

Exadata 运维实践-日常维护管理

COE

郭俊龙

May, 2020

扫码加入：
19c新特性讲座群



欢迎关注：
甲骨文云技术公众号

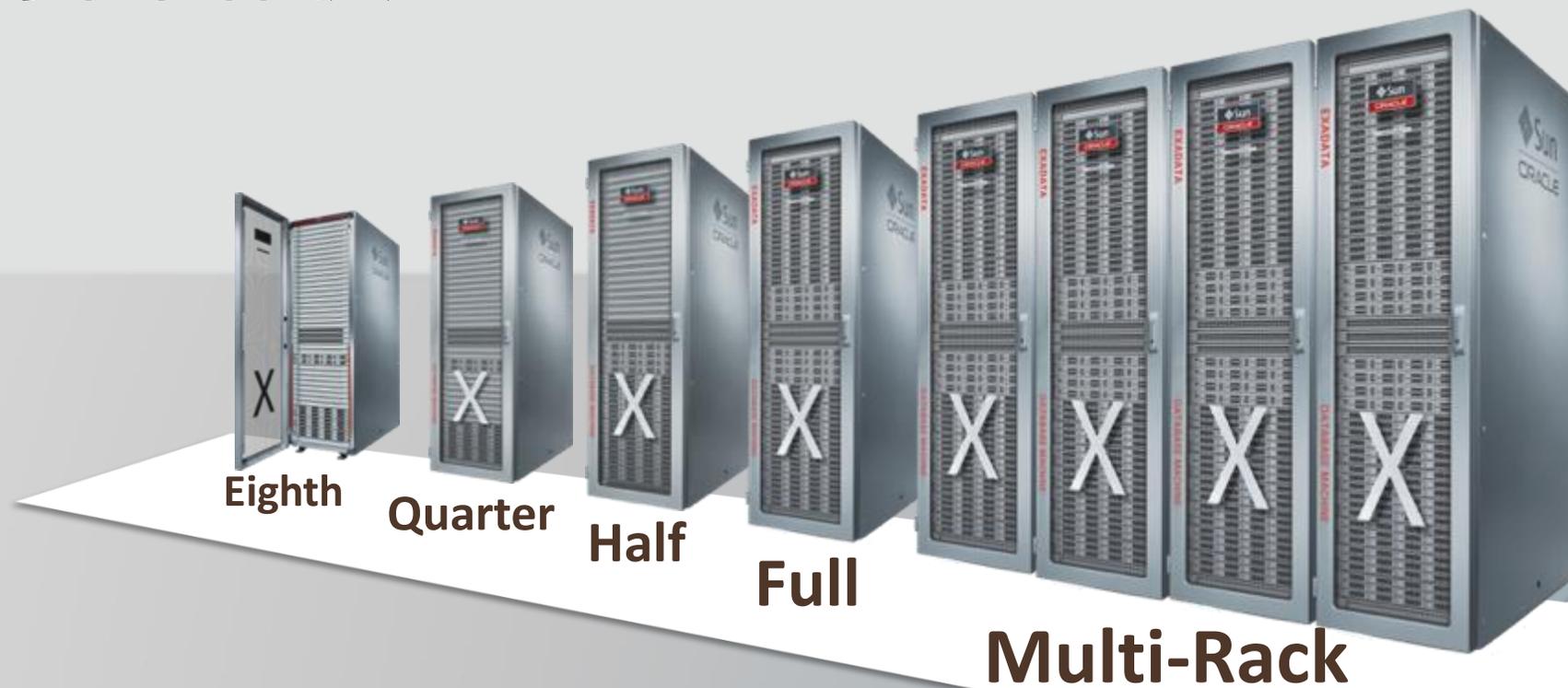


◆ Exadata 运维管理

- 1 Exadata 组件管理
- 2 Exadata 状态检查
- 3 Exadata 管理工具
- 4 Exadata 案例分享

Exadata 硬件介绍:

从第八分之一配置向多机架弹性扩展



说明:

Exadata开箱即用;

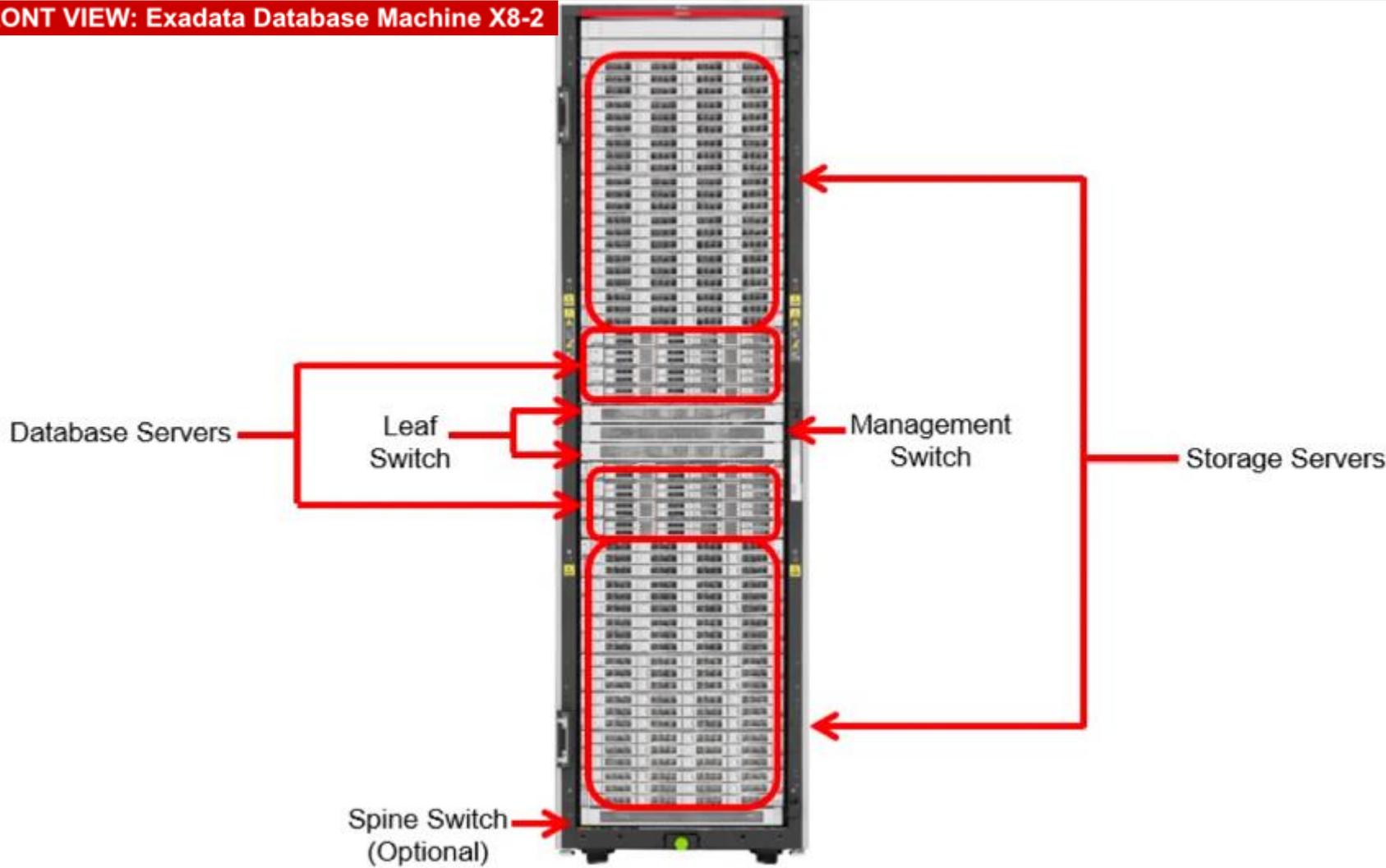
从2个数据库服务器(db server)和3个存储服务器(cell server)起步;
根据需要添加数据库或存储服务器。



Exadata 硬件介绍:

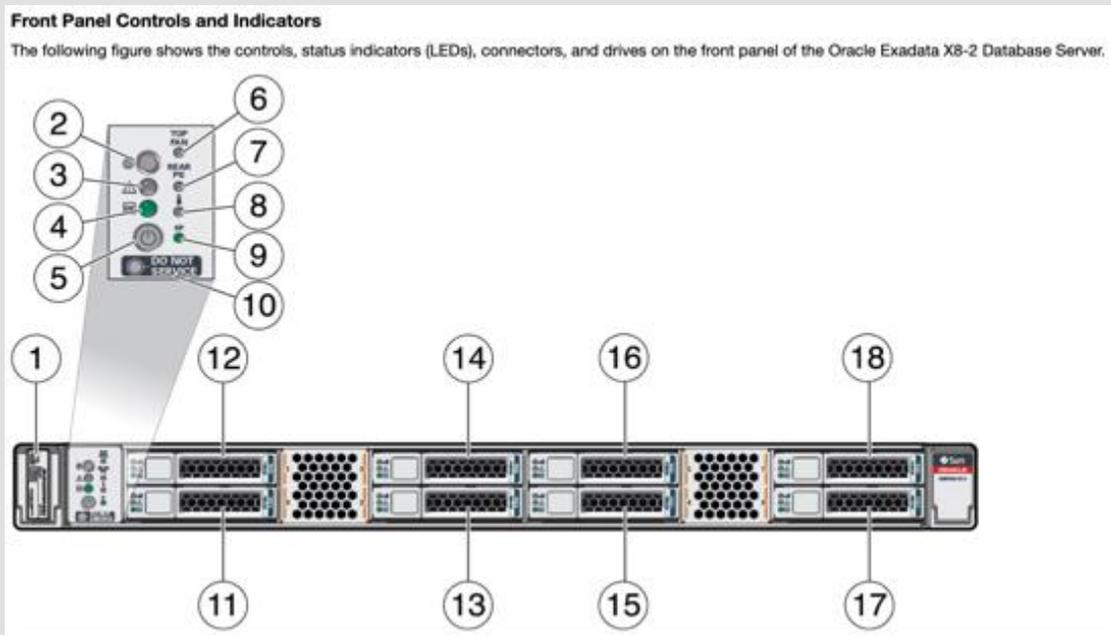
Exadata X8-2 正面/背面视图:

FRONT VIEW: Exadata Database Machine X8-2



Exadata 硬件介绍:

数据库服务器(db server)正面视图说明:



说明:

- OK to Remove 指示灯为蓝色：在热插拔过程中可以安全地卸下存储驱动器。
- Fault-Service Required LED：系统正在运行，但存储驱动器出现故障。如果系统检测到存储驱动器故障，则前面板和后面板故障-需要维修的指示灯也会亮起。
- OK/Activity（确定/活动）指示灯呈绿色：数据正在从存储驱动器读取或写入。

Call Out	Description
1	Product Serial Number (PSN) label and Radio Frequency Identification (RFID) tag
2	Locate Button/LED: white
3	Fault-Service Required LED: amber
4	System OK LED: green
5	On/Standby button
6	Fault-Service Required LED: Top: Fan Module (amber)
7	Fault-Service Required LED: Rear: Power Supply (amber)
8	Fault-Service Required LED: Overtemp Icon: System Over Temperature Warning (amber)
9	SP OK LED: green
10	DO NOT SERVICE LED: white
11	Storage drive 0 HDD
12	Storage drive 1 HDD
13	Storage drive 2 HDD
14	Storage drive 3 HDD
15	Storage drive 4 HDD
16	Storage drive 5 HDD
17	Storage drive 6 HDD
18	Storage drive 7 HDD

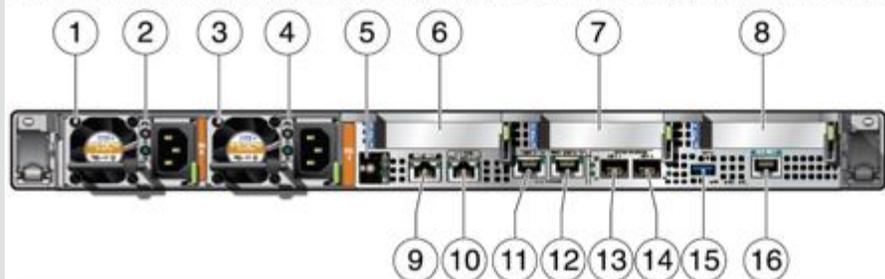


Exadata 硬件介绍:

数据库服务器(db server)背面视图说明:

Back Panel Components and Cable Connections

The following figure shows the Oracle Exadata X8-2 Database Server back panel and the location of power supplies, status indicators (LEDs), connectors, and PCIe slots.



