Capacity Planning at Scale

Kapil Goyal
Fidelity Investments
kapil.goyal@fmr.com
02/02/2017

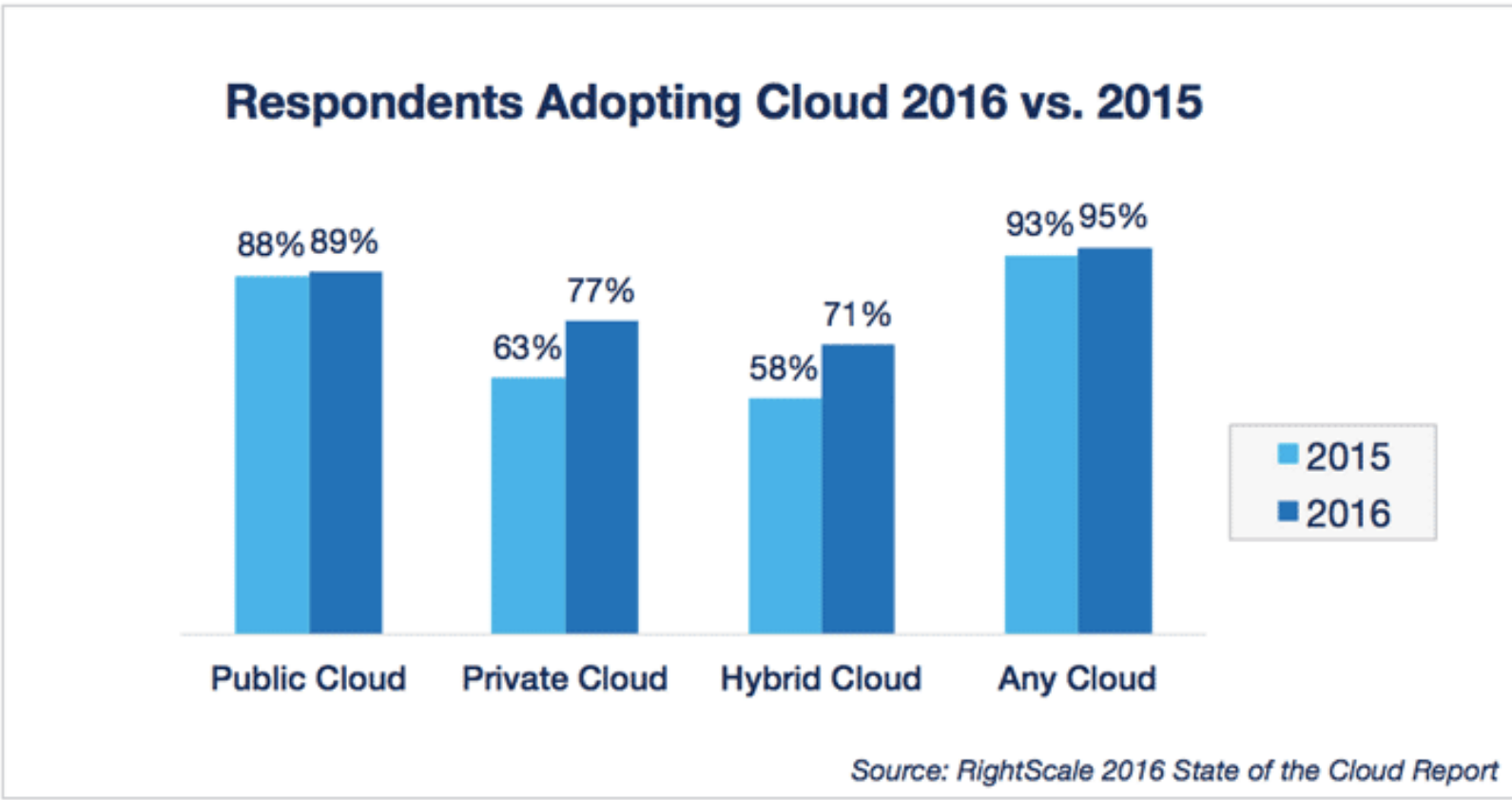
Who am I?

