

# FlashGrid

## Provisioning and Maintenance Guide for Oracle RAC in AWS

rev. 18.08-2018.10.04



# Table of Contents

1	Introduction .....	3
2	Compatibility .....	3
3	Getting Access to FlashGrid AMI .....	3
4	Creating AMI with Encryption of Boot Volume .....	3
5	Uploading Oracle Installation Files to S3 .....	4
6	Preparing the VPC .....	5
7	Provisioning a Cluster .....	6
7.1	Creating cluster with CloudFormation .....	6
7.2	Verifying cluster health .....	7
7.3	OS User Accounts .....	7
7.4	Changing ASM passwords .....	8
7.5	Enabling instance termination protection .....	8
7.6	Deleting a cluster .....	8
8	Creating a Database .....	9
8.1	Creating a database using DBCA .....	9
8.2	Configuring database memory settings .....	9
9	Enabling Strict Read-Local Mode for a New Database .....	9
10	Measuring Performance .....	11
11	Connecting Database Clients .....	12
11.1	Configuring Alias for Local Listener .....	13
11.2	Setting LOCAL_LISTENER parameter for each database .....	13
11.3	Adding DNS Records .....	14
11.4	Opening listener ports .....	14
11.5	Testing client connectivity .....	14
11.6	Configuring client connect string .....	15
12	Monitoring Cluster Health .....	16
13	Backup and Restore Best Practices .....	17
13.1	Backing up OS and software on the cluster nodes .....	17
13.2	Restoring root or software volume of a cluster node .....	17
13.3	Restoring an instance that was accidentally terminated .....	18
13.4	Backing up and restoring database files .....	18
13.5	Backing up and restoring files on ACFS .....	19
14	Maintenance Tasks .....	19
14.1	Rebooting a node .....	19
14.2	Powering off a node .....	19
14.3	Shutting down an entire cluster .....	19
14.4	Adding EBS volumes for use in ASM .....	20
14.5	Removing EBS volumes .....	20
14.6	Re-adding a lost disk .....	21
14.7	Updating FlashGrid software .....	21
14.8	Updating Linux kernel .....	22
15	FlashGrid Tools and Commands .....	23
15.1	flashgrid-ca .....	23
15.2	flashgrid-fix-grid-dg-ca .....	23
15.3	flashgrid-create-dg .....	23
15.4	flashgrid-cluster .....	23
15.5	flashgrid-node .....	24
15.6	flashgrid-dg .....	25
15.7	flashgrid-diags .....	26
16	Troubleshooting .....	26
17	Contacting FlashGrid Technical Support .....	26

# 1 Introduction

FlashGrid Cloud Area Network software and FlashGrid Storage Fabric software enable running Oracle RAC clusters in AWS 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 AWS 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 on Amazon EC2 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
- Oracle Grid Infrastructure: ver. 12.2.0.1 or 12.1.0.2
- Operating System: Oracle Linux 7, Red Hat Enterprise Linux 7

The solution can be deployed on the following Amazon EC2 instance types: M4, M5, R4, i3, X1, X1E.

## 3 Getting Access to FlashGrid AMI

To be able to create cluster your AWS account must have a subscription to a FlashGrid AMI via AWS Marketplace. FlashGrid AMIs are based on either Oracle Linux 7 or RHEL 7. Please contact your FlashGrid representative if you need to customize the AMI.

### To get access to the FlashGrid AMI

1. Open [FlashGrid product page](#) at AWS Marketplace
2. Click **Continue** button
3. Select **Manual Launch** tab
4. Click **Accept Software Terms** button

## 4 Creating AMI with Encryption of Boot Volume

The FlashGrid Cloud Provisioning tool has an option for enabling encryption on EBS volumes. However, this option does not cover the boot volume. If EBS encryption must be enabled on the boot volume too then you need to create a new encrypted AMI based on the FlashGrid AMI.

### To create an encrypted AMI

1. Launch a single instance of t2.micro (or any other) type using the FlashGrid AMI.
2. Stop the instance.
3. Create a new AMI from the instance and name it "*FlashGrid AMI not encrypted YY-MM-DD*".

4. Copy the "FlashGrid AMI not encrypted YY-MM-DD" AMI to "FlashGrid AMI Encrypted YY-MM-DD" and enable **Encryption** option when copying.

