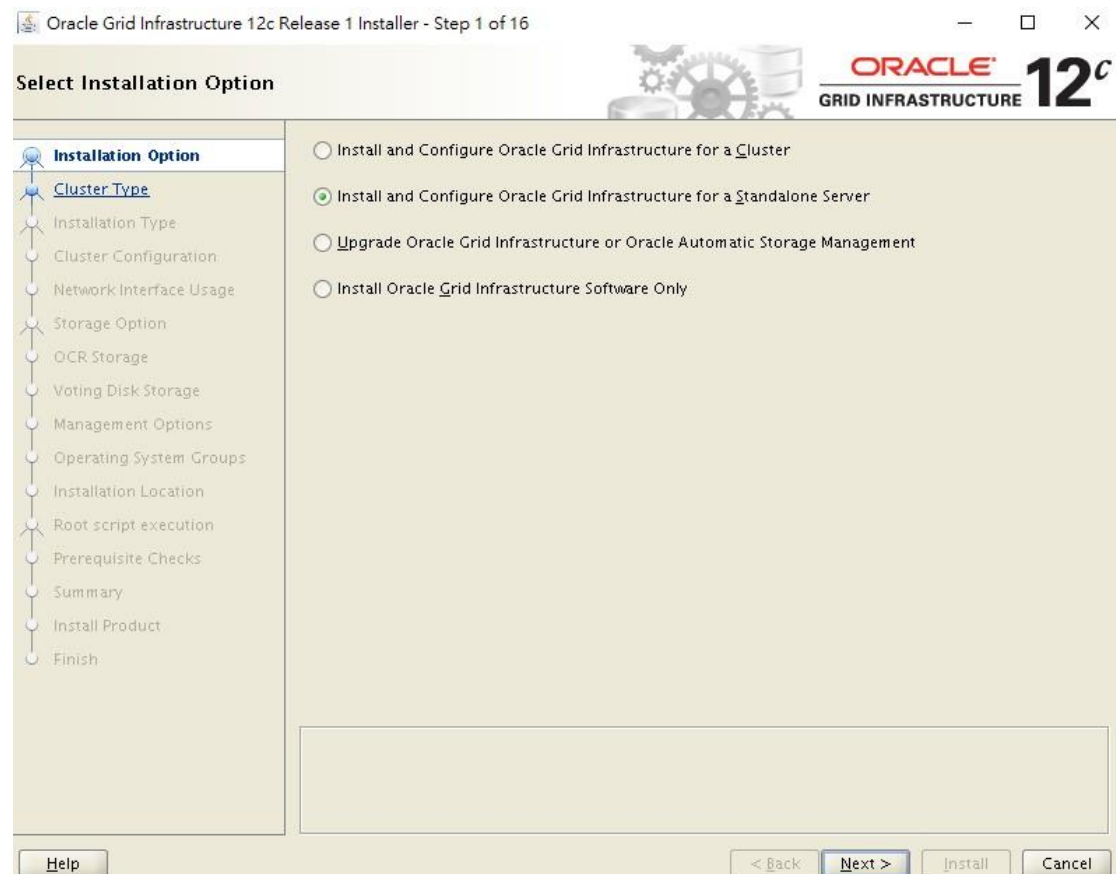
8. Oracle software Installation Location