- Lead for Oracle database performance team at Fidelity Investments
- Worked with Oracle Consulting and got exposure to many different companies' database performance challenges
- Speaker at DOUG, Hostos, ECO, Enkitec Expo, IOUG and AIOUG
- Published articles for DOUG and IOUG SELECT magazine
- Certified OCP DBA 8i to 11g

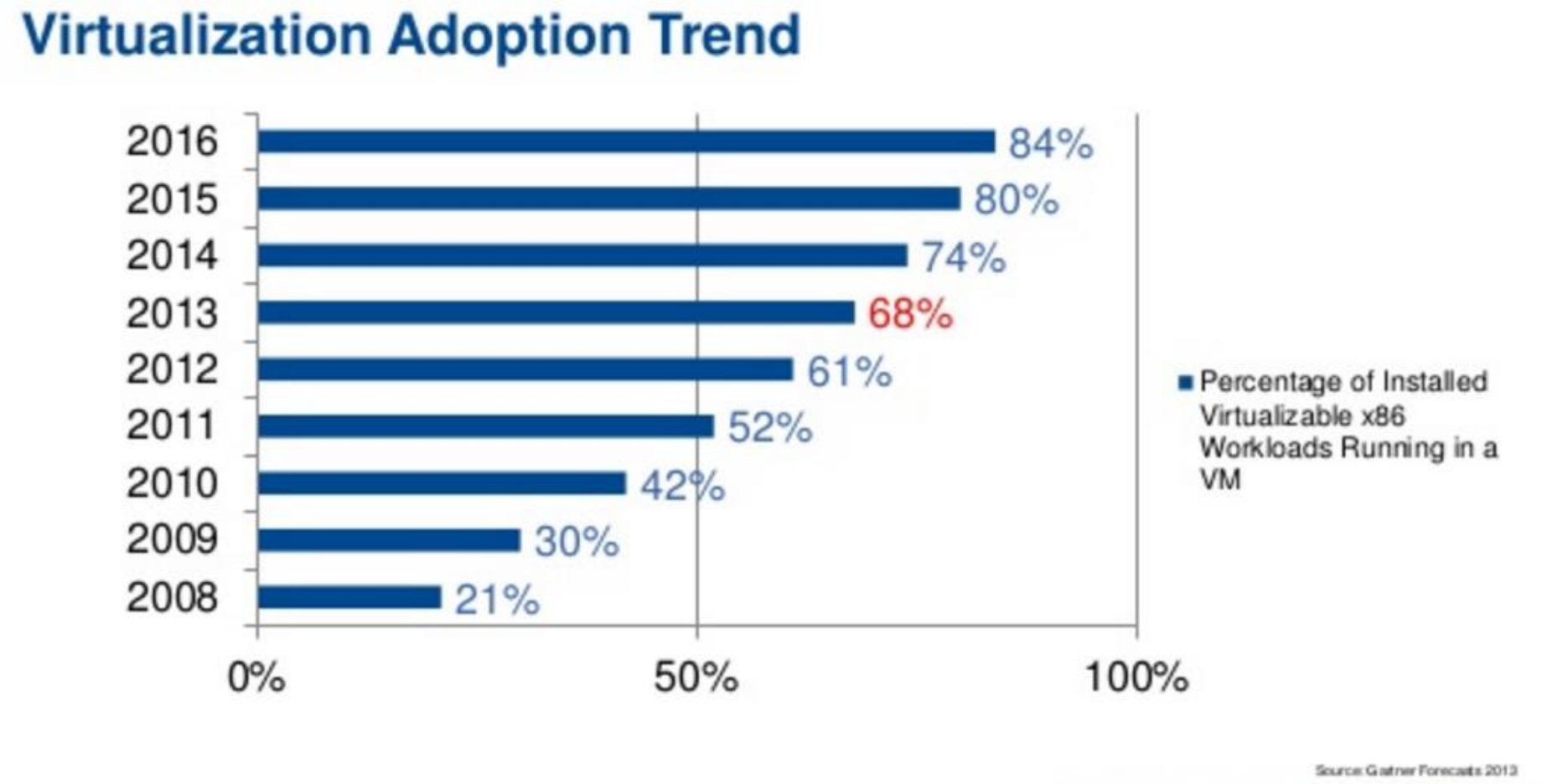
Agenda

- Some industry stats on Cloud Adoption
- Know your Exadata resource utilization
- Introduction to OEM Metrics
- How to extract data from OEM Views
- Single View across all Exadata
- Introduction to AWR Warehouse
- How to stack rank databases
- Which Oracle databases are right fit for cloud(public/private)