说明:

- 业务网络是通过数据库服务上的网卡来提供，通常使用绑定模式：主备绑定模式或链路聚合模式，依赖于上连交换机的配置。

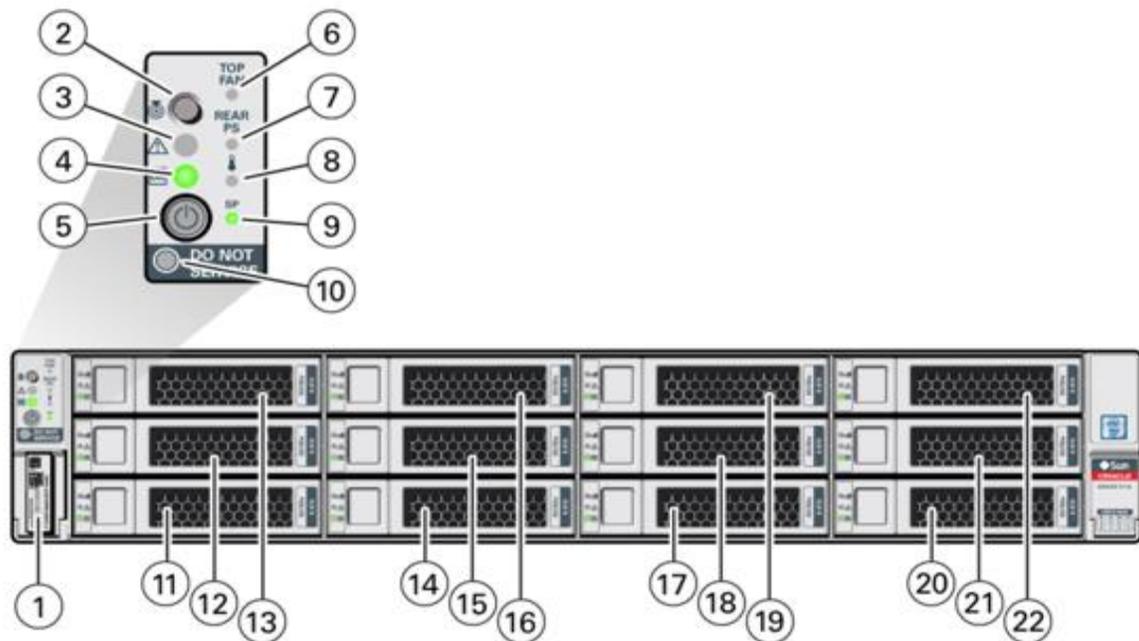
Call Out	Description
1	Power Supply (PS) 0
2	PS 0 status indicators: Fault-Service Required LED: amber; AC OK LED: green
3	Power Supply (PS) 1
4	PS 1 status indicators: Fault-Service Required LED: amber; AC OK LED: green
5	System status indicators: Locate Button/LED: white; Fault-Service Required LED: amber; System OK LED: green
6	PCIe card slot 1 (Oracle Dual Port 25 Gb Ethernet Adapter)
7	PCIe card slot 2 (Oracle Dual Port QDR InfiniBand Adapter M3 or Oracle Dual Port 100 Gb Ethernet Adapter)
8	PCIe card slots 3 (Optional Oracle Dual Port 25 Gb Ethernet Adapter) and 4 (Oracle Storage 12Gb SAS PCIe RAID HBA, Internal card)
9	Oracle Integrated Lights Out Manager (ILOM) service processor (SP) network management (NET MGT) RJ-45 10/100/1000BASE-T port
10	Network (NET) 100/1000BASE-T RJ-45 Gigabit Ethernet (GbE) port: NET 0
11	Network (NET) 10GBASE-T RJ-45 GbE port: NET 1
12	Network (NET) 10GBASE-T RJ-45 GbE port: NET 2
13	Network (NET) 10/25GbE enhanced small form-factor pluggable (SFP+) port: NET 1
14	Network (NET) 10/25GbE SFP+ port: NET 2
15	USB 3.0 connector
16	Serial management (SER MGT) RJ-45 serial port



Exadata 硬件介绍:

HC/XT 存储服务器(Cell server)正面视图说明:

HC and XT Front Panel Components



说明:

- OK to Remove指示灯为蓝色：在热插拔过程中可以安全地卸下存储驱动器。
- Fault-Service Required LED指示灯呈琥珀色：系统正在运行，但存储驱动器出现故障。如果系统检测到存储驱动器故障，则前面板和后面板故障需要维修的指示灯也会亮起。
- OK/Activity（确定/活动）指示灯呈绿色：数据正在从存储驱动器读取或写入。
- Do Not Service指示灯为白色/亮起：存储服务器不能关机进行维修；否则可能导致强制卸载Oracle自动存储管理（Oracle ASM）磁盘组并影响数据可用性。
- Do Not Service指示灯熄灭：可以关闭存储服务器以进行维修。

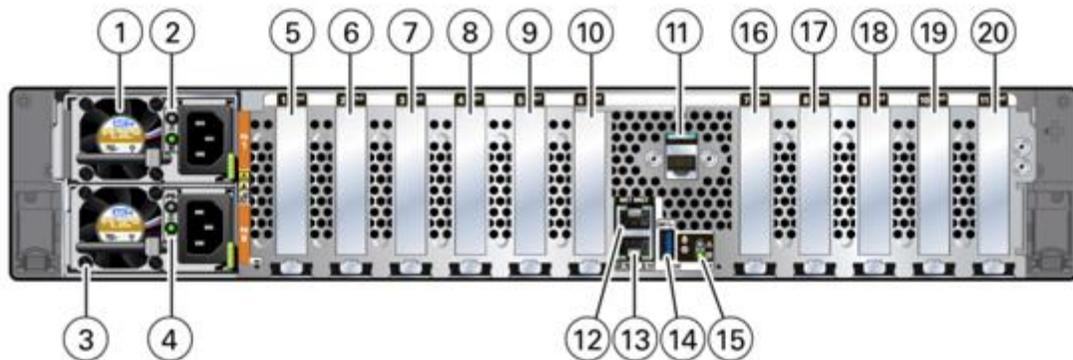
Callout	Description
1	Product Serial Number (PSN) label and Radio Frequency Identification (RFID) tag
2	Locate Button/LED: white
3	Fault-Service Action Required LED: amber
4	System OK LED: green
5	On/Standby button
6	Fault-Service Required LED: Top: Fan Module (amber)
7	Fault-Service Required LED: Rear: Power Supply (amber)
8	Fault-Service Required LED: Overtemp Icon: System Over Temperature Warning (amber)
9	SP OK LED: green
10	DO NOT SERVICE LED: white
11	Storage drive 0 HDD
12	Storage drive 1 HDD
13	Storage drive 2 HDD
14	Storage drive 3 HDD
15	Storage drive 4 HDD
16	Storage drive 5 HDD
17	Storage drive 6 HDD
18	Storage drive 7 HDD
19	Storage drive 8 HDD
20	Storage drive 9 HDD
21	Storage drive 10 HDD
22	Storage drive 11 HDD



Exadata 硬件介绍:

HC/XT存储服务器(Cell server)背面视图说明:

HC and XT Back Panel Components and Cable Connections



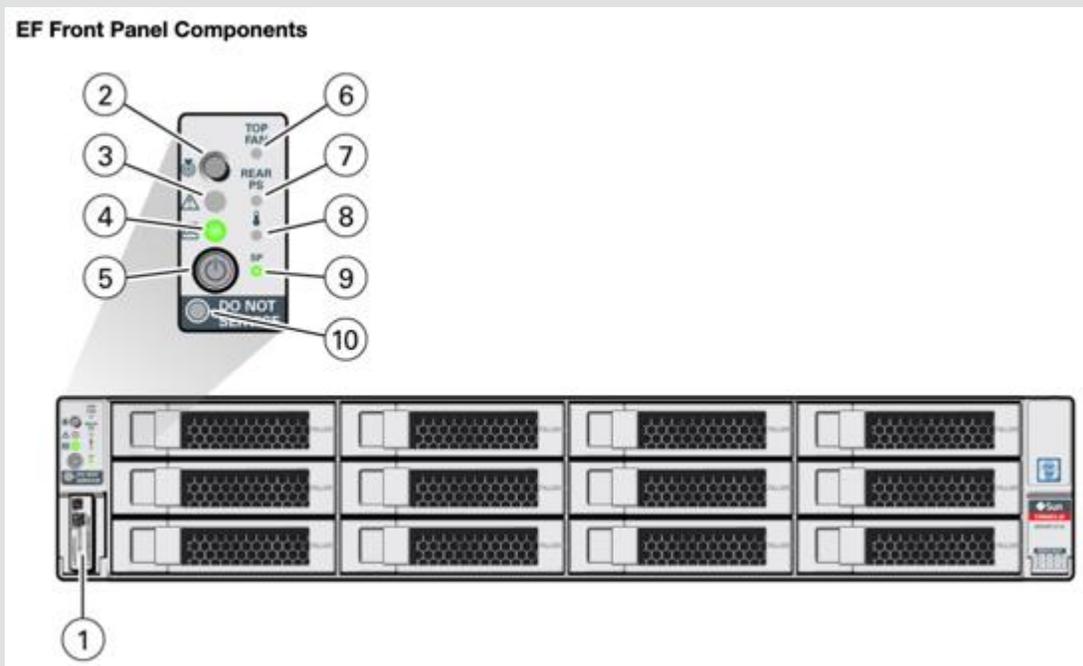
说明:

- HC存储服务器上有闪存卡, 又有大容量磁盘。
- XT存储服务器上没有闪存卡, 只有大容量磁盘。
- 不需要连接业务网络。

Callout	Description
1	Power supply (PS 1)
2	PS 1 status indicators: Fault-Service Required LED: amber; AC OK LED: green
3	Power supply (PS 0)
4	PS 0 status indicators: Fault-Service Required LED: amber; AC OK LED: green
5	PCIe slot 1 (Oracle Dual Port 100 Gb Ethernet Adapter, HC only) Note - This slot is unpopulated if the Oracle Dual Port QDR InfiniBand Adapter M3 is installed in PCIe slot 7.
6	PCIe slot 2
7	PCIe slot 3
8	PCIe slot 4 (Oracle Flash Accelerator F640, HC only)
9	PCIe slot 5 (Oracle Flash Accelerator F640, HC only)
10	PCIe slot 6 (Oracle Flash Accelerator F640, HC only)
11	Serial management (SER MGT) RJ-45 serial port
12	Oracle Integrated Lights Out Manager (ILOM) service processor (SP) network management (NET MGT) RJ-45 10/100/1000BASE-T port
13	Network (NET) 100/1000BASE-T RJ-45 Gigabit Ethernet (GbE) port: NET 0
14	USB 3.0 connector
15	System status LEDs: Locate Button/LED: white; Fault-Service Required: amber; System OK: green
16	PCIe slot 7 (Oracle Dual Port QDR InfiniBand Adapter M3) Note - This slot is unpopulated (HC only) if the Oracle Dual Port 100 Gb Ethernet Adapter is installed in PCIe slot 1.
17	PCIe slot 8
18	PCIe slot 9
19	PCIe slot 10 (Oracle Flash Accelerator F640, HC only)
20	PCIe slot 11 (Oracle Storage 12 Gb SAS PCIe RAID HBA, Internal card)

Exadata 硬件介绍:

EF存储服务器(cell server)正面视图说明:



说明:

- EF存储服务器上全部是闪存卡，没有大容量磁盘。

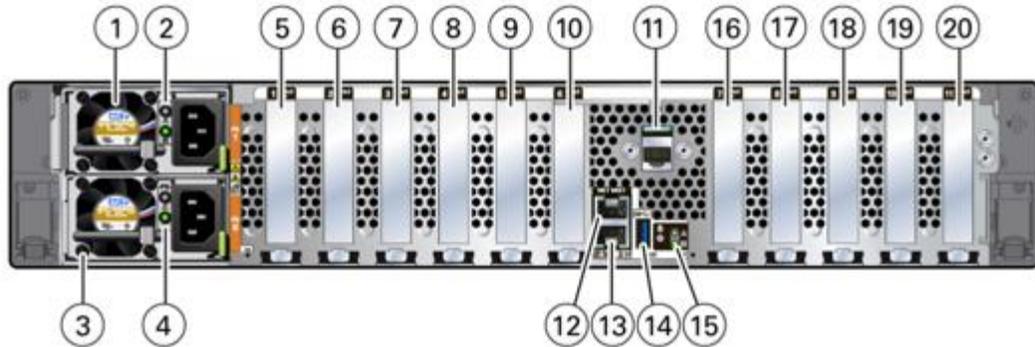
Callout	Description
1	Product Serial Number (PSN) label and Radio Frequency Identification (RFID) tag
2	Locate Button/LED: white
3	Fault-Service Action Required LED: amber
4	System OK LED: green
5	On/Standby button
6	Fault-Service Required LED: Top: Fan Module (amber)
7	Fault-Service Required LED: Rear: Power Supply (amber)
8	Fault-Service Required LED: Overtemp Icon: System Over Temperature Warning (amber)
9	SP OK LED: green
10	DO NOT SERVICE LED: white