Before installing the Oracle software, you must format the LUN for a file system. XFS is recommended with parameters as in the example below:
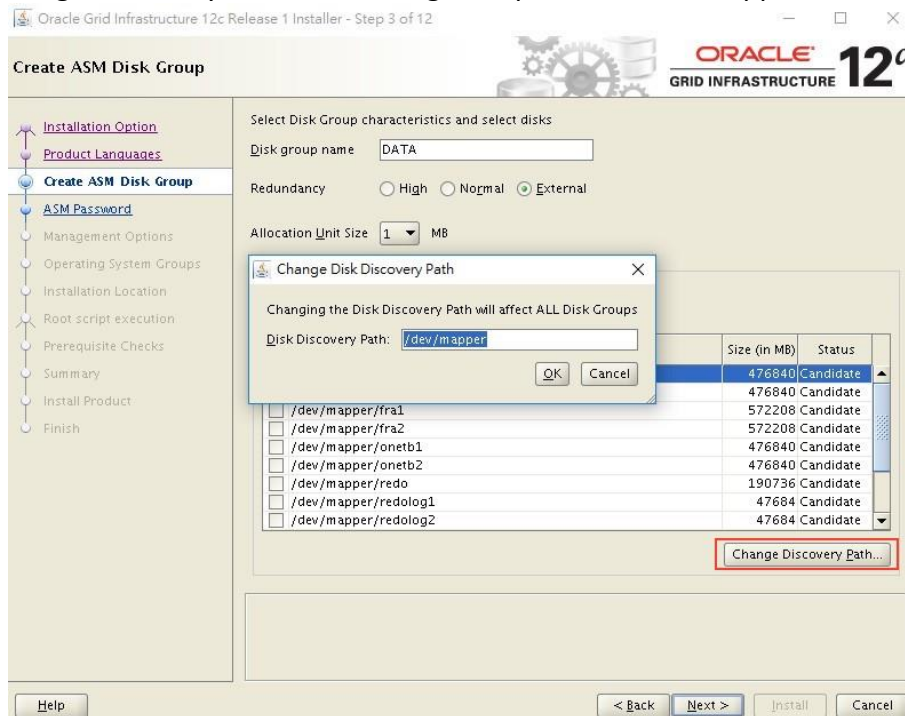

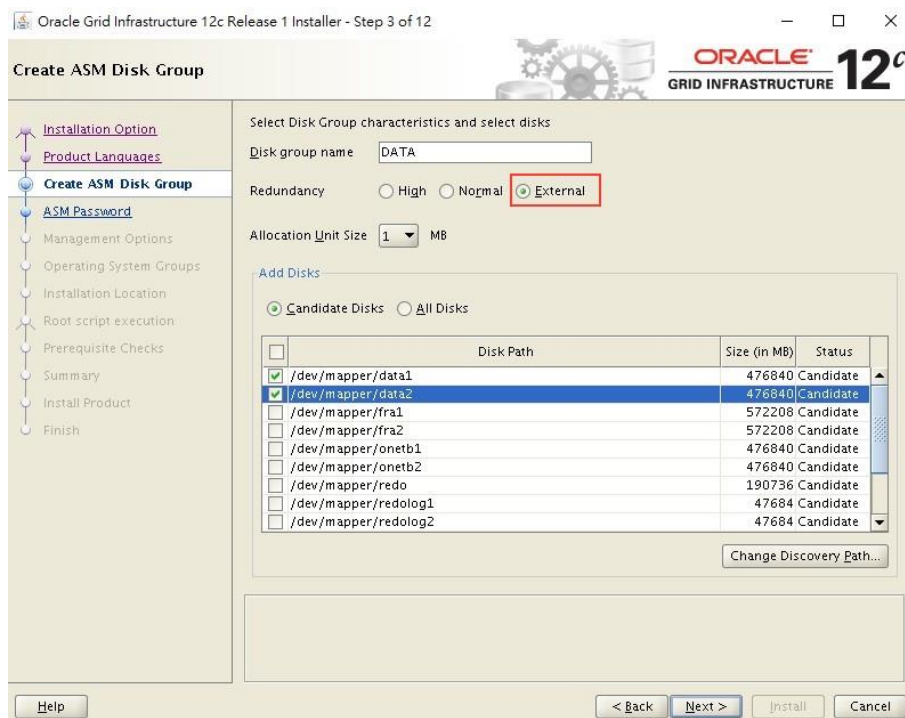## 9. Oracle Grid Infrastructure Installation

Switch to user "grid" and execute "runInstaller" to start installation of Grid Infrastructure.

The first grid infrastructure will create an ASM disk group. The rest of the disk groups can be created with the command "asmca".
Click "Change Discovery Path" and change the path to "/dev/mapper".

Input disk group name: DATA
Redundancy: External
Allocated unit size: 1MB

For more detailed information about Oracle Grid Infrastructure installation, please refer to the Oracle Database Installation Guide.

After Grid Infrastructure installation is complete, execute the command "asmca" by "grid" to create the rest of the disk groups.