## 5 Uploading Oracle Installation Files to S3

During cluster provisioning Oracle installation files will be downloaded from an S3 bucket. The list of files that must be placed in the S3 bucket is available at <https://www.flashgrid.io/cloud-provisioning-for-aws/>

Two options are available for allowing access to the files in the S3 bucket for the cluster node instances:

- Enabling public access to each file for the duration of cluster provisioning  
OR
- Assigning the cluster node instances an IAM role that has permissions for accessing files in the bucket

### To allow public access to the files in S3

1. Create an S3 bucket/folder for uploading the installation files
2. Upload the required files to the S3 bucket/folder
3. In S3 Management Console navigate to the bucket and the folder to see the list of files
4. Select all files
5. Click **More** -> **Make Public**
6. You can disable public access after the cluster completes initialization

### To use an IAM role for access to the files in S3

1. Create an S3 bucket/folder for uploading the installation files
2. Upload the required files to the S3 bucket/folder
3. In IAM Management Console create a new policy named **GetOracleFilesFromS3** that allows `s3:GetObject` action on all uploaded files. See an example below.
4. In IAM Management Console create a new role named **GetOracleFilesFromS3** and attach the **GetOracleFilesFromS3** policy to it.
5. Use the **GetOracleFilesFromS3** role when configuring cluster parameters in the FlashGrid Cloud Provisioning tool.

### Example of an IAM policy:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Stmt1508867055000",
      "Effect": "Allow",
      "Action": [
        "s3:GetObject"
      ],
      "Resource": [
        "arn:aws:s3:::mybucket/mydirectory/*"
      ]
    }
  ]
}
```

# 6 Preparing the VPC

When creating a new cluster you have two options:

- **Automatically create a new VPC.**  
This option is usually used for test clusters isolated in their own sandbox VPCs. A VPC will be created together with the required subnets, placement group(s), and security groups. By default the VPC will be created with CIDR 10.100.0.0/16
- **Create the cluster in an existing VPC.**  
This option is used for majority of production deployments where other systems (e.g. app servers) share the same VPC as the cluster. You will need to provide the VPC ID in the Cloud Provisioning tool and subnet IDs and security group IDs in the CloudFormation Manager.

If using an existing VPC then make sure that the following pre-requisites are met before creating a cluster:

- The VPC has a subnet in each of the availability zones used for the cluster nodes.
- The VPC has an S3 endpoint configured (required unless public IPs can be enabled for access to S3)
- The VPC has a security group with the following ports open for inbound traffic:
  - UDP ports 4801, 4802, 4803 between any members of the security group
  - TCP ports 1521, 1522 for SCAN and Local Listener access to the database nodes from app servers and other database clients. These are default port numbers that can be changed in the Cloud Provisioning tool.
  - TCP port 22 for SSH access to the cluster nodes
  - TCP port 5901 if you choose to use VNC for creating a database using DBCA in GUI mode
- Create a *Spread Placement Group* that will be used for the VM instances. The use of the spread placement group is not required only if each VM is placed in a separate availability zone, e.g. 2 db nodes + 1 quorum node across three AZs, or 3 db nodes + 2 quorum nodes across five AZs. It is required in all other cases to guarantee that all nodes are running on separate physical hosts.

# 7 Provisioning a Cluster

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

- Creating and configuring EC2 VPC, subnets, security group (optional)
- Creating EBS volumes and launching EC2 instances for all nodes in the cluster
- Installing and configuring FlashGrid Cloud Area Network
- Installing and configuring FlashGrid Storage Fabric
- Installing and patching Oracle Grid Infrastructure software
- Configuring Grid Infrastructure cluster
- Installing and patching Oracle Database software
- Creating ASM disk groups

## 7.1 Creating cluster with CloudFormation

### To create a cluster

1. Log in to AWS Management Console with a user account that has the following privileges:
  - AWSCloudFormationFullAccess
  - AmazonEC2FullAccess
  - AmazonVPCFullAccess (required only if creating a new VPC)
2. Open FlashGrid Cloud Provisioning tool:
  - Start with one of the standard configurations at <https://www.flashgrid.io/cloud-provisioning-for-aws>
  - or, if you have a custom configuration file, upload it at <https://1808.cloudprov.flashgrid.io>
3. Configure parameters of the cluster
4. Click *Verify* button
5. If verification passes then click *Create Cluster* button, which will take you to AWS CloudFormation Manager
6. Click *Next*
7. Select your SSH key
8. If using an existing VPC, then select a spread placement group, subnet(s), and security group.
9. Click *Next*
10. Click *Next* (if you added tags at the cluster configuration page then do not add the same tags in CloudFormation Manager)
11. Click *Create*
12. Wait until the status of the stack changes to *CREATE\_COMPLETE*
13. If creating the stack fails:
  - a) Check for the cause of the failure on the *Events* tab
  - b) Correct the cause of the error
  - c) Delete the failed stack
  - d) Repeat the steps for creating a new stack
14. Get IP addresses of the newly launched instances on the *Outputs* tab
15. SSH to the first (as it was specified on the cluster configuration page) cluster node as user `fg@`
16. 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.
17. 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.

## 7.2 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 18.07.10.46032 #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
-----
```