Exadata 硬件介绍:

EF存储服务器(cell server)背面视图说明:

EF Back Panel Components and Cable Connections



说明:

- 不需要连接业务网络。

Callout	Description
1	Power supply (PS 1)
2	PS 1 status indicators: Fault-Service Required LED: amber; AC OK LED: green
3	Power supply (PS 0)
4	PS 0 status indicators: Fault-Service Required LED: amber; AC OK LED: green
5	PCIe slot 1 (Oracle Dual Port 100 Gb Ethernet Adapter) Note - This slot is unpopulated if the Oracle Dual Port QDR InfiniBand Adapter M3 is installed in PCIe slot 7.
6	PCIe slot 2 (Oracle Flash Accelerator F640)
7	PCIe slot 3 (Oracle Flash Accelerator F640)
8	PCIe slot 4 (Oracle Flash Accelerator F640)
9	PCIe slot 5 (Oracle Flash Accelerator F640)
10	PCIe slot 6 (Oracle Flash Accelerator F640)
11	Serial management (SER MGT) RJ-45 serial port
12	Oracle Integrated Lights Out Manager (ILOM) service processor (SP) network management (NET MGT) RJ-45 10/100/1000BASE-T port
13	Network (NET) 100/1000BASE-T RJ-45 Gigabit Ethernet (GbE) port: NET 0
14	USB 3.0 connector
15	System status LEDs: Locate Button/LED: white; Fault-Service Required: amber; System OK: green
16	PCIe slot 7 (Oracle Dual Port QDR InfiniBand Adapter M3) Note - This slot is unpopulated if the Oracle Dual Port 100 Gb Ethernet Adapter is installed in PCIe slot 1.
17	PCIe slot 8 (Oracle Flash Accelerator F640)
18	PCIe slot 9 (Oracle Flash Accelerator F640)
19	PCIe slot 10 (Oracle Flash Accelerator F640)
20	PCIe slot 11

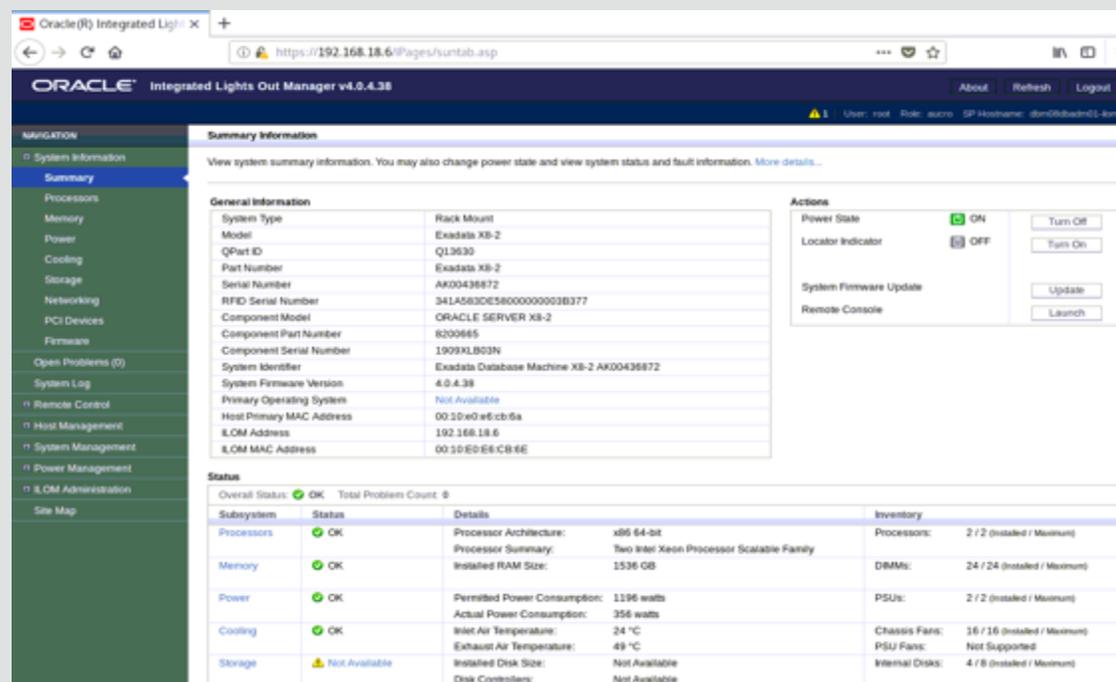
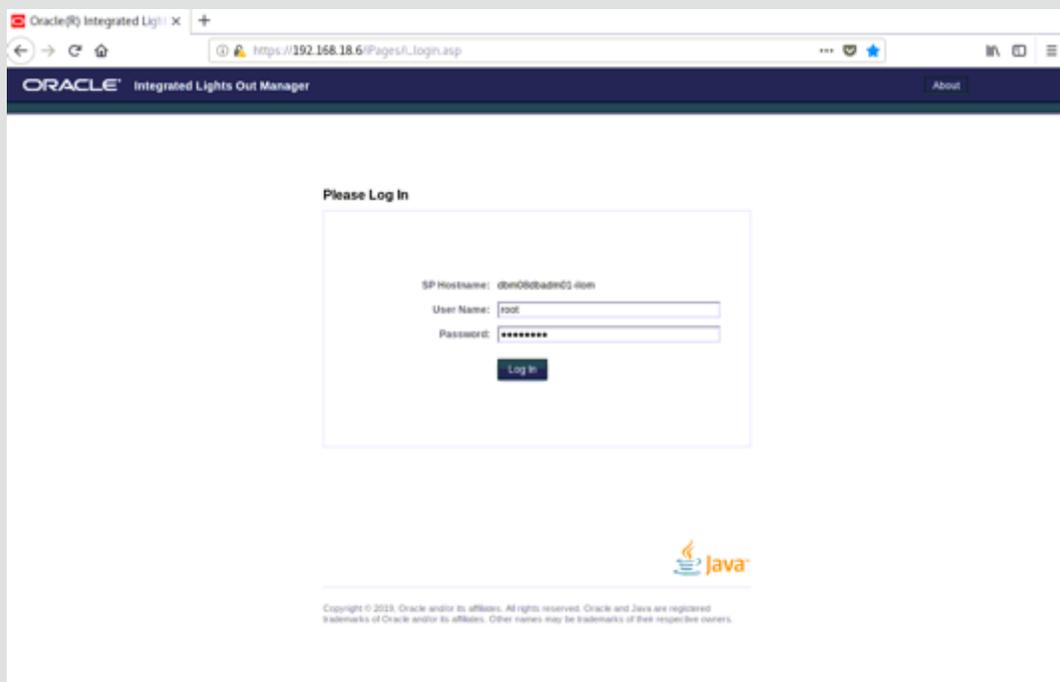
ILOM管理:

登录ILOM:

Oracle Exadata的硬件设备均配有Integrated Lights Out Manager，简称ILOM。而所有ILOM均提供两种方式登陆，既浏览器方式和命令行方式，所有硬件设备硬件均可以通过两种方式运维。

方式一：浏览器登录方式:

通过硬件的ILOM，我们可以通过浏览器对硬件实施运维，登陆方式为：<https://192.168.18.6>（ILOM的IP地址）
缺省用户名与口令为：root/welcome1，登陆界面如下：



ILOM管理：

登录ILOM：

方式二：SSH登录方式：

#ssh root@[192.168.18.6](#) (ILOM的IP地址)

缺省用户名与口令为：root/welcome1 ， 登陆成功后显示如下：

Oracle(R) Integrated Lights Out Manager

Version 4.0.4.34 r129151

Copyright (c) 2019, Oracle and/or its affiliates. All rights reserved.

Warning: HTTPS certificate is set to factory default.

Hostname: dbm0xdbadm0x-ilom

->

ILOM管理：

服务器ILOM命令行常用命令：

1、重启iLOM:

-> reset /SP

2、重启操作系统:

->reset /System

3、清除错误信息:

-> show /SP/faultmgmt

-> start /SP/faultmgmt/shell

faultmgmtsp> fmadm faulty

faultmgmtsp> fmadm repair <UUID>

faultmgmtsp> fmadm faulty

faultmgmtsp> exit

或:

-> set /SYS/MB clear_fault_action=true

4、设置下次启动设备:

-> set /HOST boot_device=cdrom|pxe



ILOM管理：

ILOM命令行常用命令：

5、登录控制台：

-> start /SP/console

输入用户名及密码登录本机的操作系统。

在操作系统中要登录本机的iLOM可以使用命令：`#ipmitool sunoem cli`

6、收集snapshot：

-> set /SP/diag/snapshot dataset=normal

-> set /SP/diag/snapshot dump_uri=sftp://root:password@IP/tmp/snapshot

-> cd /SP/diag/snapshot

-> show

待输出结果中的“result”部分显示“Snapshot Complete”时表明快照收集完成。

7、开机

-> start /SYS

8、关机

-> stop /SYS

InfiniBand交换机管理：

Infiniband交换机管理，通过ssh方式：

Infiniband交换机的管理口与ILOM口是聚合的，即是同一个口，ILOM端口的IP与也是管理口IP，登录网络管理端口：

```
# ssh root@IB交换机管理口 IP (如: #ssh root@192.168.18.12)
```

```
# spsh --切换到ILOM命令行模式
```

可以通过iLOM的命令来管理IB交换机，如：

设置时区：

```
->set /SP/clock/ timezone = CST (Asia/Shanghai)
```

```
-> show /SP/clock/ timezone
```

```
/SP/clock
```

```
Properties:
```

```
    timezone = CST (Asia/Shanghai)
```

设置NTP：

```
-> set /SP/clients/ntp/server/1 address = 192.168.2.250
```

```
(可设置两个: -> set /SP/clients/ntp/server/2 address = 192.168.2.25X)
```

InfiniBand交换机管理：

Infiniband交换机管理命令：

IB交换机版本检查： # version

IB交换机拓扑检查： # showtopology

IB交换机健康检查： # showunhealthy --正常返回“OK - No unhealthy sensors”

IB交换机环境测试： # env_test --正常返回“***returned OK、Environment test PASSED”

IB交换机状态检查： # ibcheckstate -v --正常返回“0 bad nodes found、0 ports with bad state found”