Click "Create".
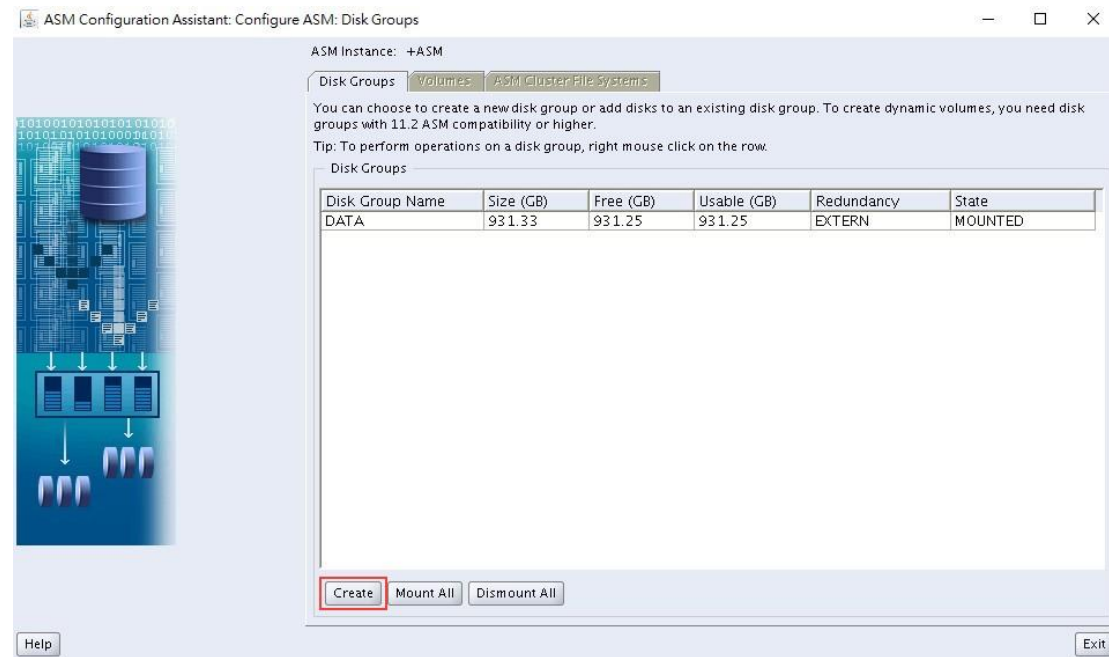
Input disk group name: FRA
Redundancy: External
Choose disk group member
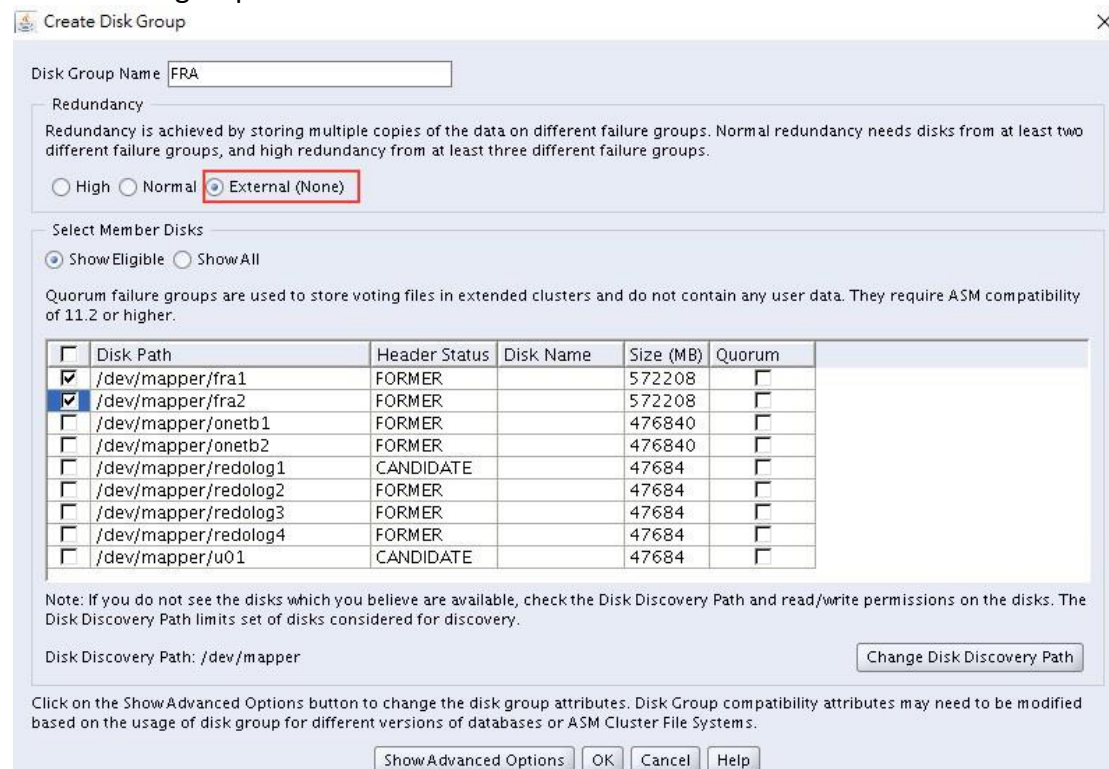
