

FlashGrid

Provisioning and Maintenance Guide for Oracle RAC in Azure Cloud

rev. 18.08-2018.10.04



Table of Contents

1	Introduction	3
2	Compatibility.....	3
3	Prerequisites	3
4	Provisioning a Cluster.....	4
4.1	Creating a cluster with Azure Resource Manager	4
4.2	Verifying cluster health	5
4.3	Verifying synchronization of clocks	5
4.4	OS User Accounts	6
4.5	Changing ASM passwords	6
4.6	Adding a protection lock for the cluster	6
4.7	Deleting a cluster	7
5	Creating a Database	8
5.1	Creating a database using DBCA	8
5.2	Configuring database memory settings	8
6	Enabling Strict Read-Local Mode for a New Database	9
7	Measuring Performance	10
8	Connecting Database Clients	11
8.1	Configuring Alias for Local Listener	12
8.2	Setting LOCAL_LISTENER parameter for each database.....	12
8.3	Adding DNS Records.....	13
8.4	Opening listener ports.....	13
8.5	Testing client connectivity.....	13
8.6	Configuring client connect string	14
9	Monitoring Cluster Health	15
10	Backup Best Practices	16
10.1	Backing up OS and software on the cluster nodes	16
10.2	Backing up and restoring database files	16
10.3	Backing up and restoring files on ACFS	16
11	Maintenance Tasks	17
11.1	Rebooting a node	17
11.2	Powering off a node	17
11.3	Shutting down an entire cluster	17
11.4	Adding disks for use in ASM	18
11.5	Removing disks.....	18
11.6	Re-adding a lost disk	18
11.7	Updating FlashGrid software.....	20
11.8	Updating Linux kernel	21
12	FlashGrid Tools and Commands.....	22
12.1	flashgrid-ca.....	22
12.2	flashgrid-fix-grid-dg-ca	22
12.3	flashgrid-create-dg	22
12.4	flashgrid-cluster.....	22
12.5	flashgrid-node	23
12.6	flashgrid-dg	24
12.7	flashgrid-diags	25
13	Troubleshooting.....	25
14	Contacting FlashGrid Technical Support	25

1 Introduction

FlashGrid Cloud Area Network software and FlashGrid Storage Fabric software enable running Oracle RAC clusters in Azure cloud. FlashGrid Cloud Provisioning simplifies the deployment process by automating configuration of multiple components required for an Oracle RAC cluster. This guide provides step-by-step instructions for system and database administrators deploying Oracle RAC in Azure using the FlashGrid Cloud Provisioning tool. Additional information about the solution is available in the following white paper: *“Mission-Critical Databases in the Cloud. Oracle RAC in Microsoft Azure Enabled by FlashGrid®.”*

2 Compatibility

The following versions of software are covered in this guide:

- FlashGrid Cloud Provisioning: ver. 18.08
- FlashGrid Storage Fabric: ver. 18.08
- FlashGrid Cloud Area Network: ver. 18.07
- FlashGrid Cluster Monitoring and Diagnostics: ver. 18.04
- Oracle Database: ver. 12.2.0.1, 12.1.0.2, or 11.2.0.4 with the latest PSU/RU
- Oracle Grid Infrastructure: ver. 12.2.0.1 with the latest Release Update
- Operating System: Oracle Linux 7, Red Hat Enterprise Linux 7

The solution can be deployed on the following Azure VM families: DSv2, DSv3, ESv3, M, GS.

3 Prerequisites

The following prerequisites are required for automated provisioning of an Oracle RAC cluster in Azure using FlashGrid Cloud Provisioning:

- **Azure Storage Blob Container** with Oracle installation files that will be downloaded to the cluster nodes during provisioning. The list of files that must be placed in the Storage Container is available at <https://www.flashgrid.io/cloud-provisioning-for-azure/>
- **Azure subscription with sufficient quotas** for creating the required number and type of VMs and sufficient number and size of Premium Managed Disks.
- **SSH key pair** that will be used for accessing the VMs. Use of passwords instead of the key pair is not supported. To create a new key pair use *ssh-keygen* in Linux or *puttygen* in Windows. In the Cloud Provisioning tool you will need to provide the public key that will be placed on the VMs. Example of a valid public key pair format (must be a single line):

```
ssh-rsa <PublicKeyBody>
```

- **Properly configured Network Security Group (NSG)** when deploying in an existing VNet. You have a choice of attaching an NSG to the VMs or using the NSG attached to the subnet. In either case, the following ports must be open between the cluster node VMs: UDP 4801, 4802, 4803 and TCP 3260. FlashGrid recommends configuring the NSG rules by using an Application Security Group (ASG) for the cluster node VMs. You can configure one ASG per cluster or a separate ASG for each cluster.