IB交换机端口链路检查： # listlinkup

查看所有IB交换机： # ibswitches

查看所有连接到IB交换机上的主机： # ibhosts

查看Subnet Manager服务信息： # sminfo

查看Subnet Manager优先级设置： # setsmpriority list

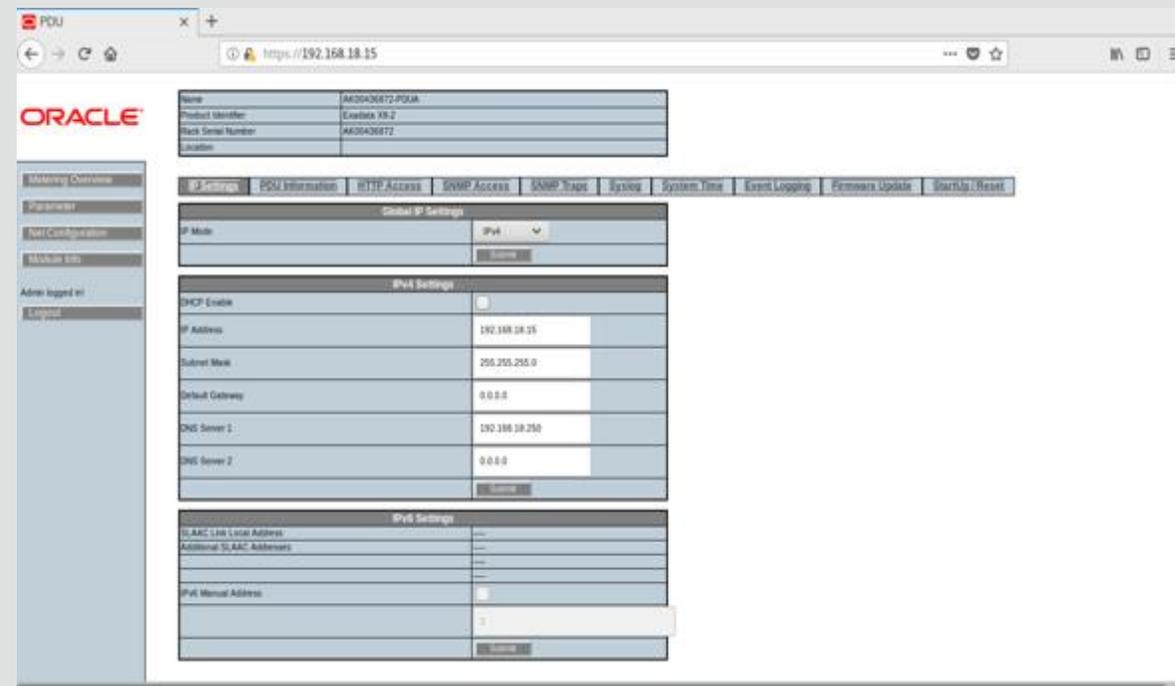
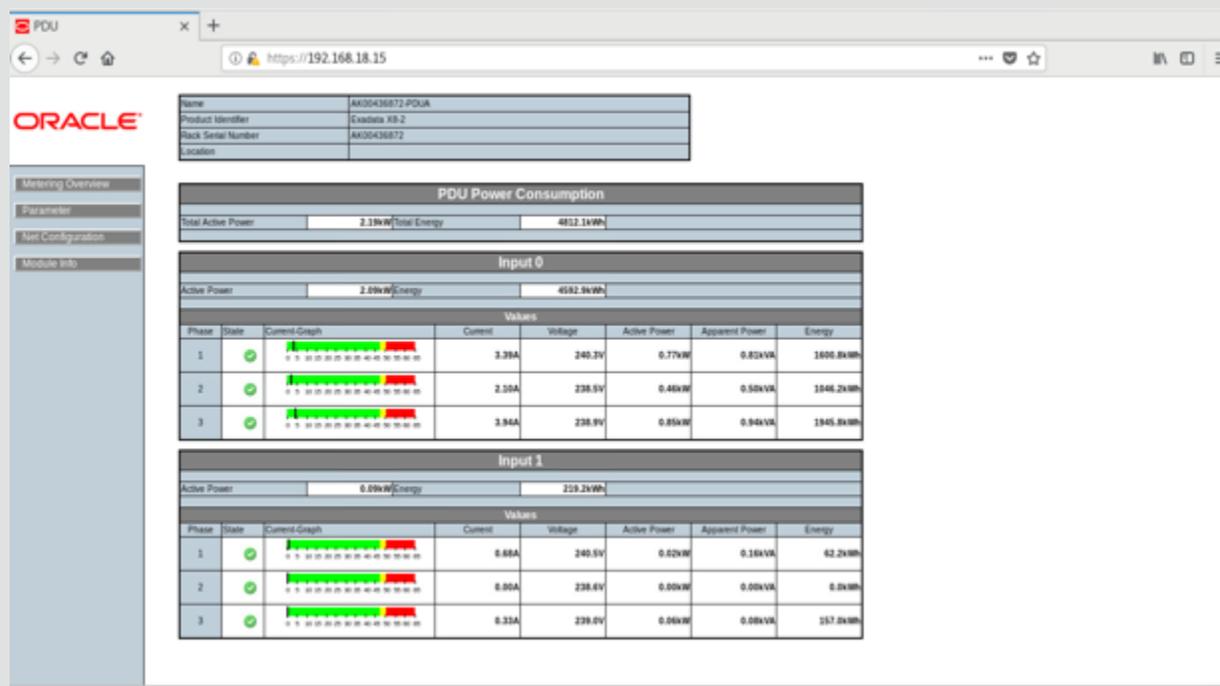
停止Subnet Manager服务： # disablesm

启动Subnet Manager服务： # enablesm

IB交换机日志查看： /var/log目录下messages、opensm.log、opensm-subnet.lst等。

PDU管理:

PDU可以通过其上的按键来查看配置和监控信息，也可以通过WEB远程登录来管理：通过web方式，直接通过在浏览器中输入PDU的IP地址，输入用户名/密码即可：登录后可以设置PDU名称、IP、DNS、NTP、SNMP等，还可以监控PDU的状态：



说明：用户名/密码是：admin/welcome1。



Cisco交换机管理:

Cisco交换机主要是连接管理网络，用来连接内部各个设备的管理口，并连接到外部的管理网络中，主要通过ssh方式登录并进行维护管理操作：

```
[root@x52-103218 ~]# ssh admin@192.168.18.11
User Access Verification
Password:

Cisco Nexus Operating System (NX-OS) Software
TAC support: http://www.cisco.com/tac
Copyright (C) 2002-2018, Cisco and/or its affiliates.
All rights reserved.
The copyrights to certain works contained in this software are
owned by other third parties and used and distributed under their own
licenses, such as open source. This software is provided "as is," and unless
otherwise stated, there is no warranty, express or implied, including but not
limited to warranties of merchantability and fitness for a particular purpose.
Certain components of this software are licensed under
the GNU General Public License (GPL) version 2.0 or
GNU General Public License (GPL) version 3.0 or the GNU
Lesser General Public License (LGPL) Version 2.1 or
Lesser General Public License (LGPL) Version 2.0.
A copy of each such license is available at
http://www.opensource.org/licenses/gpl-2.0.php and
http://opensource.org/licenses/gpl-3.0.html and
http://www.opensource.org/licenses/lgpl-2.1.php and
http://www.gnu.org/licenses/old-licenses/library.txt.
dm08sw-adm01#
```

说明： 用户名/密码是： admin/welcome1。
检查交换机状态(电源、风扇等): # show environment
查看当前运行的配置信息： # show running-config
查看启动配置信息： # show startup-config

```
dm08sw-adm01# show environment

Fan:
-----
Fan          Model          Hw    Direction    Status
-----
Fan1(sys_fan1)  NXA-FAN-30CFM-F  --    back-to-front  Ok
Fan2(sys_fan2)  NXA-FAN-30CFM-F  --    back-to-front  Ok
Fan3(sys_fan3)  NXA-FAN-30CFM-F  --    back-to-front  Ok
Fan_in_PS1     --              --    back-to-front  Ok
Fan_in_PS2     --              --    back-to-front  Ok
Fan Zone Speed: Zone 1: 0x80
Fan Air Filter : NotSupported

Power Supply:
Voltage: 12 Volts
Power
Supply      Model          Actual      Actual      Total      Status
Output     Input          Capacity
(Watts )   (Watts )      (Watts )
-----
1          NXA-PAC-350W-PE  68 W        81 W        350 W      Ok
2          NXA-PAC-350W-PE  61 W        72 W        350 W      Ok

Power Usage Summary:
-----
Power Supply redundancy mode (configured)    PS-Redundant
Power Supply redundancy mode (operational)   PS-Redundant

Total Power Capacity (based on configured mode)    350.00 W
Total Grid-A (first half of PS slots) Power Capacity 350.00 W
Total Grid-B (second half of PS slots) Power Capacity 350.00 W
Total Power of all Inputs (cumulative)             700.00 W
Total Power Output (actual draw)                   129.00 W
Total Power Input (actual draw)                    154.00 W
Total Power Allocated (budget)                     N/A
Total Power Available for additional modules       N/A

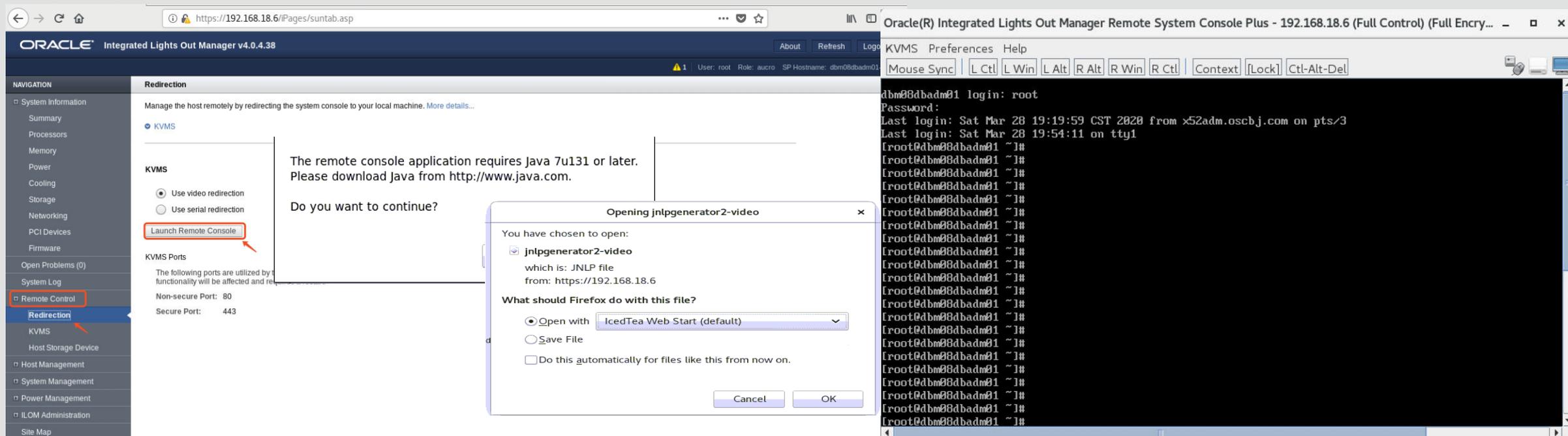
Temperature:
-----
Module  Sensor      MajorThresh  MinorThres  CurTemp  Status
(Celsius) (Celsius)   (Celsius)
-----
1       FRONT      80           70          39       Ok
1       BACK       70           42          30       Ok
1       CPU        90           80          39       Ok
1       Homewood  110          90          47       Ok
dm08sw-adm01#
```

Exadata远程终端：

远程终端：

Exadata提供两种远程终端，基于浏览器方式和基于命令行方式，通过远程终端，我们可以监控主机启动与关闭过程，并在主机网卡异常时，登陆主机。

通过浏览器登陆ILOM，点击屏幕左侧的remote control。在点击Redirection，在出现Redirection界面后，点击Launch Remote Console按钮。点击后系统提示需要的JAVA版本，请下载相应的版本，点击OK或YES，最终会弹出虚拟KVMS：



Exadata远程终端：

命令行远程终端：

通过ssh登陆ILOM，输入：

#start /SP/console,回车并输入“y”，进入字符kvms，退出按ESC和（。

```
[root@x52-103218 ~]# ssh root@192.168.18.6
Password:

Oracle(R) Integrated Lights Out Manager

Version 4.0.4.38 r130206

Copyright (c) 2019, Oracle and/or its affiliates. All rights reserved.

Warning: HTTPS certificate is set to factory default.

Hostname: dbm08dbadm01-ilom

-> start /SP/console
Are you sure you want to start /SP/console (y/n)? y

Serial console started. To stop, type ESC (

dbm08dbadm01 login:
dbm08dbadm01 login: █
```