Follow the steps above to complete all disk groups.
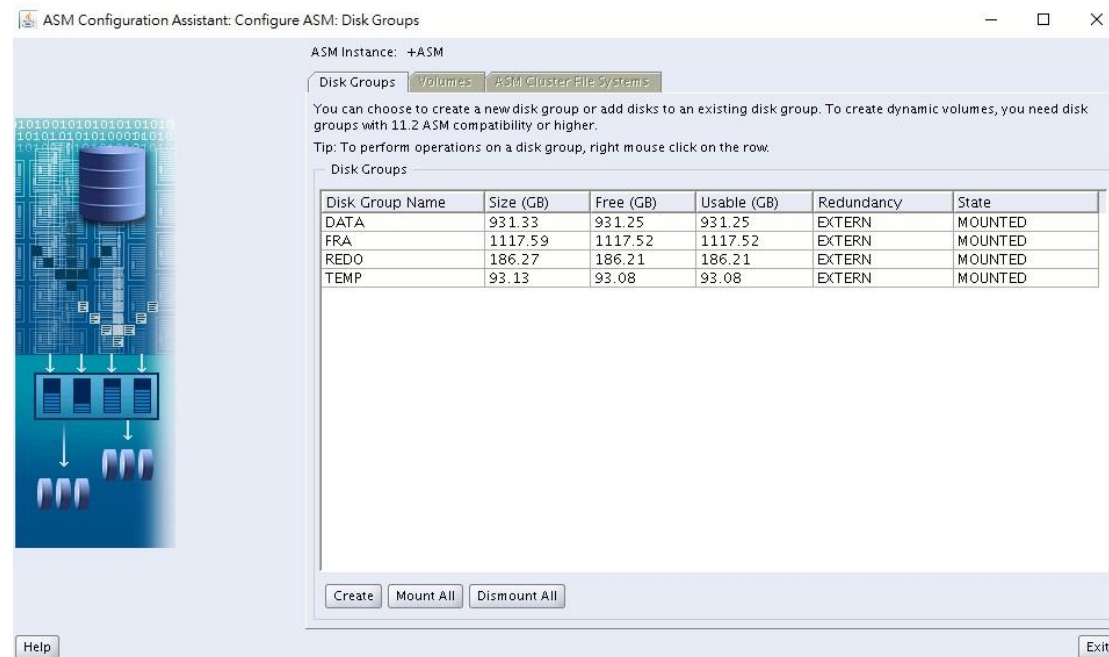
10. Oracle Database Software Installation