4 Provisioning a Cluster

The FlashGrid Cloud Provisioning tool simplifies provisioning of Oracle RAC clusters in Azure by automating the following tasks:

- Creating cloud infrastructure: VMs, storage, and optionally network
- Installing and configuring FlashGrid Cloud Area Network
- Installing and configuring FlashGrid Storage Fabric
- Installing, configuring, and patching Oracle Grid Infrastructure
- Installing and patching Oracle Database software
- Creating ASM disk groups

4.1 Creating a cluster with Azure Resource Manager

To create a cluster

1. Open FlashGrid Cloud Provisioning tool:
 - Start with one of the standard configurations at <https://www.flashgrid.io/cloud-provisioning-for-azure/>
 - or, if you have a custom configuration file, upload it at <https://1808.cloudprov.flashgrid.io>
2. Configure parameters of the cluster
3. Click *Verify* button
4. If verification passes then click *Create Stack* button, which will take you to Azure Resource Manager
5. Select *Resource group* -> *Create new*. By having the cluster in a separate resource group you can later delete the entire cluster by simply deleting the resource group.
6. Enter a name for the new resource group that will contain the cluster. A name matching the cluster name is recommended.
7. Select your target location (region)
8. Check *'I agree to the terms and conditions state above'*
9. Click *Purchase*
10. Open list of Notifications (bell icon) and click *'Deployment in progress...'*
11. Wait until the deployment status changes to *Succeeded*
12. If the deployment fails:
 - a) Check for the cause of the failure in the *Operation details*
 - b) Correct the cause of the error
 - c) Delete the failed resource group
 - d) Repeat the steps for creating a new resource group
13. Get IP address of the newly created VM corresponding to the first node
14. SSH to the first (as it was specified on the cluster configuration page) cluster node VM as user `fg@`
15. If the cluster initialization is still in progress or failed then you will see a corresponding welcome message. If there is no welcome message then the cluster initialization has completed successfully.
16. Wait for cluster initialization (including Oracle software installation and configuration) to complete. You will receive a broadcast message when initialization completes or fails. Cluster initialization takes approximately 90 minutes.

4.2 Verifying cluster health

On any of the cluster nodes run `flashgrid-cluster` command to verify that the cluster status is *Good* and all checks are passing.

```
[fg@rac1 ~]$ flashgrid-cluster
FlashGrid 18.07.15.48564 #95f2b5603f206af26482ac82386b1268b283fc3c
License: via Marketplace Subscription
Support plan: 24x7
~~~~~
FlashGrid running: OK
Clocks check: OK
Configuration check: OK
Network check: OK

Querying nodes: quorum, rac1, rac2 ...

Cluster Name: myrac
Cluster status: Good
-----
Node      Status  ASM_Node  Storage_Node  Quorum_Node  Failgroup
-----
rac1     Good   Yes       Yes           No           RAC1
rac2     Good   Yes       Yes           No           RAC2
racq     Good   No        No            Yes          QUORUM
-----
-----
GroupName  Status  Mounted  Type      TotalMiB  FreeMiB  OfflineDisks  LostDisks  Resync  ReadLocal  Vote
-----
GRID       Good   AllNodes  NORMAL    12588     3376     0              0          No     Enabled    3/3
DATA       Good   AllNodes  NORMAL    2048000   2048000  0              0          No     Enabled    None
FRA        Good   AllNodes  NORMAL    1024000   1024000  0              0          No     Enabled    None
-----
```

4.3 Verifying synchronization of clocks

Starting with FlashGrid Cloud Provisioning version 18.03 the Chrony service is used for synchronizing cluster node clocks with external NTP servers. Without active clock synchronization service the clocks are likely to get out of sync. Oracle CTSS service synchronizes system clocks while CRS is running. However, it cannot synchronize clocks before CRS is started or on quorum nodes. Clock synchronization is important for FlashGrid Storage Fabric. Clocks difference of 30 seconds or more will cause warnings. Clocks difference of 60 seconds or more will cause disconnect of remotely attached disks.

To check the current clock difference between the cluster nodes

```
$ flashgrid-cluster verify
```

To check status of CHRONYD service

```
$ chronyc sources
```

Example:

```
[fg@rac1 ~]$ chronyc sources
210 Number of sources = 6
MS Name/IP address          Stratum Poll Reach LastRx Last sample
=====
^+ time1.google.com         1   9   377   197  +4820us[+4820us] +/-  23ms
^* time2.google.com         1  10   377   838  -3017us[-3363us] +/-  14ms
^- time3.google.com         1   9   377   313  -5459us[-5459us] +/-  60ms
^- time4.google.com         1   9   377    41   +13ms[ +13ms] +/- 100ms
-- rac2-ext.example.com     2   9   367   398  +109us[ +109us] +/-  14ms
-- racq-ext.example.com     2  10   357   263  -341us[ -341us] +/-  14ms
```

Note that the '*' character shows which NTP server is currently used for synchronization. Normally this should be one of the external NTP servers. If it shows that one of the cluster nodes is used for synchronization then this means that the external NTP servers are not accessible.

The cluster nodes are configured as peers for the Chrony service to provide synchronization between the nodes even when the external NTP servers are not accessible. However, this is only a temporary solution for cases when external NTP servers become inaccessible. In production environments the peer nodes should not be relied upon as a permanent clock synchronization solution.

Public NTP servers, e.g. *timeX.google.com*, can be used only if public IPs are enabled on the VMs (not recommended in production use for security reasons) or if NAT is configured in the network. If needed, the list of NTP servers can be modified in `/etc/chrony.conf` after the cluster is configured.

4.4 OS User Accounts

During cluster initialization the following OS user accounts are created:

- *fg* - the user account used to SSH to the VMs with the SSH key that was selected when creating the cluster configuration. It can also be used for running FlashGrid Storage Fabric or FlashGrid Cloud Area Network utilities. The user *fg* has sudo rights.
- *grid* - Grid Infrastructure owner. GI environment variables are preconfigured.
- *oracle* - Database home owner. Database environment variables, except ORACLE_SID and ORACLE_UNQNAME, are preconfigured. After creating a database you can configure ORACLE_SID and ORACLE_UNQNAME by editing `/home/oracle/.bashrc` file on each database node.

Note that no passwords are configured for any users. Also password-based SSH authentication is disabled in `/etc/ssh/sshd_config`. Key-based authentication is recommended for better security. Creating passwords for any user is not recommended.

User *fg* has sudo rights and allows switching to any other user without requiring a password (which is not configured by default). Example:

```
$ sudo su - grid
```

Users *fg*, *grid*, and *oracle* have key-based SSH access configured between the nodes of the cluster. The corresponding key pairs are generated automatically during cluster initialization. For example, if you are logged in to *node1* as user *fg* then you can SSH into *node2* by simply running `'ssh node2'` without entering a password or providing a key.

4.5 Changing ASM passwords

The automatic configuration process sets a temporary password for ASM users SYS and ASMSNMP. Use the following commands to set new passwords:

```
$ sudo su - grid
$ sqlplus / as sysasm
SQL> alter user sys identified by "MyNewPassword";
SQL> alter user asmsnmp identified by "MyNewPassword";
```

4.6 Adding a protection lock for the cluster

It is strongly recommended to add a lock to the cluster resource group to protect it against accidental deletion or modification.

4.7 Deleting a cluster

To delete a cluster

1. Delete any protection lock(s) for the resource group
2. Delete the resource group corresponding to the cluster

5 Creating a Database

Use standard procedures for creating, restoring, or importing a database. Select ASM storage and use DATA disk group for the database files and FRA disk group for Fast Recovery Area. Do not use the GRID disk group as storage for a database. The GRID disk group is dedicated for OCR files, voting files, and MGMTDB. It must not be used for any other purposes.

5.1 Creating a database using DBCA

For running DBCA in GUI mode use X-forwarding or VNC. If running Oracle Database version 11.2.0.4, note that the DBCA utility has problems when working via VNC - use X-forwarding or silent mode instead.

To run DBCA via VNC, perform the following steps on database node 1

1. Make sure VNC port 5901 (TCP) is either open via the security group settings or forwarded via SSH tunnel.
2. As root, run `vncserver` command and set VNC password
3. Connect your VNC client (e.g. RealVNC) to the node. Use one of the following connection strings:
 - `PublicIP:5901` - if connecting via public IP
 - `PrivateIP:5901` - if connecting via VPN to the VPC private IP
 - `localhost:5901` - if using port forwarding via SSH
4. Run the following commands in the x-terminal inside VNC client

```
# xhost localhost
# su oracle
$ dbca
```

DBCA is expected to show a warning about insufficient swap size. Swap is intentionally disabled and the warning should be ignored. If the swap is enabled than swapping may happen when system is low on memory even when there is some physical memory still available. Swapping will cause system being unresponsive and potentially causing time outs at various levels including storage. While swapping may prevent or delay running out of memory, the results of swapping are likely to be worse than the results of running out of memory.

5.2 Configuring database memory settings

Larger SGA and PGA allocations can help with achieving higher database performance. However, it is critical to ensure that the settings are configured correctly to avoid swapping or running out of memory. Running out of memory will result in processes being killed with unpredictable results to the system stability. A typical mistake leading to low available memory condition is having the PGA size parameters set too high or too many HugePages configured.

On systems with 60 GiB or larger physical memory FlashGrid recommends allocating 80% of the total memory for use by the database(s). The remaining 20% must be reserved for the OS, Grid Infrastructure, and FlashGrid software. On systems with less than 60 GiB of physical memory, 12 GiB must be reserved for the OS, Grid Infrastructure, and FlashGrid software.

The optimal ratio of SGA and PGA is different for different types of databases. However, the sum of SGA and PGA allocations for all databases must not exceed the total memory *DatabaseMemory* value as calculated above.

If configuring HugePages then the amount of memory allocated as HugePages must match the SGA target (or sum of all SGA targets for multiple databases) plus 2 GiB for GIMR. Note that PGA and other software cannot use HugePages. Allocating too many HugePages may result in running out of memory.

Example of memory allocations calculation for a 20% PGA / 80% SGA ratio for a single database:

- 1) Calculate total database memory (for all databases)

If `Total_Memory_GiB` >= 60 GiB: $Database_Memory_GiB = 0.8 \times Total_Memory_GiB$

If `Total_Memory_GiB` < 60 GiB: $Database_Memory_GiB = Total_Memory_GiB - 12\ GiB$

- 2) Calculate PGA size parameters

$PGA_AGGREGATE_LIMIT_GiB = round(0.2 \times Database_Memory_GiB)$

$PGA_AGGREGATE_TARGET_GiB = round(0.5 \times PGA_AGGREGATE_LIMIT_GiB)$

Note: In database version 11.2.0.4 explicitly setting `PGA_AGGREGATE_LIMIT` parameter is not supported. It is calculated automatically from `PGA_AGGREGATE_TARGET`.

- 3) Calculate SGA max size

$SGA_MAX_SIZE_GiB = round(0.8 \times Database_Memory_GiB)$

- 4) Calculate number of huge pages

$Number_HugePages = (SGA_MAX_SIZE_GiB + 2) \times 512$

In case of multiple databases sharing the same nodes, the **sum of all** PGA aggregate limit/target parameters must be equal to or lower than the values calculated using the formulas above.

6 Enabling Strict Read-Local Mode for a New Database

It is recommended that *Strict Read-Local* mode is enabled for every new database.

ASM does not allow reads from disks that are resynchronizing data (SYNCING state) after being offline. As a result, if database is running on a node whose local disks are in SYNCING state, all reads will be performed remotely over the network. This may result in lower performance of the database instance on a node that has just rebooted and is still resynchronizing its data.

Strict Read-Local mode prevents such performance asymmetry between nodes. When the *Strict Read-Local* mode is enabled, a database instance start will be delayed until its local disks complete resynchronization.

Use the following commands to enable, disable, and show status of Strict Read-Local mode:

```
$ flashgrid-cluster strict-read-local-enable
$ flashgrid-cluster strict-read-local-disable
$ flashgrid-cluster strict-read-local-show
```

Note that enabling *Strict Read-Local* mode changes the setting only for existing databases. Re-running the *enable* command is required after creating new database(s).

Note that in order to unmount a disk group while Strict Read-Local mode is enabled, `srvctl stop diskgroup` command with `-force` option must be used. Example:

```
$ srvctl stop diskgroup -diskgroup DGNAME -node rac1,rac2 -force
```

7 Measuring Performance

DBMS_RESOURCE_MANAGER.CALIBRATE_IO procedure provides an easy way for measuring storage performance including maximum bandwidth, random IOPS, and latency. The CALIBRATE_IO procedure generates I/O through the database stack on actual database files. The test is read-only and it is safe to run it on any existing database. It is also a good tool for directly comparing performance of two storage systems because the CALIBRATE_IO results do not depend on any non-storage factors, such as memory size or the number of CPU cores.

To measure storage performance with CALIBRATE_IO

1. Create or load a database on the corresponding ASM disk group
2. Make sure the total size of the database files is larger than 5 GB per disk. If needed, create an additional large table space / data file.
3. Customize the first parameter in the SQL code below with the number of disks corresponding to your storage setup. Keep the second parameter (max latency) with the minimum allowed value of 10 milliseconds.
4. Connect to the database with sqlplus and run the customized SQL code.
5. Wait for the CALIBRATE_IO to complete. This may take 10 to 30 minutes.

Example of running CALIBRATE_IO on a 2-node cluster with E64s_v3 VMs and 16 x 1024 GB Premium disks per node

```
SQL> SET SERVEROUTPUT ON;
DECLARE
lat INTEGER;
iops INTEGER;
mbps INTEGER;
BEGIN DBMS_RESOURCE_MANAGER.CALIBRATE_IO (32, 10, iops, mbps, lat);
DBMS_OUTPUT.PUT_LINE ('Max_IOPS = ' || iops);
DBMS_OUTPUT.PUT_LINE ('Latency = ' || lat);
DBMS_OUTPUT.PUT_LINE ('Max_MB/s = ' || mbps);
end;
/
Max_IOPS = 140605
Latency = 1
Max_MB/s = 3250

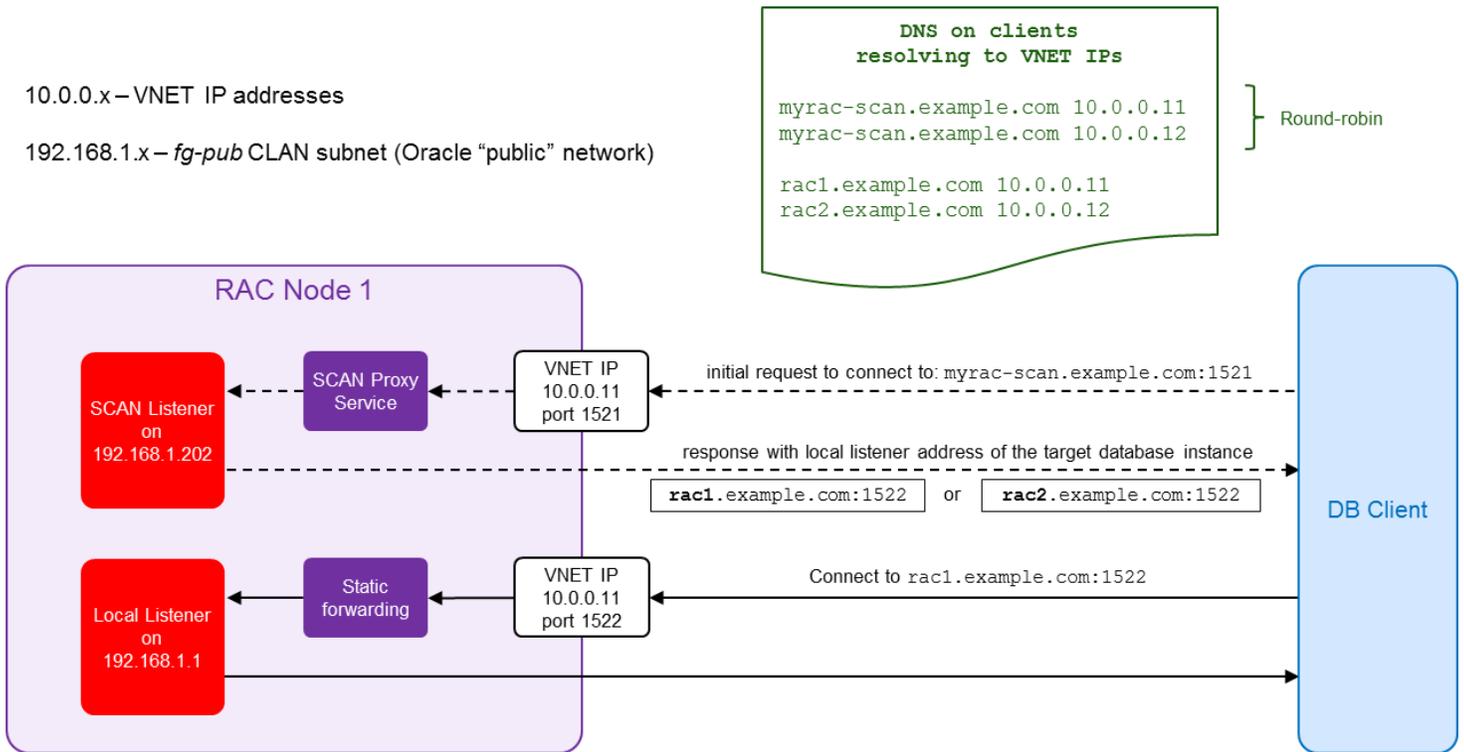
PL/SQL procedure successfully completed.
```

8 Connecting Database Clients

Clients (e.g. app servers) connecting to the database need connectivity to SCAN listeners and local listeners. However, in the FlashGrid architecture the listeners are configured on the *fg-pub* CLAN virtual network, which is isolated from the rest of the VNet where the clients are located. FlashGrid SCAN Proxy service running on each database node of the cluster facilitates connectivity between the clients on the VNet and the SCAN listeners on the *fg-pub* subnet. Port forwarding facilitates connectivity between the clients on the VNet and the local listeners on the *fg-pub* subnet.

Because of the two separate IP address spaces, host name resolution must be configured with split-DNS. Within the cluster, DNSMASQ service running on each node of the cluster provides host name resolution to the 192.168.1.x address space. For the clients outside of the cluster, DNS servers must be configured with the database host names resolving to the corresponding VNET IP addresses.

The diagram below shows an example of the process of establishing a client connection via a SCAN address.



Establishing client connection via SCAN address

8.1 Configuring Alias for Local Listener

Use of an alias for the local listener simplifies use of a split horizon DNS resolution and allows setting same LOCAL_LISTENER parameter for all database instances.

In most cases the default configuration of the local listener alias is sufficient and does not require any changes. If automatic installation of the database software was selected in FlashGrid Cloud Provisioning then the alias entry is created in `<database_home>/network/admin/tnsnames.ora`

Example of a default alias entry on `rac1` node:

```
DONOTDELETE,NodeFQDN = (ADDRESS = (PROTOCOL= TCP) (Host=rac1.example.com) (Port=1522))
```

In certain cases listed below these alias entries must be modified or added manually.

- **When default domain is configured in sqlnet.ora.** If you configure default domain in sqlnet.ora then you also need to replace `NodeFQDN` with `NodeFQDN.domainname` in the `<database_home>/network/admin/tnsnames.ora` file.
- **When manually installing database software.** If you did not select automatic installation of database software in the FlashGrid Cloud Provisioning tool then the `<database_home>/network/admin/tnsnames.ora` file with the corresponding entries must be created manually.
- **When same `tnsnames.ora` file must be used across all database nodes.** By default the `tnsnames.ora` file is different on each node and contains same alias name `NodeFQDN`, but a different address corresponding to the host. This allows setting the LOCAL_LISTENER parameter to the same value on all database instances. If you need to use same `tnsnames.ora` file on all nodes then it must contain a separate alias entry for each node, e.g. `NodeFQDN1` and `NodeFQDN2` instead of a single `NodeFQDN`, and the LOCAL_LISTENER parameter of each database instance must be configured with the alias corresponding to the host where the database instance is running.

Note that in all cases a DNS address of the host must be used in the Host attribute of the alias. Do not replace the DNS address with an IP address. Also, the Port parameter must match the local listener port number selected in the FlashGrid Cloud Provisioning tool (1522 by default).

8.2 Setting LOCAL_LISTENER parameter for each database

Configuring the LOCAL_LISTENER parameter is required for accessing database services via the SCAN Listener. It must be set separately for each database.

To set the parameter, log in to each database and run the following command:

```
SQL> ALTER SYSTEM SET LOCAL_LISTENER="NodeFQDN" scope=both sid='*';
```

Note: Do not replace the `"NodeFQDN"` parameter with a hostname. Run the command as is.

8.3 Adding DNS Records

On the DNS server(s) used by clients, for each database node add two records resolving to the VNet Private IP address of the node VM:

- 1) Hostname of the database node
- 2) SCAN address

Example for a 2-node cluster:

```
rac1.example.com 10.100.0.1
rac2.example.com 10.100.0.2
myrac-scan.example.com 10.100.0.1
myrac-scan.example.com 10.100.0.2
```

IMPORTANT! Do not modify the `/etc/resolv.conf` files on the database servers. Within the cluster the DNS names are resolved to IP addresses on the *fg-pub* CLAN subnet 192.168.1.x via local DNSMASQ service.

8.4 Opening listener ports

The FlashGrid architecture requires separate port numbers for the SCAN Listener (default: 1521) and for the Local Listener (default: 1522). The default ports can be changed when creating a cluster configuration. Make sure that both ports are open in the corresponding Network Security Group settings for inbound connections from the clients.

8.5 Testing client connectivity

To test client connectivity via the SCAN address

1. Confirm that Local Listeners are reachable via the hostnames (testing with IP address is not enough)

```
$ tnsping rac1.example.com:1522
$ tnsping rac2.example.com:1522
```

2. Connect to all database instances via Local Listeners (replace *orcl* with the corresponding database name)