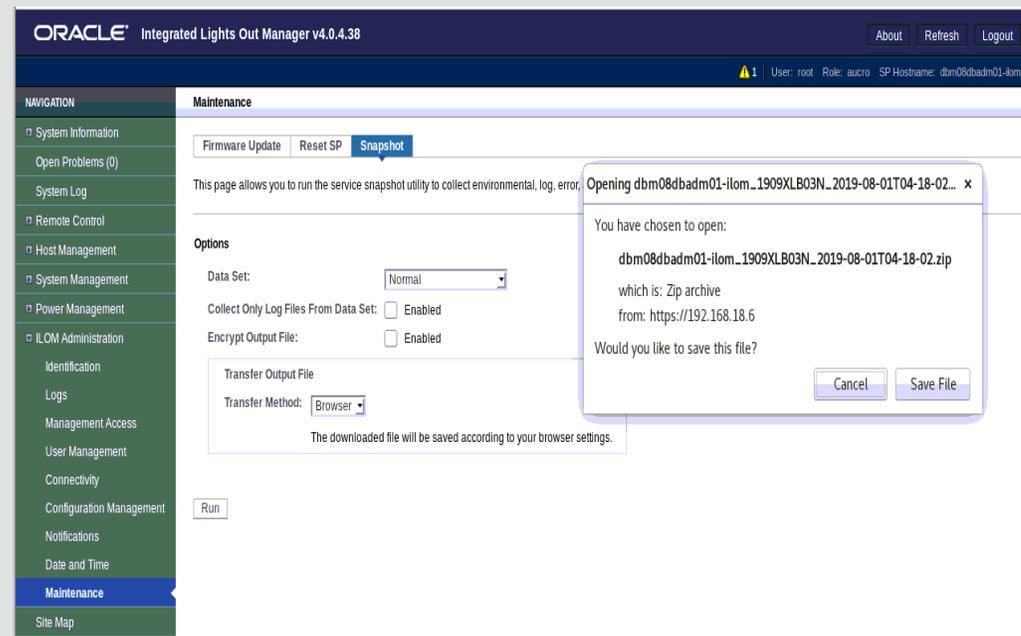
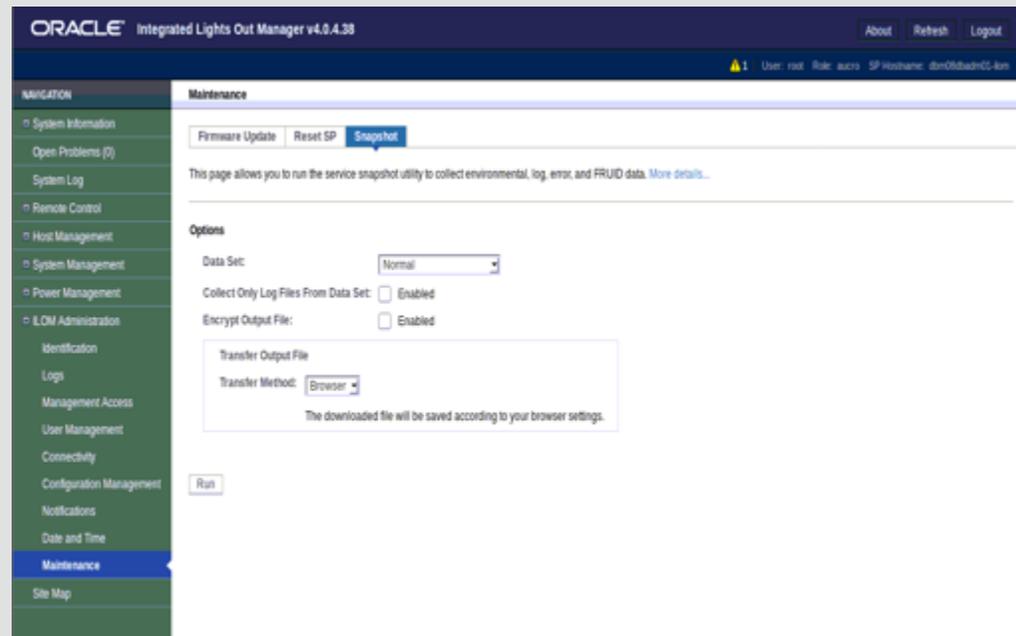
Exadata硬件信息收集:

在系统出现报警后，Oracle Support一般会要求收集系统信息进行远程诊断，通常会收集两种信息：ILOM snapshot和sundiag信息。

ILOM snapshot:

通过浏览器收集ILOM snapshot信息（也可以使用命令行来收集），通过浏览器登陆ILOM后，按如下顺序找到snapshot生成界面：

ILOM Administrator → Maintenance → snapshot → RUN



在点击Run按钮后，系统会自动收集相关信息，并生成下载文件，保存即可。



Exadata硬件信息收集:

Sundiag: 使用虚拟kvm或主机管理网络登陆主机，需要在系统启动状态下执行，直接运行 /opt/oracle.SupportTools/sundiag.sh命令，收集相关硬件信息。结果输出在/var/log/exadatatmp/目录下，文件名称为sundiag*.tar.bz2。

```
dbm08dbadm01 login: root
Password:
Last login: wed Jul 31 21:08:59 CST 2019 from x52adm.oscbj.com on pts/5
Last login: Thu Aug  1 13:25:02 on ttys0
[root@dbm08dbadm01 ~]# /opt/oracle.SupportTools/sundiag.sh

Oracle Exadata Database Machine - Diagnostics Collection Tool

Gathering Linux information

SecureBoot enabled
Skipping collection of OSwatcher/Exawatcher logs, Cell Metrics and Traces
Skipping ILOM collection. Use the ilom or snapshot options, or login to ILOM
over the network and run Snapshot separately if necessary.

/var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25
Gathering dbms information
cp: will not overwrite just-created '/var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25/asr/index.html' with '/var/lo
cp: will not overwrite just-created '/var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25/asr/dbm08dbadm01.oscbj.com_in
_12_15_21_49_02h00m00s_0/dbm08dbadm01.oscbj.com_inc.html'
cp: will not overwrite just-created '/var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25/asr/dbm08dbadm01.oscbj.com.ht
15_21_49_02h00m00s_0/dbm08dbadm01.oscbj.com.html'
cp: will not overwrite just-created '/var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25/asr/index.html' with '/var/lo
s_0/index.html'
cp: will not overwrite just-created '/var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25/asr/dbm08dbadm01.oscbj.com_me
7_12_15_21_49_02h00m00s_0/dbm08dbadm01.oscbj.com_menu.html'
cp: will not overwrite just-created '/var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25/asr/exawchart.trc' with '/var
cp: will not overwrite just-created '/var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25/asr/exawchart.log' with '/var
cp: will not overwrite just-created '/var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25/asr/alertHistory' with '/var
Generating diagnostics tarball and removing temp directory

=====
Done. The report files are bzip2 compressed in /var/log/exadatatmp/sundiag_dbm08dbadm01_1909XLB03N_2019_08_01_13_25.tar.bz2
=====
```

Exadata 开关机:

节点主机开关机:

由于Exadata数据库一体机上运行的是Oracle GI集群系统和Oracle RAC数据库，所以有严格的开关机顺序:

系统开机:

1、为机柜加电

打开PDU开关进行加电，加电后:

- 1) db服务器、cell服务器的SP(service processor)自动加电：指示灯都变绿，慢闪;
- 2) cisco交换机、Infiniband交换机自动加电开机;

2、依次启动cell服务器（存储服务器）

按住其面板开关5秒。

3、待所有cell服务器完全启动后再依次启动db服务器（数据库服务器）

按住其面板开关5秒。

4、启动其他数据库、应用等。

```
#crsctl start cluster
```

```
#srvctl start database -d xxxx
```

上面2、3步也可以也可在ILOM进行。

注意：启动顺序先cell server后db server。

Exadata开关机:

系统关机:

1、关闭db服务器:

1) 登录到要关闭db服务器

2) 关闭Oracle Clusterware

```
# GRID_HOME/grid/bin/crsctlstop [-f] cluster
```

3) 关闭节点

```
# shutdown -h now
```

2、关闭cell服务器:

1) 登录到要关闭cell服务器

2) 执行命令# shutdown -h now

关闭cell节点时，所有存储相关服务会自动停掉；

3、关闭PDU开关；

此时可以拔掉电源，进行硬件维护等操作。

注意:

1) 如果需要关闭一个以上的cell节点，则需要停止所有数据库和clusterware；

2) 只关闭一个cell节点不会影响到正在运行的数据库进程或ASM；

3) 关闭顺序先db server后cell server。

◆ Exadata 运维管理

- 1 Exadata 组件管理
- 2 **Exadata 状态检查**
- 3 Exadata 管理工具
- 4 Exadata 案例分享

Exadata 系统状态检查：

检查系统报错信息：

通过iLOM来查看系统状态，以Web方式登录如下：

Oracle(R) Integrated Light x +

← → ↻ 🏠 <https://192.168.18.6/iPages/suntab.asp> ... 📄 ☆

ORACLE Integrated Lights Out Manager v4.0.4.38 [About](#) [Refresh](#) [Logout](#)

⚠️ 1 | User: root | Role: auro | SP Hostname: dbm08dbadm01-ilo

NAVIGATION

- System Information
 - Summary
 - Processors
 - Memory
 - Power
 - Cooling
 - Storage
 - Networking
 - PCI Devices
 - Firmware
- Open Problems (0)**
- System Log
- Remote Control
- Host Management
- System Management
- Power Management
- ILOM Administration
- Site Map

Open Problems

There are no open problems to report.

Open Problems

ID	Date/Time	Subsystem	Component
No Items To Display			

说明：如果系统有报警信息，在屏幕左侧的Open problems(x)中，有非零数字，代表报警数量，点击可进入报错信息浏览界面，通常为0.

Exadata 系统状态检查：

在iLOM中进行硬件状态检查：

点击屏幕左侧的system information，可以监控系统硬件信息，包括汇总信息和各个硬件组件的状态信息，当Open Problems(1)提示CPU0故障，点击Processors查看详细信息：

The screenshot shows the iLOM web interface for the Processors section. The left sidebar contains navigation options like System Information, Processors, Memory, Power, Cooling, Storage, Networking, PCI Devices, Firmware, Open Problems (1), System Log, Remote Control, Host Management, System Management, Power Management, ILOM Administration, and Site Map. The main content area shows the following details:

- Health:** ✖ Service Required
- Health Details:** P0 (CPU 0) is faulty. See the [Open Problems](#) page for more information.
- Architecture:** x86 64-bit
- Summary Description:** Two Intel Xeon Processor E5 V3 Series
- Installed CPUs:** 2
- Maximum CPUs:** 2

CPUs

CPU #	Health	Health Details	Location	Maximum Clock Speed	Total Cores	CPU Details
CPU 0	✖ Service Required	A malfunctioning component has been detected by the Complex Programmable Logic Device (CPLD). See the Open Problems page for more information.	P0 (CPU 0)	2.400 GHz	8	Details
CPU 1	✔ OK	-	P1 (CPU 1)	2.400 GHz	8	Details

说明： 在db server上显示的storage信息，为本地磁盘信息，该磁盘是本地文件系统，不保存数据库数据。在cell server上显示的storage信息，为Exadata的数据库数据存储磁盘信息。



Exadata 系统状态检查:

通过命令行方式登录iLOM检查是否有故障告警:

✓ 可以直接ssh登录iLOM的地址:

```
#ssh root@host_ilom_addr
```

✓ 如果本机操作系统可以登录, 也可以从本机的操作系统直接登录iLOM:

```
#ipmitool sunoem cli
```

```
Connected. Use ^D to exit.
```

```
-> show /SP/faultmgmt
```

```
-> start /SP/faultmgmt/shell
```

```
faultmgmtsp> fmadm faulty
```

或:

```
# ipmitool sunoem cli 'show faulty'
```

Exadata 系统状态检查-DB节点:

DB节点当前镜像检查: # imageinfo

DB节点历史镜像检查: # imagehistory

DB节点操作系统检查: # cat /etc/os-release 或 #cat /etc/redhat-release

DB节点硬盘状态检查: # dbmcli -e list physicaldisk --正常返回“normal”

DB节点硬盘RAID信息检查: # /opt/oracle.SupportTools/reclaimdisks.sh -check

DB节点硬件firmware检查: # /opt/oracle.SupportTools/CheckHWnFWProfile -c strict --正常返回“[SUCCESS]”

DB节点IB网卡状态检查: # ibstatus --正常返回“state: ACTIVE phys state: LinkUp”

DB节点IB网络状态检查: # /opt/oracle.SupportTools/ibdiagtools/verify-topology --正常返回“[SUCCESS]”

DB节点业务网卡状态检查: # ethtool ethX --ethX为业务网卡名称, Speed: 显示网络速率, Link detected: 正常状态是yes

DB节点文件系统使用率检查: # df -Th --建议Use%在90%以下

DB节点大页设置及使用检查: # cat /proc/meminfo|grep -i HugePages

DB节点当前时区检查: # timedatectl

DB节点交换分区检查: # free --正常Swap: used应该为0, 否则需要检查内存的使用状况

DB节点DNS解析顺序检查: # cat /etc/nsswitch.conf -- hosts: files dns默认先/etc/hosts后resolv.conf

DB节点DNS配置信息检查: # cat /etc/resolv.conf --配置正确的nameserver

DB节点系统日志检查: /var/log目录下messages、lastlog、maillog等。

Exadata 系统状态检查-CELL节点:

CELL节点当前镜像检查: # imageinfo

CELL节点历史镜像检查: # imagehistory

CELL节点操作系统检查: # cat /etc/os-release 或 #cat /etc/redhat-release

CELL节点硬盘, 闪存卡状态检查: # cellcli -e list physicaldisk --正常返回“normal”

CELL节点闪存工作模式状态: # cellcli -e list cell detail|grep flashCacheMode --正常返回“WriteBack”

CELL节点各数据库占用闪存情况检查: # cellcli -e list metriccurrent DB_FC_BY_ALLOCATED

CELL节点griddisk状态检查: # cellcli -e list griddisk attributes name, asmmodestatus, status

CELL节点磁盘组缓存策略检查: # cellcli -e list griddisk attributes name,asmDiskgroupName,cachingPolicy

CELL节点系统盘RAID信息检查: # mdadm -Q -D /dev/mdX --mdX为系统盘, 使用df -h命令查看根文件系统所用的盘符即是。

CELL节点磁盘擦洗操作周期检查: #cellcli -e list cell attributes harddiskscrubinterval

CELL节点硬件firmware检查: # /opt/oracle.SupportTools/CheckHWnFWProfile -c strict --正常返回“[SUCCESS]”