There are 2 steps to install Oracle Database. First, install Oracle software.
Switch to user "oracle" and run the command below.

```
oracle@astor ~]$ cd database
oracle@astor ~]$ ./runInstaller
```

Select the type of database installation, e.g., "Single instance database installation".

For more detailed information about Oracle Database software installation, please refer to the Oracle Database Installation Guide.

11. Oracle Database Creation
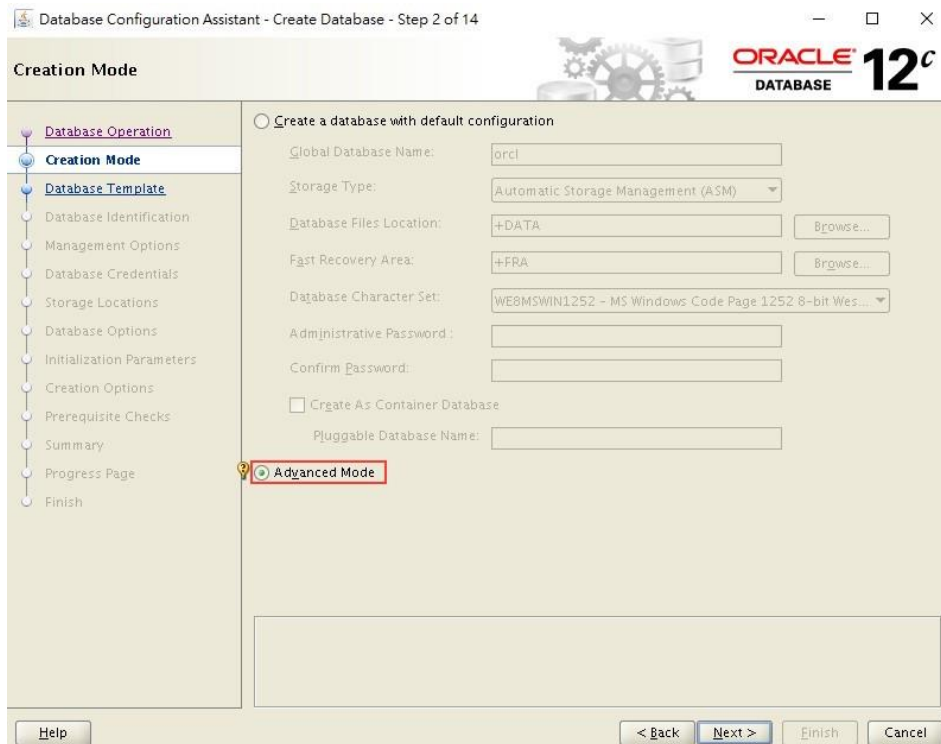
After the Oracle Database software installation is complete, the second step is to create an Oracle database and assign the location of data files, redo log files, and a fast recovery area.

Switch to user "oracle" and run command "dbca".