```
sqlplus "sys/password@(DESCRIPTION=(ADDRESS=(PROTOCOL=tcp)(HOST=rac1.example.com)(PORT=1522))
(CONNECT_DATA=(SERVICE_NAME=orcl)))" as sysdba

sqlplus "sys/password@(DESCRIPTION=(ADDRESS=(PROTOCOL=tcp)(HOST=rac2.example.com)(PORT=1522))
(CONNECT_DATA=(SERVICE_NAME=orcl)))" as sysdba
```

3. Confirm that SCAN Listeners are reachable via the SCAN address (run three times to cycle through all nodes)

```
$ tnsping myrac-scan.example.com:1521
```

4. Connect to a database service via SCAN Listener (replace *salesservice* with the corresponding service name)

```
sqlplus "sys/password@(DESCRIPTION=(TRANSPORT_CONNECT_TIMEOUT=3)(RETRY_COUNT=6)
(ADDRESS=(PROTOCOL=tcp)(HOST=myrac-scan.example.com)(PORT=1521))(CONNECT_DATA=(SERVICE_NAME=
salesservice)))" as sysdba
```

8.6 Configuring client connect string

The following parameters must be specified in the client connect string:

- `TRANSPORT_CONNECT_TIMEOUT=3`
The time, in seconds, for a client to establish a TCP connection to the database server. The default value is 60 seconds. It must be changed to avoid a long wait in case one of the database servers is down.
- `RETRY_COUNT=6`
The number of connection attempts before the connection is terminated.

Example of a TNSNAMES.ORA entry:

```
SALESservice=(DESCRIPTION=(TRANSPORT_CONNECT_TIMEOUT=3)(RETRY_COUNT=6)
  (ADDRESS=(PROTOCOL=tcp)(HOST=myrac-scan.example.com)(PORT=1521))
  (CONNECT_DATA=(SERVICE_NAME=saleservice)))
```

Example of a JDBC thin connect string:

```
jdbc:oracle:thin:@(DESCRIPTION=(TRANSPORT_CONNECT_TIMEOUT=3)(RETRY_COUNT=6)
  (ADDRESS=(PROTOCOL=tcp)(HOST=myrac-scan.example.com)(PORT=1521))
  (CONNECT_DATA=(SERVICE_NAME=saleservice)))
```

For more details about the connect string parameters see <https://docs.oracle.com/database/121/NETRF/tnsnames.htm>

9 Monitoring Cluster Health

The following methods of monitoring cluster health are available:

- `flashgrid-cluster` utility displays status of the cluster and its main components
- Alerts about failures are recorded in system log and can be analyzed by 3rd-party tools
- Email alerts can be sent to one or several email addresses
- ASM disk group monitoring and alerting via Oracle Enterprise Manager

To test email alerts

1. On all nodes (including quorum node) run

```
$ flashgrid-node test-alerts
```

2. Check that test alert emails were received from all cluster nodes at each of the configured email addresses.

To modify the list of email alert recipients

As user `fg@` on any database node run

```
$ flashgrid-cluster set-email-alerts name1@host1 name2@host2 ...
```

Note that by default the *From* address is set to `flashgrid@localhost.localdomain`. This will ensure that delivery failure notifications are sent to root's mailbox on the originating node, which can help with troubleshooting delivery issues. It is recommended to add this address to the whitelist of senders on the receiving email server and in the email clients.

10 Backup Best Practices

This section covers backup procedures that must be used for different types of data including:

- OS and software on the cluster nodes
- Database files
- Files on ACFS, if used

10.1 Backing up OS and software on the cluster nodes

It is strongly recommended to back up the OS disk and software disk (LUN 0) of all cluster nodes after the initial cluster configuration and before and after applying any changes such as patch installation or security settings changes. Taking disk snapshots is the recommended method for backing up the disks.

Note that Azure Backup service is not currently supported due to its existing restrictions.

To create a backup of a cluster node VM

1. If the node is a database node, stop Oracle services.

```
# crsctl stop crs
```

2. Gracefully stop the instance.

```
# flashgrid-node poweroff
```

3. Create snapshot of the OS disk
4. Create snapshot of the software disk (LUN 0)

10.2 Backing up and restoring database files

Use standard RMAN procedures for backing up and restoring database files. A disk attached to one of the database node VMs is recommended as a backup destination. The disk must be attached with a LUN number 50 through 64 (LUNs 1 through 49 are automatically shared by Storage Fabric). The disk must be formatted with a local file system, such as XFS.

The disk image can be exported for storing the backup data outside of Azure.

10.3 Backing up and restoring files on ACFS

For backing up and restoring files on ACFS use same tools and procedures that you would normally use for file-level backup and restore.

11 Maintenance Tasks

11.1 Rebooting a node

To reboot a node in a running cluster

1. Make sure there are no other nodes that are in offline or re-syncing state. All disk groups must have zero offline disks and *Resync = No*.

```
# flashgrid-cluster
```

2. If the node is a database node, stop all local database instances running on the node.
3. Reboot the node using flashgrid-node command. It will gracefully put the corresponding failure group offline.

```
# flashgrid-node reboot
```

4. After the nodes boots up, wait until re-synchronization completes for all disk groups before rebooting or powering off any other node.

11.2 Powering off a node

To power off a node in a running cluster

1. Make sure there are no other nodes that are in offline or re-syncing state. All disk groups must have zero offline disks and *Resync = No*.

```
# flashgrid-cluster
```

2. If the node is a database node, stop all local database instances running on the node.
3. Stop FalshGrid services on the node. It will gracefully put the corresponding failure group offline, stop CRS, and stop Flashgrid services.