CELL节点IB网卡状态检查: # ibstatus --正常返回“state: ACTIVE phys state: LinkUp”

CELL节点IB网络状态检查: # /opt/oracle.SupportTools/ibdiagtools/verify-topology --正常返回“[SUCCESS]”

CELL节点防火墙状态检查: # systemctl status cellwall --正常返回“Active: active”

CELL节点IORM状态检查: # cellcli -e list iormplan detail

CELL节点性能统计数据检查: #cellsrvstat -interval=2 -count=10 --间隔2秒执行一次, 共执行10次

CELL节点磁盘IO检查: #iostat -dxk 1 100 --每秒运行1次, 执行100次

CELL节点文件系统使用率检查: # df -Th --建议Use%在90%以下

CELL节点系统日志检查: /var/log目录下messages、lastlog、maillog等。

CELL节点alert日志检查: /var/log/oracle/diag/asm/cell/<cell_nodename>/trace/alert.log

Exadata 数据库检查(1):

数据库集群alert日志:

```
/u01/app/grid/diag/crs/<nodename>/crs/trace/alert.log
```

ASM alert日志:

```
/u01/app/grid/diag/asm/+asm/+ASM*/trace/alert_+ASM*.log
```

DB alert日志:

```
/u01/app/oracle/diag/rdbms/<dbname>/<SID>/trace/alert_<SID>.log
```

CELL alert日志:

```
/var/log/oracle/diag/asm/cell/<cell_nodename>/trace/alert.log
```

集群OCR状态检查:

```
$ $ORACLE_HOME/bin/ocrcheck --使用grid用户执行
```

集群OCR备份检查:

```
$ $ORACLE_HOME/bin/ocrconfig -showbackup --使用grid用户执行, 默认会自动进行备份: 每4小时 (CRS会保留最后3个副本); 每天结束时 (保留最后2个副本); 每周结束时 (CRS保留最后2个副本)。
```

集群vote disk检查:

```
$ $ORACLE_HOME/bin/crsctl query css votedisk --使用grid用户执行
```

检查集群网络配置信息:

```
$ $ORACLE_HOME/bin/oifcfg getif --显示正确的公网和私网配置信息
```

数据库集群状态检查:

```
# $ORACLE_HOME/bin/crsctl stat res -t --使用root用户执行
```

数据库监听检查:

```
$ $ORACLE_HOME/bin/lsnrctl status [LISTENER|LISTENER_SCAN1|LISTENER_SCAN2|LISTENER_SCAN3] --可以通过$ps -ef|grep tnslsnr命令查看监听名称:
```

数据库状态检查:

```
SQL> select name,dbid,open_mode,log_mode,flashback_on,supplemental_log_data_min,force_logging from v$database;
```

数据库时区及会话时区检查:

```
SQL>select dbtimezone,sessiontimezone from dual; --使用oracle用户执行
```



Exadata 数据库检查(2):

数据库磁盘组使用情况检查:

```
SQL> select name, total_mb, free_mb, total_mb - free_mb used_mb, round(100*(total_mb - free_mb)/total_mb,2) "pct_used(%)" ,offline_disks, state, type from v$asm_diskgroup;
```

或使用asmcmd命令: \$ \$ORACLE_HOME/bin/asmcmd lsdg --使用grid用户执行

数据库磁盘组属性检查:

例如查看disk_repair_time、failgroup_repair_time的信息:

```
SQL>select a.name diskgroup_name,b.name attribute_name,b.value from v$asm_diskgroup a,v$asm_attribute b where a.group_number=b.group_number and b.name like '%repair%';
```

或使用asmcmd命令: \$ \$ORACLE_HOME/bin/asmcmd lsattr -G DATA1 -l

磁盘组failgroup检查:

```
SQL> select a.group_number group_number,b.name dgname,a.failgroup failgroup,a.disk_number,a.name diskname,a.mode_status status,a.voting_file voting_file from v$asm_disk a,v$asm_diskgroup b where a.group_number=b.group_number order by 1,2,3,4;
```

磁盘Rebalance操作检查:

```
SQL>select dg.name, o.* from gv$asm_operation o, v$asm_diskgroup dg where o.group_number = dg.group_number;
```

或使用asmcmd命令: \$ \$ORACLE_HOME/bin/asmcmd lsop

数据库允许智能扫描的内置函数或操作符检查:

```
SQL>select name,disp_type,descr from v$sqlfn_metadata where offloadable='YES';
```

数据库直接路径读取参数检查:

```
SQL> select a.ksppinm "parameter", b.ksppstvl "session value", c.ksppstvl "instance value" from x$ksppi a, x$ksppcv b, x$ksppsv c where a.indx = b.indx and a.indx = c.indx and a.ksppinm = '_serial_direct_read';
```

数据库备份检查:

```
SQL>select session_key,start_time,end_time,output_device_type,status,input_type,input_bytes_display, output_bytes_display,time_taken_display from v$rman_backup_job_details;
```

AWR配置检查:

```
SQL> select * from dba_hist_wr_control; SQL> show parameter statistics_level
```



Exadata 数据库检查 (3) :

收集数据库的性能报告:

```
SQL> @?/rdbms/admin/awrrpt.sql
SQL> @?/rdbms/admin/addmrpt.sql
SQL> @?/rdbms/admin/ashrpt.sql
SQL> @?/rdbms/admin/awrsqrpt.sql
SQL> @?/rdbms/admin/awrddrpt.sql
SQL> @?/rdbms/admin/awrgrpt.sql
SQL> @?/rdbms/admin/awrrpti.sql
SQL> @?/rdbms/admin/awrextr.sql
```

...

Exadata Alerts Summary

- Number of open alerts at the end snapshot
- Stateless alerts are restricted to those opened in the past 24 hours

Cell Name	Total	Critical	Warning	Info
Total	3	1	2	0
dbmx6hcceladm01	1	0	1	0
dbmx6hcceladm02	1	0	1	0
dbmx6hcceladm03	1	1	0	0

[Back to Exadata Server Health Report](#)

Exadata Alerts Detail

- Number of open alerts at the end snapshot
- only the 10 most recent open alerts per cell are displayed
- Stateless alerts are restricted to those opened in the past 24 hours
- ordered by Cell Name, Begin Time desc

Cell Name	Alert Time	Severity	Stateful	Message
dbmx6hcceladm01	04/02/2020 03:07:26	warning	Y	*Advanced Intrusion Detection Environment (AIDE) detected potential changes to software on this system. The changes are listed in /var/log/aide/aide.log and also at the end of this alert message. Summary : : Total number of files : 46542 Added files : 6 Removed files : 0 Changed files : 1 *
dbmx6hcceladm02	04/02/2020 03:07:22	warning	Y	*Advanced Intrusion Detection Environment (AIDE) detected potential changes to software on this system. The changes are listed in /var/log/aide/aide.log and also at the end of this alert message. Summary : : Total number of files : 46542 Added files : 6 Removed files : 0 Changed files : 1 *
dbmx6hcceladm03	03/31/2020 12:45:11	critical	Y	*Flash check has detected the following issue(s): Attribute Name : FlashCardCount Required : 4/2 Found : 3*

[Back to Exadata Server Health Report](#)

Main Report

- [Report Summary](#)
- [Wait Events Statistics](#)
- [SQL Statistics](#)
- [Instance Activity Statistics](#)
- [IO Stats](#)
- [Buffer Pool Statistics](#)
- [Advisory Statistics](#)
- [Wait Statistics](#)
- [Undo Statistics](#)
- [Latch Statistics](#)
- [Segment Statistics](#)
- [Dictionary Cache Statistics](#)
- [Library Cache Statistics](#)
- [Memory Statistics](#)
- [Replication Statistics \(GoldenGate, XStream\)](#)
- [Advanced Queuing](#)
- [Resource Limit Statistics](#)
- [Shared Server Statistics](#)
- [Initialization Parameters](#)
- [Active Session History \(ASH\) Report](#)
- [ADDM Reports](#)

RAC Statistics

- [RAC Report Summary](#)
- [Global Messaging Statistics](#)
- [Global CR Server Statistics](#)
- [Global Current Server Statistics](#)
- [Global Cache Transfer Statistics](#)
- [Interconnect Statistics](#)
- [Dynamic Remastering Statistics](#)

Exadata Configuration and Statistics

- [Exadata Report Summary](#)
- [Exadata Server Configuration](#)
- [Exadata Server Health Report](#)
- [Exadata Statistics](#)

[Back to Exadata Server Health Report](#)

Exadata Statistics

- [Performance Summary](#)
- [Exadata Resource Statistics](#)
- [Exadata Smart Statistics](#)
 - [Smart IO](#)
 - [Write Offload](#)
 - [Flash Log](#)
 - [Flash Cache](#)
 - [Memory Cache](#)
 - [PMEM Cache](#)
- [Exadata IO Reasons](#)
 - [Top IO Reasons by Requests](#)
 - [Top IO Reasons by MB](#)
- [Exadata Top Database Consumers](#)
 - [Top Databases by Requests](#)
 - [Top Databases by Requests - Details](#)
 - [Top Databases by Throughput](#)
 - [Top Databases by Requests per Cell](#)
 - [Top Databases by Requests per Cell - Details](#)
 - [Top Databases by Throughput per Cell](#)
- [Exadata IO Latency Capping](#)
 - [Cancelled IOs - Client](#)
 - [Cancelled IOs - Cells](#)
- [Exadata Flash Wear](#)

说明: 针对Exadata专门有一部分“Exadata Configuration and Statistics”可以用来分析Exadata的性能状况。

◆ Exadata 运维管理

- 1 Exadata 组件管理
- 2 Exadata 状态检查
- 3 Exadata 管理工具
- 4 Exadata 案例分享

Exadata管理工具:

- OEM监控Exadata
- dcli
- dbmcli
- cellcli
- exachk
- exawatcher
- exacli
- exadcli

Exadata管理工具:

OEM监控Exadata:

OEM目前是Exadata最好的监控工具，它能够做到集中监控和管理，能够持续的监控大量的配置，发现变更，能够找出配置错误，监控显示如下:

The screenshot displays the Oracle Enterprise Manager Cloud Control 13c interface. On the left, the 'Databases' section is active, showing an 'Overview of Incidents and Problems' and a 'Database Load Map' with a green bar chart. A red box highlights the 'Exadata' option in the 'Targets' menu. On the right, a '数据库计算机示意图' (Database Machine Diagram) is shown, illustrating the physical architecture of an Exadata system. The diagram includes components like 'EXADATA', 'sn01sw-pdu0', 'sn01sw-ib0', 'sn01dbadm02-4cm', 'sn01dbadm01-4cm', and 'sn01celadm03'. A legend on the right side of the diagram provides details on viewing options (e.g., '查看' - View, '缩放' - Zoom), status indicators (e.g., '状态' - Status), and component types (e.g., '组件' - Components).



Exadata管理工具:

dcli工具:

dcli工具可以分布式地同时指定的所有节点上运行相同的命令，并展现命令的输出，它是一个python脚本工具，它可以下载到任何有python环境的系统上调用，常用命令如下:

命令格式: `dcli [options] [command]`

首先，在某一节点上生成密钥:

```
#ssh-keygen -t dsa
```

建立此节点到其它节点的信任关系:

```
#dcli -g /tmp/node_group -l root -k
```

(其中-g参数指定的node_group中存放着所有节点的节点名称或IP，每行一个，也可以把计算节点与存储节点分开到不同的文件中如cell_group、db_group; 参数-l指定目标节点的登录用户。)

在此节点上运行如下命令，则会返回所有节点的主机名:

```
# dcli -g /tmp/node_group -l root date
```

如果同时运行多条命令，需要使用分号分隔，同时用单引号或双引号引起来:

```
# dcli -g /tmp/node_group -l root 'hostname;date'
```

Exadata管理工具:

dcli工具:

显示各个计算节点的状态:

```
#dcli -g db_group -l root dbmcli -e list dbserver
```

显示各个存储节点的状态:

```
#dcli -g cell_group -l root cellcli -e list cell
```

使用-f把指定文件拷贝到所有节点的-d指定的目录下:

```
#dcli -g db_group -l root -f /tmp/test.sh -d /tmp
```

使用-x 执行指定的可执行文件, 并将此文件拷贝到所有节点的-d指定的目录下:

```
#dcli -g db_group -l root -x /tmp/run.sh -d /tmp
```

使用-c直接指定节点名称或IP地址, 多个节点之前使用逗号分隔:

```
#dcli -c dbm09dbadm01,dbm09dbadm02,dbm09celadm02 -l root hostname
```

关闭数据库集群:

```
#dcli -g db_group -l root /u01/app/19.0.0.0/grid/bin/crsctl stop crs
```

关闭CELL服务:

```
#dcli -g cell_group -l root "su - celladmin -c \"cellcli -e alter cell shutdown services all\""
```

Exadata管理工具:

dbmcli工具:

dbmcli工具是用于配置计算节点以及计算节点中管理对象的命令行管理工具，用来管理和监控计算节点的运行情况，dbmcli工具在计算节点中从Linux命令行直接调用，在root、dbmmonitor、dbmadmin或dbmsvc操作系统用户下，直接输入dbmcli命令即可，常用命令如下:

命令格式: dbmcli [port] [-n] [-e <command>]

DBMCLI>help 或 #dbmcli -e help

DBMCLI>list dbserver detail

DBMCLI>list physicaldisk detail

DBMCLI>list lun detail

DBMCLI>list metricdefinition attributes name,description

DBMCLI>list metriccurrent

DBMCLI>list metrichistory [where ..]