oracle@astor ~]$ dbca
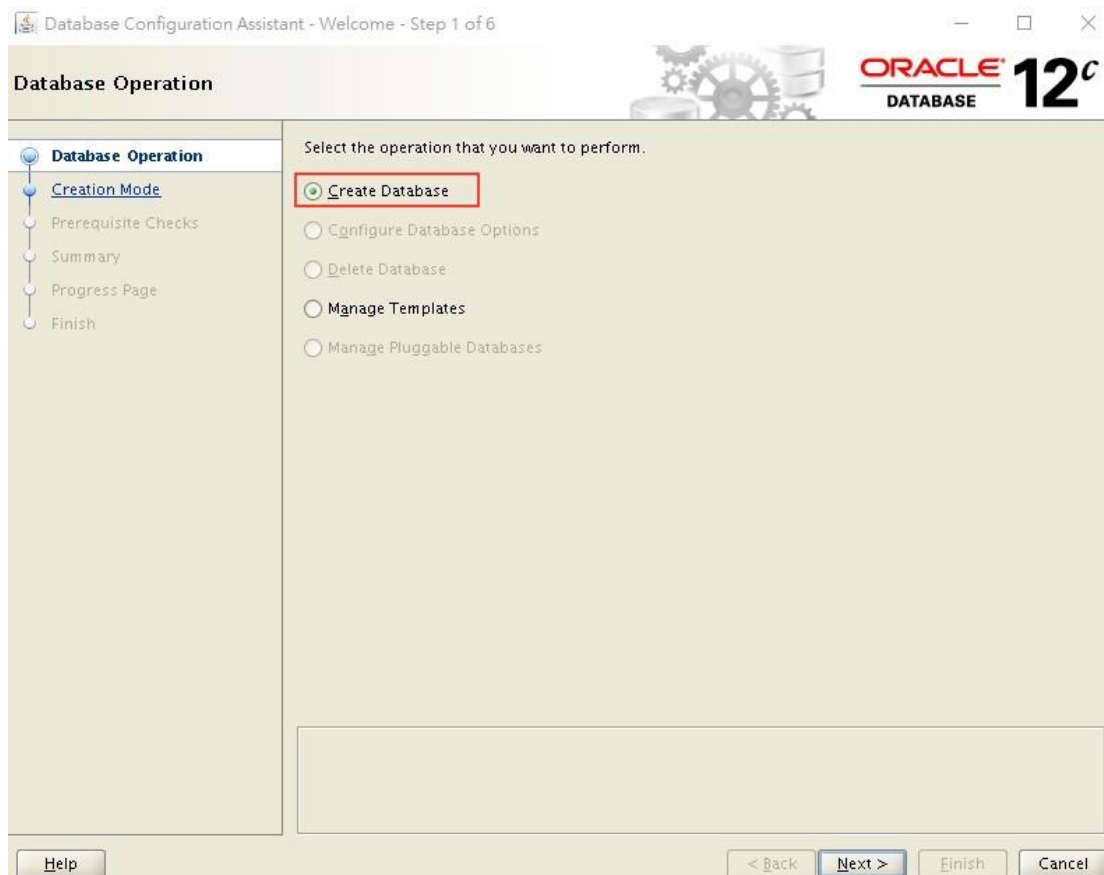
Under Creation Mode, click "Advanced Mode".
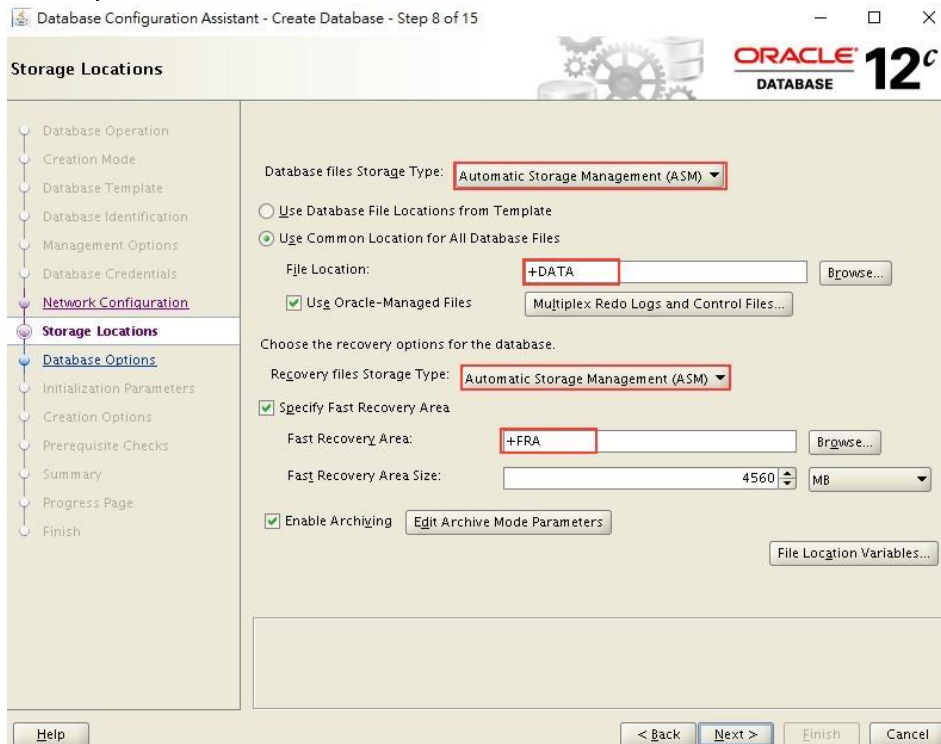
Under Storage Locations, choose the settings below:
Database files Storage Type: Automatic Storage Management (ASM)