```
# flashgrid-node stop
```

4. Stop the node VM using Azure console.
5. After powering up the node, wait until re-synchronization completes for all disk groups before rebooting or powering off any other node.

11.3 Shutting down an entire cluster

To shut an entire cluster down

1. Stop all running databases
2. Stop Oracle cluster services on all nodes.

```
# crsctl stop cluster -all
```

3. Stop all cluster node VMs using Azure console.

11.4 Adding disks for use in ASM

When adding new disks make sure that each disk group has disks of the same size and that the number of disks per node is the same.

To add new disks in a running cluster

1. Create and attach new disks to the database node VMs. Attach the disks using LUN numbers 1 through 40 - these LUNs will be shared automatically by FlashGrid Storage Fabric.
2. Confirm FlashGrid names of the new drives, e.g. *rac2.lun3*

```
$ flashgrid-cluster drives
```

3. If the new disks are not listed then you might need to reload FlashGrid Storage Fabric configuration on the corresponding node:

```
$ sudo flashgrid-node reload-config
```

4. Create a new disk group: `$ flashgrid-create-dg`
or
5. Add the new disks to an existing disk group. Example:

```
$ flashgrid-dg add-disks -G MYDG -d /dev/flashgrid/rac[12].lun[3-5]
```

11.5 Removing disks

To remove disks from a running cluster

1. Determine FlashGrid names of the disks to be removed, e.g. *rac2.lun3*

```
$ flashgrid-cluster drives
```

2. If the drives are members of an ASM disk group then drop the drives from the disk group. Example:

```
SQL> alter diskgroup MYDG  
drop disk RAC1$LUN3  
drop disk RAC2$LUN3  
rebalance wait;
```

3. Prepare the drives for removal. Example:

```
[fg@rac1 ~] $ sudo flashgrid-node stop-target /dev/flashgrid/rac1.lun3  
[fg@rac2 ~] $ sudo flashgrid-node stop-target /dev/flashgrid/rac2.lun3
```

4. Detach the disks from the VMs.

11.6 Re-adding a lost disk

ASM will drop a disk from a disk group if the disk stays offline for longer than the disk repair time. If the disk was taken offline because of an intermittent problem, for example a network problem, then you can re-add such disk to the disk group. Force option must be used for re-adding such disk because it already contains ASM metadata.

Example of re-adding a regular disk:

```
$ flashgrid-dg add-disks -G MYDG -d /dev/flashgrid/rac2.lun3 -f
```

Example of re-adding a quorum disk:

```
$ flashgrid-dg add-disks -G MYDG -q /dev/flashgrid/racq.lun2 -f
```


11.7 Updating FlashGrid software

The following procedure applies to minor updates. Minor updates are those that have the same first two numbers in the version number, for example, from 17.05.31 to 17.05.50. However, update from 17.05 to 17.09 is considered major and may require a different procedure. Contact FlashGrid support for assistance if you need to do a major version update.

To update FlashGrid software on a running cluster repeat the following steps on each node, one node at a time

1. Make sure there are no other nodes that are in offline or re-syncing state. All disk groups must have zero offline disks and *Resync = No*.

```
# flashgrid-cluster
```

2. If the node is a database node,
 - a. Stop all local database instances running on the node.
 - b. Stop Oracle CRS:

```
# crsctl stop crs
```

3. Stop the FlashGrid Storage Fabric services:

```
# flashgrid-node stop
```

4. Stop the FlashGrid Cloud Area Network service:

```
# systemctl stop flashgrid-clan
```

5. Update the *flashgrid-sf*, *flashgrid-clan*, *flashgrid-diags* RPMs using yum or rpm tool

6. Start the FlashGrid Cloud Area Network service:

```
# systemctl start flashgrid-clan
```

7. Start the FlashGrid Storage Fabric service

```
# flashgrid-node start
```

8. Restart the FlashGrid Monitoring service:

```
# systemctl restart flashgrid-node-monitor
```

9. If the node has ASM installed on it then start Oracle services:

```
# systemctl start oracle-ohasd
```

```
# crsctl start crs -wait
```

10. Wait until all disks are back online and resyncing operations complete on all disk groups before updating the next node. All disk groups must have zero offline disks and *Resync = No*.

```
# flashgrid-cluster
```

11.8 Updating Linux kernel

Before updating the Linux kernel confirm that a LIS (Azure Linux Integration Services) kmod RPM for the new kernel is available. Failure to install the updated LIS kmod RPM may result in significant network performance degradation.

To update Linux kernel on a running cluster repeat the following steps on each node, one node at a time

1. Install the new kernel and the corresponding kernel-devel package