DBMCLI>**list alerthistory**

Exadata管理工具:

cellcli工具:

cellcli工具是存储节点中的一个命令行工具，它是用户与Exadata存储软件交互的一个接口，cellcli工具在存储节点中从Linux命令行直接调用，在root、cellmonitor或celladmin操作系统用户下，直接输入cellcli命令即可，常用命令如下:

命令格式: cellcli [port] [-n] [-e <command>]

CELLCLI>help 或 #cellcli -e help

CELLCLI>list cell detail

CELLCLI>list physicaldisk [where diskType=HardDisk AND status like ".*failure.*"] detail

CELLCLI>list lun detail

CELLCLI>list celldisk detail

CELLCLI>list griddisk detail

CELLCLI>list flashlog detail

CELLCLI>list flashcache detail

CELLCLI>list metricdefinition attributes name,description

CELLCLI>list alerthistory

Exadata 管理工具:

cellcli 工具(续):

重点介绍一下使用 **cellcli** 命令查看性能监控指标:

查看存储节点所有监控对象的当前指标值:

```
CellCLI>list metriccurrent
```

查看某个监控对象的当前指标值, 如瞬时温度:

```
CellCLI>list metriccurrent cl_temp detail
```

查看 **metriccurrent** 的所有属性:

```
CellCLI>describe metriccurrent
```

查看某一类监控对象的当前指标值, 如 **celldisk** 的所有监控对象当前指标:

```
CellCLI>list metriccurrent attributes name,metricobjectname,metricvalue,collectiontime where  
objecttype='CELLDISK' --还可以指定 cell|flashcache|flashlog|smartio|iorm_database 等
```

查看存储节点所有监控对象的历史指标值:

```
CellCLI>list metrichistory
```

查看某个监控对象的历史指标值, 如温度:

```
CellCLI>list metrichistory cl_temp detail
```

查看某一类监控对象的历史指标值, 如 **celldisk** 的所有监控对象历史指标:

```
CellCLI>list metrichistory attributes name,metricobjectname,metricvalue,collectiontime where  
objecttype='CELLDISK'
```

查看某个具体监控对象的详细含义, 如查看温度 **cl_temp** 这个监控项的详细信息, 使用如下命令:

```
CellCLI>list metricdefinition cl_temp detail
```

Exadata管理工具:

exachk工具:

exachk工具审查**exadata**上各组件的重要配置信息和性能配置是否满足要求，在MOS上下载最新的版本（文档号：**1070954.1**）当前是**Oracle Autonomous Health Framework (AHF) - with ORAchk, EXAchk & Trace File Analyzer**。定期的运行**exachk**工具，收集机器上的系统信息，并结合最佳实践和建议值，及时的发现潜在问题，然后把这些隐患消除，保障**Exadata**系统的稳定运行。

exachk主要检查:

- (1) 硬件和**Firmware**版本
- (2) 操作系统的内核参数
- (3) 操作系统补丁包
- (4) 针对**RAC**数据库的特定操作系统配置
- (5) **GI**集群
- (6) 数据库
- (7) **ASM**
- (8) 最大高可用性架构

使用方法：用**root**用户执行：**#!/opt/oracle.SupportTools/exachk/exachk**

Exadata 管理工具:

exachk 工具:

exachk 执行过程中按提示进行相应操作，命令执行完之后，会产生一个后缀为 .zip 的压缩文件，放在 /opt/oracle.SupportTools/exachk 目录下，下载到本地并解压，使用浏览器打开后缀为 .html 的文件，可以看到系统健康情况的得分，可以根据给出的建议信息来调整当前系统的配置，以达到最佳实践的目标。

Oracle Exadata Assessment Report

System Health Score is 90 out of 100 (detail)

Cluster Summary

Cluster Name	Cluster-c1
OS/Kernel Version	LINUX X86-64 OELRHHEL 7 4.1.12-124.20.3.el7uek.x86_64
CRS Home - Version	/u01/app/18.0.0.0/grid - 18.0.0.0.0
DB Home - Version - Names	/u01/app/oracle/product/18.0.0.0/dbhome_1 - 18.0.0.0.0 - graphdb database
EM Agent Home	/u01/app/oem/agent_13.3.0.0.0
Exadata Version	19.1.0.0.0
Number of nodes	7
Database Servers	2
Storage Servers	3
IB Switches	2
EXAchK Version	19.2.0_20190403
Collection	exachk_dbm02dbadm01_graphdb_071519_145131
Duration	16 mins, 20 seconds
Executed by	root
Arguments	
Collection Date	15-Jul-2019 14:58:14

Database Server

Status	Type	Message	Status On	Details
CRITICAL	OS Check	Database control files are not configured as recommended	All Database Servers	View
CRITICAL	OS Check	Database parameter db_create_online_log_dest_n is not set to recommended value	All Database Servers	View
CRITICAL	OS Check	One or more Ethernet network cables are not connected.	All Database Servers	View
CRITICAL	OS Check	Database parameter CLUSTER_INTERCONNECTS is not set to the recommended value	All Database Servers	View
CRITICAL	SQL Parameter Check	Database parameter USE_LARGE_PAGES is not set to recommended value	All Instances	View
CRITICAL	ASM Check	There is not enough diskgroup freespace for a rebalance operation	All ASM Instances	View
CRITICAL	ASM Check	There should be no failed diskgroup rebalance operations	All ASM Instances	View

Storage Server

Status	Type	Message	Status On	Details
CRITICAL	Storage Server Check	Storage Server Flash Memory is not configured as Exadata Smart Flash Cache	dbm02coladm03	View
CRITICAL	Storage Server Check	The callioth configuration on flash memory devices should match Oracle best practices	dbm02coladm03	View

InfiniBand Switch

Status	Type	Message	Status On	Details
CRITICAL	Switch Check	System is exposed to Exadata Critical Issue IS - All InfiniBand Switches		View



Exadata管理工具:

exawatcher工具:

exawatcher工具: 在原有OSWatcher的基础上增加了对infiniBand、存储磁盘等相关系统资源的监控, 从11.2.3.3版本开始, Oracle重新设计了一款全新系统信息收集及诊断工具**ExaWatcher**来代替OSWatcher, 主要用来收集操作系统资源使用情况(采样频率默认5秒, 保留一周时间), 用来诊断节点驱逐等故障, 判断是否因为操作系统资源问题导致故障发生或当数据库出现性能问题时提供相关的系统资源使用信息。

查看ExaWatcher工具的帮助信息:

```
/opt/oracle.ExaWatcher/ExaWatcher.sh --help
```

关闭ExaWatcher工具:

```
#/opt/oracle.ExaWatcher/ExaWatcher.sh --stop (系统默认是启动的)
```

启动ExaWatcher工具:

```
#/opt/oracle.ExaWatcher/ExaWatcher.sh --fromconf
```

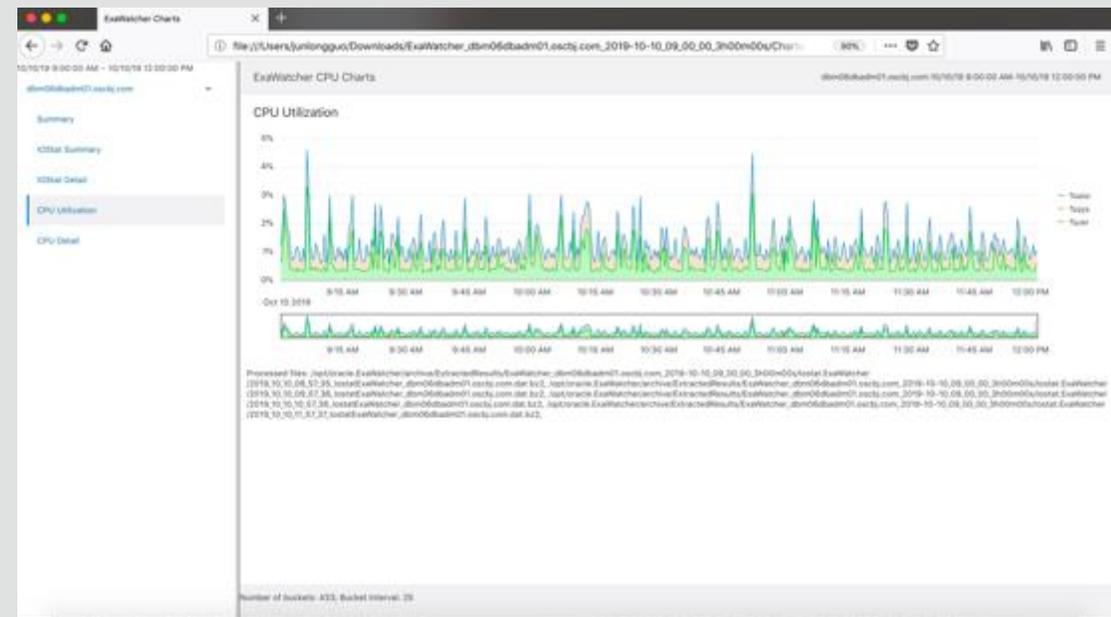
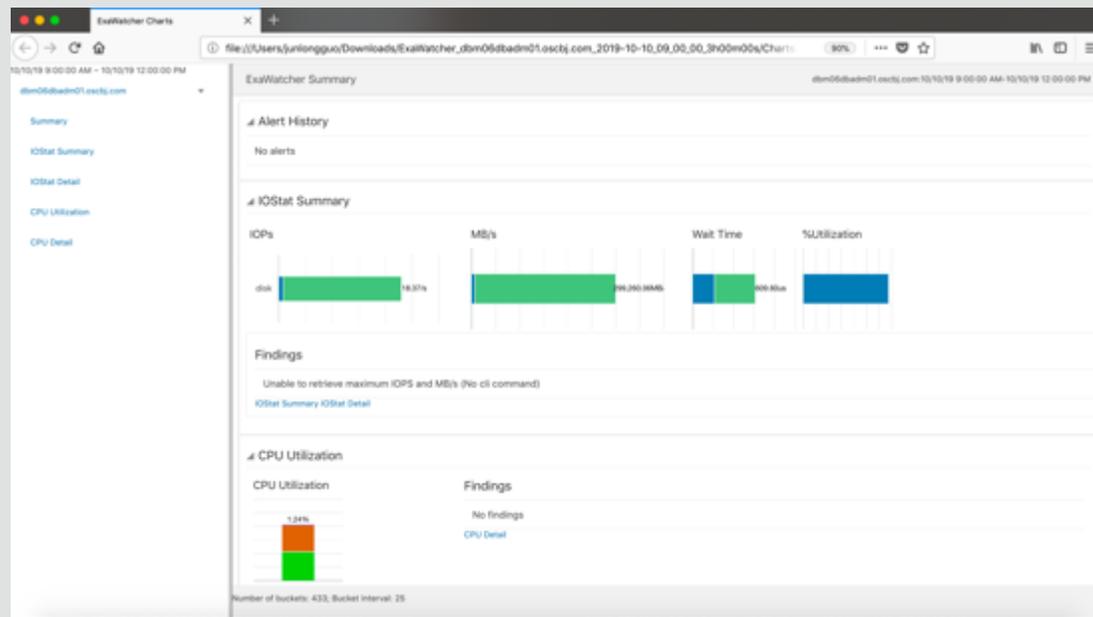
收集某个时间段的ExaWatcher日志:

```
#/opt/oracle.ExaWatcher/GetExaWatcherResults.sh --from 2020-05-20_9:00:00 --to 2020-05-20_12:00:00
```

Exadata管理工具:

exawatcher工具:

使用GetExaWatcherResults.sh命令收集某个时间段的ExaWatcher日志，命令执行完之后，会产生一个后缀为.tar.bz2的压缩文件，此文件默认放在/opt/oracle.ExaWatcher/archive/ExtractedResults目录下，下载到本地并解压，使用浏览器打开Charts.ExaWatcher.<hostname>/<timestamp>_<duration>/index.html:



注意: 打开index.html时需要确保自己的电脑能够上公网。



Exadata管理工具:

exacli工具:

exacli工具: 它可以运行在计算节点也可运行在存储节点上, 可以在一个节点上运行来远程的管理其它的计算节点或存储节点, 有的客户不允许使用**SSH**登录存储节点或计算节点来运行**cellcli**或**dbmcli**命令, 或禁用节点上的**root**访问或**SSH**服务, 这样就需要使用**exacli**工具(**exacli**只能一次管理一台主机, 无法同时并行管理多个主机)。

exacli命令格式: `exacli -c [username@]HOST [-l <username>] [--xml] [--cookie-jar [<jarfilename>]] [-e <command>]`

exacli需要使用**java1.8**或更高版本, 在使用**exacli**时首先设置正确的**JAVA_HOME**环境变量, 还需要在远程节点上先创建角色, 给角色分配权限, 然后创建用户再给用户分配角色, 在计算节点使用**dbmcli**完成, 在存储节点使用**cellcli**完成, 具体操作如下:

Exadata管理工具:

exacli工具:

存储节点设置用户和角色:

```
CELLCLI>create role administrator
```

```
CELLCLI>grant privilege all actions on all objects all attributes with all options  
to role administrator
```

```
CELLCLI>create user celladministrator password=XXXXXX
```

```
CELLCLI>grant role administrator to user celladministrator
```

```
CELLCLI>
```

计算节点设置用户和角色:

```
DBMCLI>create role administrator
```

```
DBMCLI>grant privilege all actions on all objects all attributes with all options to  
role administrator
```

```
DBMCLI>create user dbnodeadministrator password=XXXXXX
```

```
DBMCLI>grant role administrator to user dbnodeadministrator
```

```
DBMCLI>
```

Exadata管理工具:

exacli工具:

查看存储节点cell及celldisk状态:

```
#exacli -l celladministrator -c dbm06celadm01 --cookie-jar -e 'list cell;list celldisk'
```

查看数据库节点database server的详细信息:

```
#exacli -l dbnodeadministrator -c dbm06dbadm01 --cookie-jar -e 'list dbserver detail'
```

说明: **-c**指定需要维护主机, **-l**指定登录用户, **-e**指定要执行的命令或批处理脚本, **--cookie-jar[filename]**将用户名和密码发送到远程节点进行认证, 认证成功后把这个cookie以文件方式保存在本主机上, 如没指定文件名则默认存储到本机的 **\$HOME/.exacli/cookiejar**文件中, 此保留会在**24**小时后自动失效, 失效前再次输入命令时不会出现输入密码的提示, 失效后要求再次输入密码进行认证。

部分**cellcli**的功能不能使用**exacli**命令来完成, 如: 重启/启动/关闭存储节点服务、用户/权限/角色管理、**calibrate/describe**命令等。

Exadata管理工具:

exadcli工具:

如果各个节点关闭SSH功能，则dbmcli/cellcli/dcli不能执行（都需要具备SSH功能），只能用exacli工具，但是exacli只能一次管理一台主机，无法同时并行管理多个主机，要实现同时管理多个主机那就只能使用exadcli工具，此工具只能执行exacli的相关命令，其它命令（如shell命令）不能使用exadcli工具执行，要执行exadcli工具，必须也先在远程节点上设置好角色和用户：

查看多个存储节点database server的状态：

```
#exadcli -c cellnode1,cellnode2,cellnode3 -l celladministrator list cell detail
```

或：

```
#exadcli -g cell_group -l celladministrator list cell detail
```

查看多个数据库节点database server的状态：

```
#exadcli -c dbnode1,dbnode2 -l dbnodeministrator list dbserver detail
```

或：

```
#exadcli -g db_group -l dbnodeministrator list dbserver detail
```

◆ Exadata 运维管理

- 1 Exadata 组件管理
- 2 Exadata 状态检查
- 3 Exadata 管理工具
- 4 Exadata 案例分享

运维案例分享 (1) :

ASM磁盘组RECOC1的默认CachingPolicy导致其上数据对象IO性能差

现象描述:

在Exadata X8 HC上运行一套数据库，性能一直很好，刚刚上线了一个新的OLTP业务模块，有人反映此业务模块在业务高峰时性能很差，业务模块的前端界面十分卡顿，而其它先业务系统运行正常，需要运维人员帮忙分析原因。

原因分析:

先找业务人员，梳理整个过程，新业务模块采用新数据模型，所有业务表都是新创建的，并放到单独的新创建的表空间上，表的基础数据已在上线前批量导入，此时正好业务高峰，登录到数据库定位了几条与此业务模块相关的SQL语句，检查发现这几条语句执行时有大量的物理读，这是一个OLTP类型的模块，如此频繁操作的对象推测与flashcache缓存有关，查看flashcache中的内容:CellCLI> list flashcachecontent ...，发现与此业务模块相关表及索引都没有缓存到flashcache，进一步检查发现此业务模块对应表空间的数据文件创建在RECOC1磁盘组上，查看CellCLI> list griddisk attributes name,asmDiskgroupName,cachingPolicy 发现RECOC1的 cachingPolicy设置是none，表示默认情况下RECOC1磁盘上的数据文件不会使用flashcache资源，而DATAC1设置是default表示允许使用flashcache资源，由此基本定位问题是由于新创建表空间放到RECOC1上无法使用flashcache导致。

处理方法:

把此业务模块对应的表空间迁移到DATAC1磁盘组上，迁移之后，业务人员反映性能提升很大，前端界面很流畅，其实RECOC1的 cachingPolicy也可以设置为default，但是**不建议**这样做。

运维案例分享 (2) :

磁盘擦洗及修复操作导致存储IO很高

现象描述:

在Exadata X6 HC上刚刚迁移了一套数据库准备进行业务测试, 某天上午数据库管理登录OEM监控Exadata的使用情况, 发现3个CELL节点的磁盘使用率很高, 然而检查数据库的负载却很低, 基本上没运行什么SQL语句, 而业务人员也反馈前端业务程序由于要修改, 于前一天就全部停止了, 奇怪为何CELL节点的IO会这么高呢?

原因分析:

直接登录到CELL节点查看CELL节点的alert日志:
`/var/log/oracle/diag/asm/cell/<cell_nodename>/trace/alert.log`
在日志中发现: 当天8点时刻有大量Begin scrubbing CellDisk:CD_xx_dbm01celadm01, 原来是在做磁盘的擦洗和修复操作, 这个行为是CELL节点自动发启的, 旨在定期的检测和修复磁盘的数据坏块。

处理方法:

这是CELL节点的自动磁盘擦洗和修复特性, 它是CELL节点自动发启, 默认是两周运行一次, 8点发启操作, 可以使用命令来查看和修改此特性值:
查看磁盘擦洗操作开始时间: --默认8点开始
`CellCLI>list cell attributes harddiskscrubstarttime`
查看磁盘擦洗操作周期: --默认周期为每两周执行一次 (biweekly)
`CellCLI>list cell attributes harddiskscrubinterval`
如果磁盘擦洗操作正在进行, 需要立即停止且以后都不进行磁盘擦洗操作, 可以使用如下命令:
`CellCLI>alter cell harddiskscrubinterval=none` --为了确保数据安全, **建议不要关闭此特性**

运维案例分享 (3) :

FlashCache分配不合理导致数据库性能问题

现象描述:	在Exadata X5 HC上运行着两套OLTP数据库，刚上线时两套库对应的业务系统性能很好，比之前的传统架构性能提升了数倍，运行两个月后发现，一套库的性能突然变差，而前台人员反映业务高峰时情况越来越差，已有大量投诉，引起领导高度重视，要求尽快解决。
原因分析:	首先，收集此数据库业务高峰时段的AWR，在TOP事件中发现的“cell single block physical read”等待事件从刚上线时的1ms突然上升到当前的10ms，这是OLTP业务，先查看了一下flashcache的命中率， $\text{cell flash cache read hits/physical read total IO requests} * 100\%$ ，发现才83%（很低，理想应该接近100%），由于此Exadata上运行着两套库，接下来使用db_fc_by_allocated指标来查看每个数据库占用的flashcache的大小： <code>#cellcli -e list metriccurrent DB_FC_BY_ALLOCATED</code> ，结果发现性能好的库占用了90%的flashcache空间，而性能变差的库只用到了10%，至此基本定位是由于flashcache中分配问题导致此数据库的性能变差。
处理方法:	针对运行在Exadata上的两套库或多套库需要合理的分配flashcache的空间，这就需要规划好Exadata的IORM，通过设置iormplan来实现对flashcache资源的控制，设置IORM参考样例如下： <code>CellCLI>alter iormplan objective='auto',dbplan=((name=db1,share=5,flashcache=on,flashcachesize=xxxxG,flashcachemin=xxxxG), (name=db2,share=5,flashcache=on,flashcachesize=xxxxG,flashcachemin=xxxxG));</code>

运维案例分享（4）：

并行设置问题导致数据库CPU耗尽

现象描述：	一套业务数据库刚刚迁移到新购的Exadata X7 HC运行，割接当天晚上测试各业务功能都正常，第二天早上业务高峰时发现大量业务卡顿，前台已无法进行正常的处理流程，需要业务维护人员迅速定位并处理问题，否则后果不堪想象。
原因分析：	首先，通过监控平台收到主机的CPU有告警，登录数据库主机查看CPU使用率接近100%，此时查看数据库服务器进程发现有大量的ora_pxxx开头的并行进程，为何会持续的有这么多的并行进程，收集AWR报告定位Top SQL发现这些SQL中并没有加parallel之类的hint，进一步通过dba_tables，dba_indexes等视图查看SQL中涉及对象时，发现多个核心表及索引上面的degree字段都为32，后来查看昨天晚上的割接脚本发现，迁移核心大表时通过create table xxx nologging parallel 32 as select...方式迁移，创建索引使用create index xxx ... parallel 32命令；然而迁移完之后确没有把表及索引的并行度设置为noparallel，因此造成系统中启动了大量并行进程，并最终耗尽CPU。
处理方法：	把表及索引的并行度修改为noparallel: SQL>alter table xxx noparallel; SQL>alter index xxx noparallel; 注意： 这里需要强调一下，有些人员为了利用上Exadata上智能扫描以提升统计分析类查询的性能，就把SQL语句都加上并行16或更高的hint，这对高负载的数据库来说很危险，其实一般根据实际情况设置为2即可，还有就是一定要注意规范，如果有一个完善的规范完全可以避免此类事故的发生。

运维案例分享 (5) :

大页设置导致DB节点无法启动

现象描述:	在X8-2 HC上做不同级别配置 (CPU/MEM) 的性能测试: 原DB节点配置为48Core/768G, 调整为24Core/384G之后, 重启数据库主机结果系统无法启动。
原因分析:	首先, 通过登录到DB节点的iLOM: <code>#ssh root@ilom_ip</code> 重启系统: <code>->reset /SYS</code> 登录远程终端: <code>->start /SP/console</code> 检查发现启动过程有报错提示有许多服务/文件无法启动, 卡住不能继续启动, 接下来重启系统到单用户模式下进一步排查问题, 通过journalctl命令查看日志信息, 发现有无法分配内存的报错, 检查大页设置为400G, 而当前内存设置为384G, 找到原因是大页设置过大而导致系统无法启动。
处理方法:	启动到单用户: 当启动到系统引导条目“Exadata_DBM_0: LINUX_BOOT_0”时按e进行编辑, 提示要求输入用户名密码 (Enter username: root/Enter password: sos1Exadata), 在linux16 /vmlinuz-xxx.el7uek.x86_64...这一行的最后添加上single init=/bin/bash, 然后按Ctrl-x启动进入到单用户模式, 以读写方式挂载根目录 (<code>#mount -o remount,rw /</code>), 接着修改/etc/sysctl.conf文件把其中的大页参数vm.nr_hugepages设置为合理值 (比ASM实例加所有数据库实例的SGA大点即可), 最后输入exit重启, 系统正常启动。再次强调一下SGA很大、连接进程又特别多时, 强烈建议数据库SGA使用大页来提升性能 (最佳实践use_large_pages=ONLY), 但是操作系统上大页的设置一定要合理。

For More Information

[Oracle.com/Exadata](https://www.oracle.com/Exadata) (External)
[Exadata.us.oracle.com](https://exadata.us.oracle.com) (Internal)
[Docs.oracle.com](https://docs.oracle.com)