Use Common Location for All Database Files; File Location: **+DATA**.
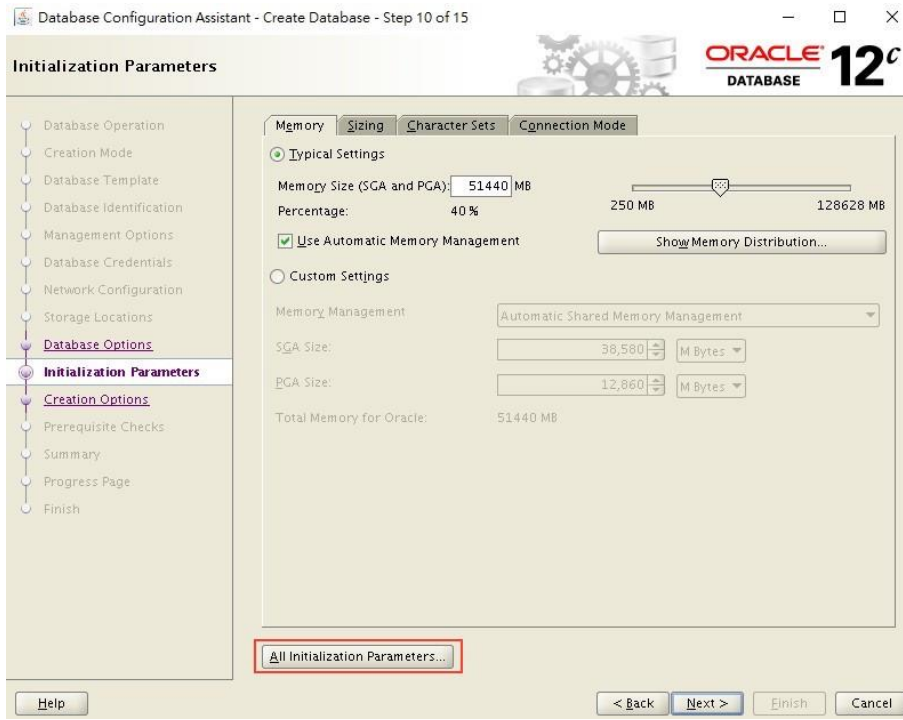
Recovery files Storage Type: Automatic Storage Management (ASM)
Fast Recovery Area: **+FRA**.
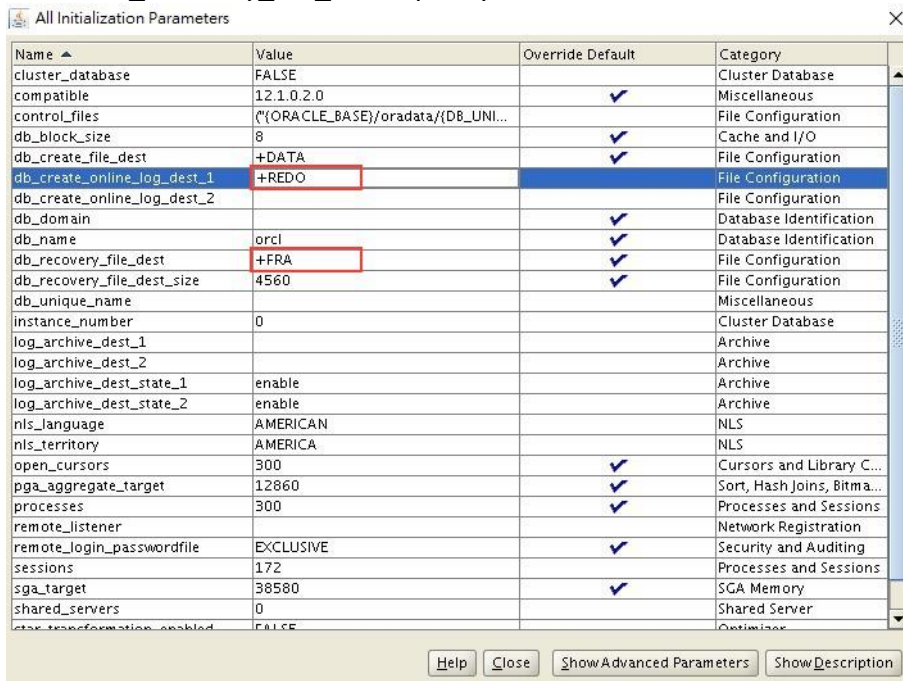
Under Initialization Parameters, click "All Initialization Parameters".