```
# yum update kernel kernel-devel
```

2. Install LIS kmod package for the new kernel
3. Follow the steps for rebooting a node

12 FlashGrid Tools and Commands

12.1 flashgrid-ca

FlashGrid Configuration Assistant helps configure a new FlashGrid cluster in a few easy steps.

Usage: flashgrid-ca [-h] [-f]

Options:

-h Show this help message and exit
-f Ignore terminal size check

12.2 flashgrid-fix-grid-dg-ca

Assistant tool for replacing temporary disks with permanent disks after GI installation.

Usage: flashgrid-fix-grid-dg-ca [-h] [-f]

Options:

-h Show this help message and exit
-f Ignore terminal size check

12.3 flashgrid-create-dg

FlashGrid Disk Group Configuration Assistant helps configure ASM disk groups with FlashGrid disks in a few easy steps.

Usage: flashgrid-create-dg [-h] [-f]

Options:

-h Show this help message and exit
-f Ignore terminal size check

12.4 flashgrid-cluster

CLI tool for cluster-level monitoring and management of all cluster nodes and disk groups.

Usage: flashgrid-cluster [command]

Commands:

show	Show cluster status (default if no command provided)
drives	List all drives in the cluster
net	List all NICs in the cluster
verify	Verify FlashGrid cluster configuration
deploy-config	Deploy configuraiton from /etc/flashgrid.cfg on the local node to all nodes
deploy-config-local	Deploy configuraiton from /etc/flashgrid.cfg on the local node to the local node only

fix-readlocal Fix Read-Local settings on all ASM instances

fix-disk-names Generate ASM SQL for fixing ASM disk names if any are invalid

set-email-alerts Set a list of email alert recipients

strict-read-local-enable
Enable strict Read-Local mode. Database auto-start is delayed until all local disks are online.

strict-read-local-disable
Disable strict Read-Local mode and restore original database startup dependencies.

strict-read-local-show
For all databases show whether strict Read-Local mode is enabled or disabled.

help Show this help message and exit

12.5 flashgrid-node

CLI tool for monitoring and management of the local node.

Usage: flashgrid-node [-f] [command]

Commands:

show Show node details (default if no command provided)

test-alerts Send test email alert.

collect-diags Create diagnostics archive file for the local node.

stop Gracefully offline the local failgroup and stop flashgrid services. '-f' option forces stop even if graceful offline fails.

start Start flashgrid service and online local disks in ASM

reboot Gracefully offline the local failgroup and reboot the node. '-f' option forces reboot even if graceful offline fails.

poweroff Gracefully offline the local failgroup and power the node off. '-f' option forces reboot even if graceful offline fails.

offline Gracefully offline the local failgroup.

online Bring the local failgroup online for each mounted disk group.

stop-waiting Stop waiting for all storage nodes and allow Oracle services to start.

stop-target	Gracefully offline the corresponding disk in ASM and stop the target. '-f' option forces target stop even if graceful offline fails.
reload-config	Reload flashgrid config without restart. Note that network configuration is not reloaded.
restart-services	Gracefully offline the local failgroup, restart flashgrid service and online local disks in ASM. '-f' option forces restart even if graceful offline fails.
restart-targets	Gracefully offline the local failgroup, restart flashgrid targets service, and online local disks in ASM. '-f' option forces target service restart even if graceful offline fails.
restart-initiators	If there are no mounted ASM disk groups using FlashGrid disks then restart flashgrid initiators service. '-f' option forces restart even if there are mounted groups.
restart-asm-connector	Restart the FlashGrid ASM connector service.
restart-cluster-connector	Restart FlashGrid cluster connector service.
verify-config	Verify validity of the local config file /etc/flashgrid.cfg
shell	Start shell (for internal and tech support use).
asm-state	Show ASM state in JSON format (for internal use).
help	Show this help message and exit

12.6 flashgrid-dg

CLI tool for monitoring and management of individual disk groups.

Usage: flashgrid-dg [command]

Commands:

list	List all FlashGrid disk groups (default if no command provided).
show	Show details of the disk group.
create	Create a new ASM disk group.
add-disks	Add disks to the disk group.
replace-disk	Replace a disk with a new one.
help	Show this help message and exit

12.7 flashgrid-diags

CLI tool for creating a diagnostics archive file.

Usage: flashgrid-diags {--all | --local}

Options:

- `--all` Collect diags from all nodes of the cluster in a single file.
- `--local` Collect diags from the local node only.

13 Troubleshooting

The following troubleshooting steps are recommended in case of any issues with FlashGrid cluster configuration or operation:

1. Check status of all FlashGrid nodes, network, and disk groups by running 'flashgrid-cluster' on any node
2. For any ASM disk group that has a *Warning* or *Critical* status, run the following command to check the list nodes where the disk group is mounted and the list of disks with their status:

```
$ flashgrid-dg show -G <DGNAME>
```

3. On any node that has a *Warning*, *Critical*, or *Inaccessible* status:
 - a. Check whether the FlashGrid service is active:

```
# systemctl status flashgrid
```
 - b. Check status of NICs, local disks, and remote disks:

```
# flashgrid-node
```
 - c. Check that the configuration has no errors:

```
# flashgrid-node verify-config
```
4. If network verification fails then run 'flashgrid-cluster verify' to get more detailed information
5. Check FlashGrid log files on the affected nodes. The log files are located in /opt/flashgrid/log

14 Contacting FlashGrid Technical Support

For help with troubleshooting an issue on an existing FlashGrid cluster please use Technical Support Request form located at <https://www.flashgrid.io/support/>

To expedite troubleshooting please also collect diagnostic data by running 'flashgrid-diags --all' command on any node and upload it using a secure upload form provided to your company by FlashGrid technical support.

Customers with Mission-Critical SLA subscription may also use the 24x7 telephone hotline for reporting *emergency* type of issues that require immediate attention: +1-650-641-2421 ext 7. Please note that use of the 24x7 hotline is reserved for emergency situations only.

Copyright © 2016-2018 FlashGrid Inc. All rights reserved.

This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document.

FlashGrid is a registered trademark of FlashGrid Inc. Amazon and Amazon Web Services are registered trademarks of Amazon.com Inc. and Amazon Web Services Inc. Oracle and Java are registered trademarks of Oracle and/or its affiliates. Red Hat is a registered trademark of Red Hat Inc. Other names may be trademarks of their respective owners.