## 7.3 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.

## 7.4 Changing ASM passwords

A temporary password for ASM users SYS and ASMSNMP is configured during cluster initialization. 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";
```

## 7.5 Enabling instance termination protection

It is strongly recommended to enable instance termination protection for all cluster nodes.

## 7.6 Deleting a cluster

### To delete a cluster

1. Disable instance termination protection for all cluster nodes
2. Open AWS CloudFormation Manager console
3. Delete the stack corresponding to the cluster

## 8 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.

### 8.1 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.

### 8.2 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.

## 9 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
```

# 10 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 i3.16xlarge instances and eight NVMe SSDs per node

```
SET SERVEROUTPUT ON;
DECLARE
  lat INTEGER;
  iops INTEGER;
  mbps INTEGER;
BEGIN DBMS_RESOURCE_MANAGER.CALIBRATE_IO (16, 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 = 1375694
Latency = 0
Max_MB/s = 27338

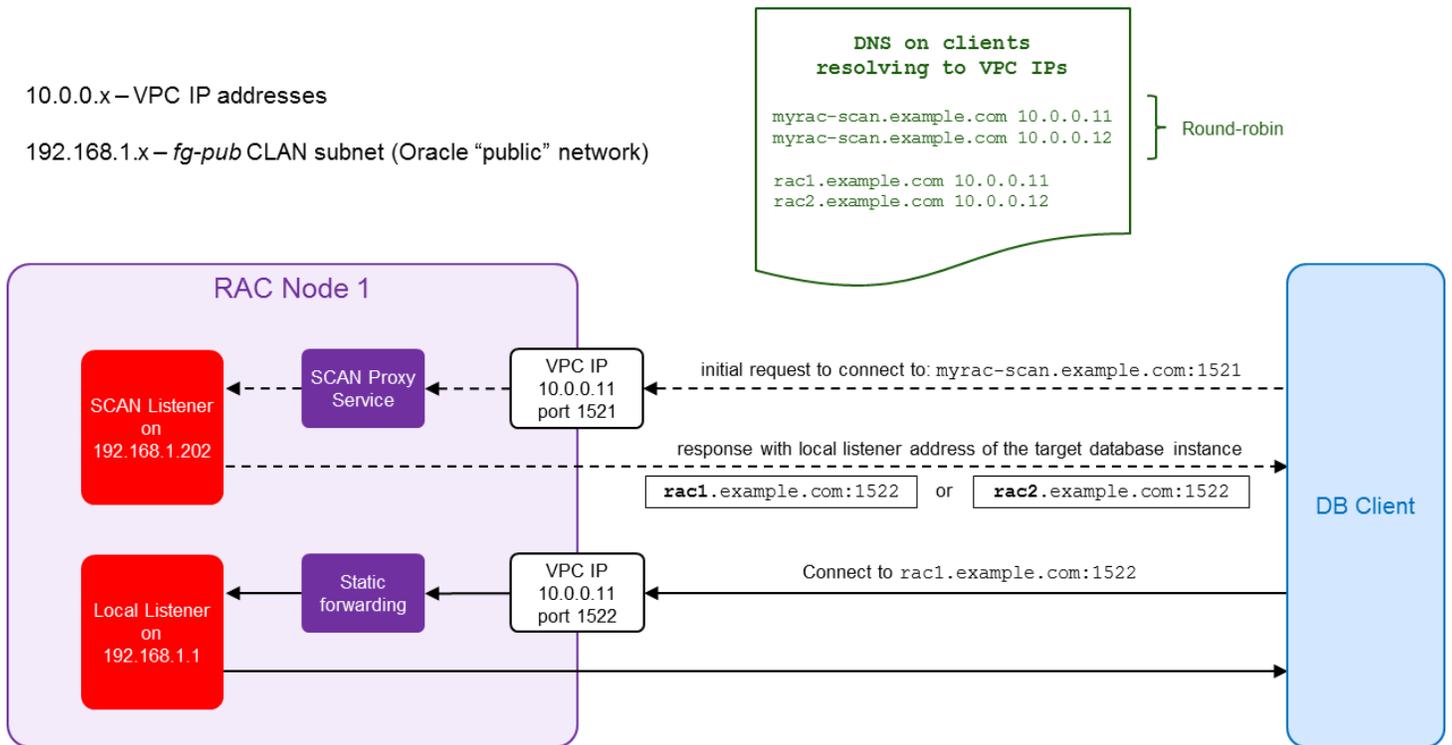
PL/SQL procedure successfully completed.
```