For the destination of db_create_online_log_dest_1, specify +REDO. For the destination of db_recovery_file_dest, specify +FRA.

For more detailed information about creating a database, please refer to the Oracle Database Installation Guide.

12. Relocate TEMP Tablespace

Relocate TEMP tablespace to +temp. You will need to restart the database after the steps below are completed.

```
SQL> create temporary tablespace temp123 tempfile '+temp'
  2  size 20M autoextend on next 640k maxsize unlimited;
Tablespace created.

SQL> alter database default temporary tablespace temp123;
Database altered.

SQL> alter tablespace temp tempfile offline;
Tablespace altered.

SQL> drop tablespace temp including contents and datafiles;
Tablespace dropped.

SQL> alter tablespace temp123 rename to temp;
Tablespace altered.
```

13. 4KB Redo Log Files

The online redo log files are crucial for the database to perform transactions reliably. All changed transaction records must be written to redo log files.
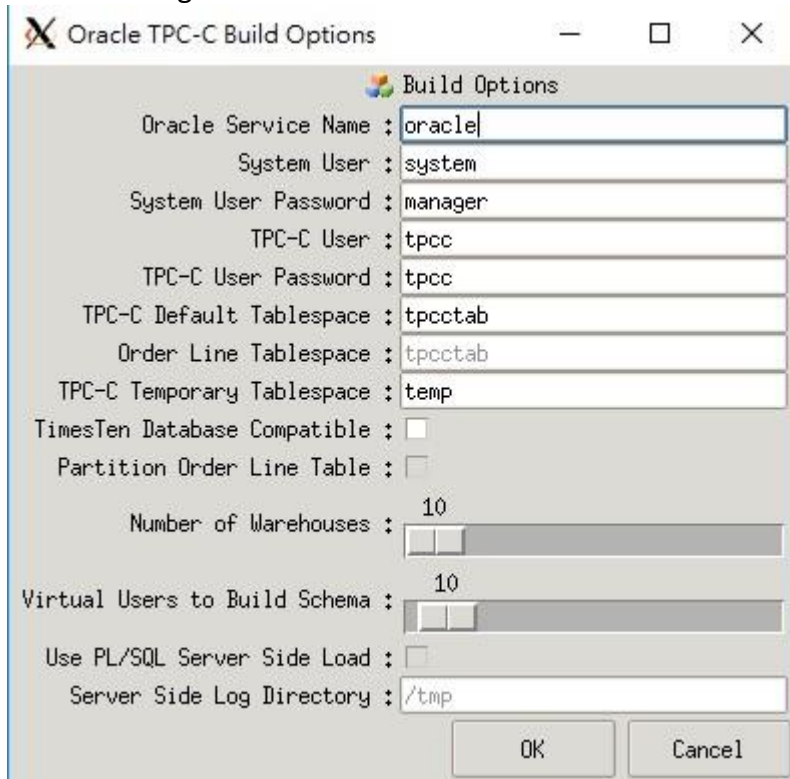
Traditionally, the default block size of redo log files is equal to the physical block size of storage. Typically, it is 512 bytes. With the proprietary algorithm of FlexiRemap, the block size is increased to 4KB. The larger block size allows for higher capacity storage as well as improved performance.

It is strongly recommended to set the block size to 4096 bytes when creating redo log files. Note the example below:

```
SQL> alter database add logfile group 2 (
  2  '+REDO/ORCL/ONLINELOG/group21',
  3  '+REDO/ORCL/ONLINELOG/group22') size 15G blocksize 4096 reuse;
```
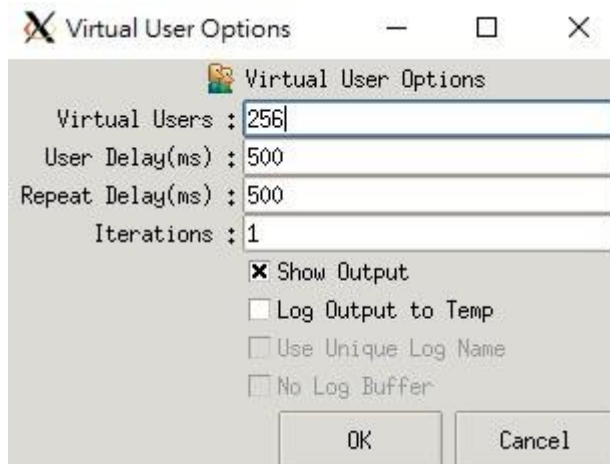
14. HammerDB Configuration

Remember to create the user TPC-C with dba privileges in the database. Here it is only necessary to adjust the number of Warehouse and Virtual Users to build Schema. The other settings remain as default.

This value may exceed default maximum processes of the database. If HammerDB reports a warning, it is necessary to increase the number of maximum processes. Virtual Users: 256.

For more detailed information, please download HammerDB Oracle OLTP Load Testing.

15. Performance

The NeoSapphire 3605 offers 5TB usable capacity, while the NeoSapphire 3611 provides 11TB in a very compact 1U rack-mount form factor. The low latency and high reliability of Fibre Channel makes these new arrays ideal for demanding, time-critical applications including banking, finance, and media production. Reliability is further enhanced by hot-swappable drives and redundant power supplies. NeoSapphire series are equipped with AccelStor's exclusive FlexiRemap software-defined flash technology. Designed for sustained performance and reliability even in the random-access scenarios typical of enterprise storage needs, FlexiRemap software technology remaps random data writes as sequential whenever beneficial, before passing them to the underlying SSDs. The high performance of 1.2M TPM from the HammerDB benchmark and NeoSapphire's powerful features will boost IT staff productivity and make deployment easier.