# 11 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 VPC where the clients are located. FlashGrid SCAN Proxy service running on each database node of the cluster facilitates connectivity between the clients on the VPC and the SCAN listeners on the *fg-pub* subnet. Port forwarding facilitates connectivity between the clients on the VPC 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 VPC 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*

## 11.1 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).

## 11.2 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.

## 11.3 Adding DNS Records

On the DNS server(s) used by clients, for each database node add two records resolving to the VPC 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.

## 11.4 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.

## 11.5 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
```

## 11.6 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>

## 12 Monitoring Cluster Health

The following methods of monitoring cluster health are available:

- The *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 3<sup>rd</sup>-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.

# 13 Backup and Restore 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

## 13.1 Backing up OS and software on the cluster nodes

It is strongly recommended to back up the OS and software volumes of all cluster nodes after the initial cluster configuration and before and after applying any changes such as patch installation or security settings changes. OS and software on a cluster node can be backed up by creating an AMI for the instance or by creating snapshots of individual EBS volumes. AMI-based backup is recommended as it allows easier way to recover an instance in case it is terminated because of a failure or a human error.

### To create a backup AMI for a cluster node

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

```
# crsctl stop crs
```

2. Gracefully stop the instance.

```
# flashgrid-node poweroff
```

3. Create AMI for the instance.

IMPORTANT! Remove data volumes from the list of volumes that will be backed up to the AMI. Typically only the root volume (`/dev/sda1`) and the software volume (`/dev/xvdz`), if present, must be included in the backup AMI. Other volumes must be excluded. Failure to exclude data volumes from the AMI will create inconsistency in ASM disk groups if the data volumes are later recovered from the AMI. Data volumes should never be backed up or recovered at the volume level. Instead, database/ACFS file backup procedures must be used as described further in this section.

## 13.2 Restoring root or software volume of a cluster node

Root volume of an instance is an EBS volume that has the OS installed on it. Device name of the root volume is `/dev/sda1`.

Software volume of an instance is an EBS volume where Oracle software binaries are installed (contains the `/u01` directory). Typically the software volume has device name `/dev/xvdz`.

The root volume or the software volume may need to be restored in case the volume fails, has file system corruption, or has logical corruption.

### To recover root volume (sda1) or software volume (xvdz)

1. In the backup AMI for the affected instance identify the snapshot id for the affected volume
2. Using the snapshot id, create a new volume in the availability zone where the affected instance is located
3. Stop the affected instance
  - If the OS is running then stop the instance gracefully using `flashgrid-node poweroff` command
  - If the OS is not running then stop the instance using AWS console or CLI

4. Detach the affected volume from the instance
5. Attach the newly created volume to the instance using the same device name (*/dev/sda1* for root volume, */dev/xvdz* for the software volume)
6. Start the instance

## 13.3 Restoring an instance that was accidentally terminated

Setting instance termination protection is strongly recommended to prevent accidental termination of a cluster node instance.

### To restore an instance that was terminated

1. Launch a new instance using the backup AMI for the instance that was terminated
  - Make sure that correct instance type, VPC, subnet, and security group are selected.
  - Configure the same Primary IP that was used on the terminated instance
  - If the cluster is in a single availability zone then also select the placement group corresponding to the cluster
  - Make sure that only */dev/sda1* and */dev/xvdz* (only on database nodes) volumes are configured. Remove any other volumes if they are present in the AMI.

2. Attach data volumes to the new instance using the same device names (such as *xvdba*) that were previously used

3. Log in to the instance and bring the data disks online:

```
$ flashgrid-node online
```

## 13.4 Backing up and restoring database files

Use standard RMAN procedures for backing up and restoring database files. The two recommended options for backup storage destination are:

- **Amazon S3.** Provides maximum flexibility with easy shared access to the backup files.
- **An EBS volume with a local file system.** Provides maximum performance, with up to 500 MB/s of read/write bandwidth on a *st1* type of volume.

For information about backing up to S3 see the following documentation from Oracle and AWS:

- [White paper: Oracle Database Backup To Cloud: Amazon Simple Storage Service \(S3\)](#)
- [Getting Started with Oracle Secure Backup and RMAN with Amazon S3](#)
- [Oracle Secure Backup \(OSB\) Cloud Module](#)

### To configure an EBS volume as a backup storage destination

1. Create an EBS volume in the availability zone where the instance running RMAN is located. *st1* volume type is recommended.
2. Attach the volume to the instance running RMAN. Select a device name in the *xvdc* to *xvdg* range - disks in this name range will be treated as local and will not be shared by FlashGrid Storage Fabric.
3. Format the volume with a local file system (XFS recommended) and create a mount point for it.
4. Use standard RMAN procedures to configure backup to the local file system.

Note that an EBS volume can be moved only between instances in the same availability zone. However, snapshot of the volume can be used to clone the volume to a different availability zone.

## 13.5 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.

# 14 Maintenance Tasks

## 14.1 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.

## 14.2 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. Power off the node using `flashgrid-node` command. It will gracefully put the corresponding failure group offline.

```
# flashgrid-node poweroff
```

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

## 14.3 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. Power all nodes off.

## 14.4 Adding EBS volumes for use in ASM

When adding new volumes 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 EBS volumes in a running cluster

1. Create and attach new volumes to the database node instances. Attach the volumes using disk names *xvdba* through *xvdbz* or *xvdca* through *xvdcz* - these disks will be shared automatically by FlashGrid Storage Fabric.

**Note:** Clusters deployed before 2018-09-01 may require configuration update to use the *xvdc* - *xvdcz* disks. Please contact FlashGrid support for help with updating the configuration.

2. Confirm FlashGrid names of the new drives, e.g. *rac2.xvdbb*

```
$ 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].xvdb[a-c]
```

## 14.5 Removing EBS volumes

### To remove EBS volumes from a running cluster

1. Determine FlashGrid names of the drives to be removed, e.g. *rac2.xvdbb*

```
$ 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$XVDBB  
drop disk RAC2$XVDBB  
rebalance wait;
```

3. Prepare the drives for removal. Example:

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

4. Detach the EBS volumes from the instances.

## 14.6 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.xvdbb -f
```

Example of re-adding a quorum disk:

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

## 14.7 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 17.10 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
```

## 14.8 Updating Linux kernel

**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. Follow the steps for rebooting a node

# 15 FlashGrid Tools and Commands

## 15.1 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

## 15.2 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

## 15.3 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

## 15.4 flashgrid-cluster

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

Usage: flashgrid-cluster [command]

Commands:

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

## 15.5 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.

<b>stop-target</b>	Gracefully offline the corresponding disk in ASM and stop the target. '-f' option forces target stop even if graceful offline fails.
<b>reload-config</b>	Reload flashgrid config without restart. Note that network configuration is not reloaded.
<b>restart-services</b>	Gracefully offline the local failgroup, restart flashgrid service and online local disks in ASM. '-f' option forces restart even if graceful offline fails.
<b>restart-targets</b>	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.
<b>restart-initiators</b>	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.
<b>restart-asm-connector</b>	Restart the FlashGrid ASM connector service.
<b>restart-cluster-connector</b>	Restart FlashGrid cluster connector service.
<b>verify-config</b>	Verify validity of the local config file /etc/flashgrid.cfg
<b>shell</b>	Start shell (for internal and tech support use).
<b>asm-state</b>	Show ASM state in JSON format (for internal use).
<b>help</b>	Show this help message and exit

## 15.6 flashgrid-dg

CLI tool for monitoring and management of individual disk groups.

Usage: flashgrid-dg [command]

Commands:

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

## 15.7 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.

## 16 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

## 17 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.