Significant Growth in Hybrid Cloud Adoption

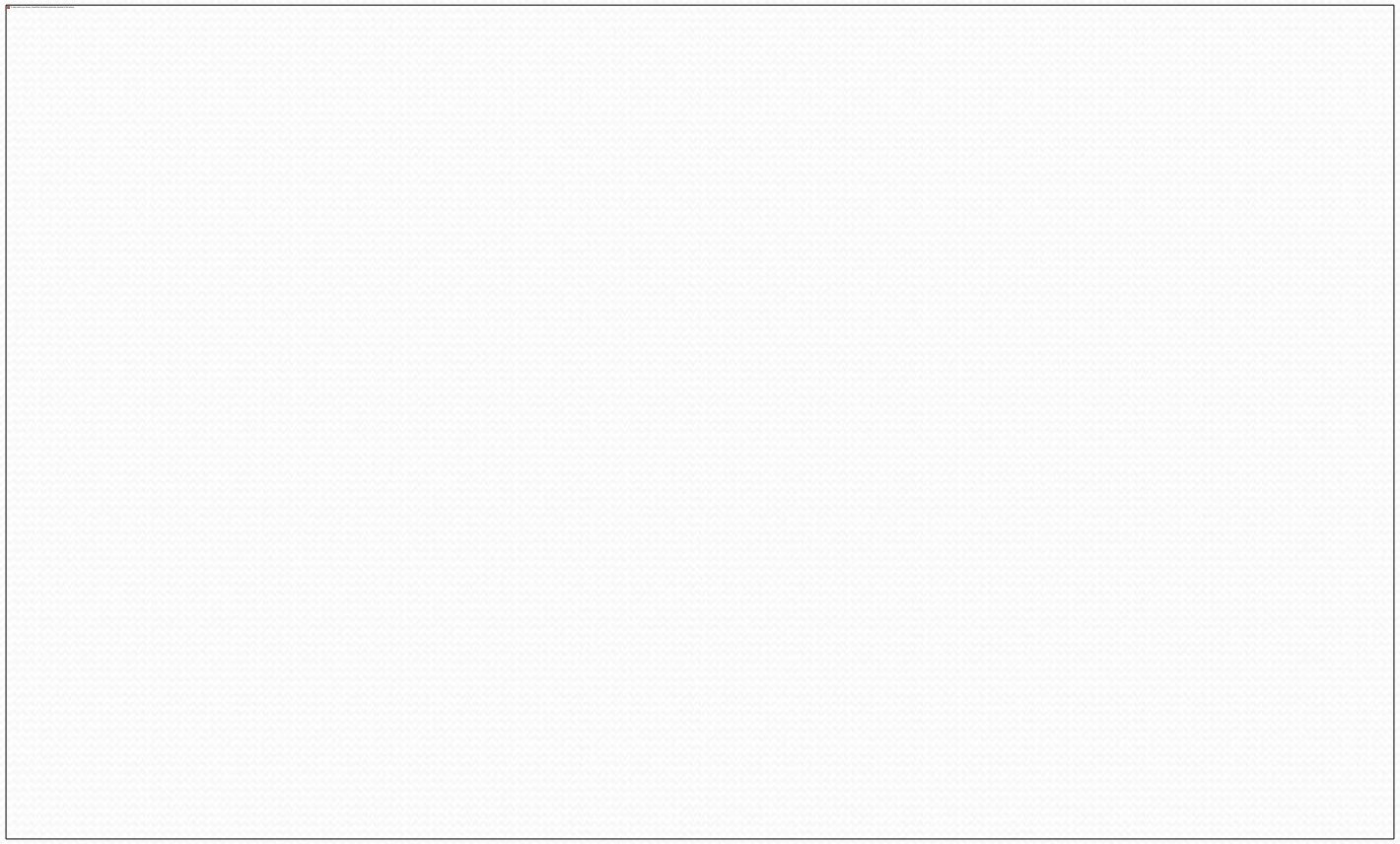


Virtualization adoption trend

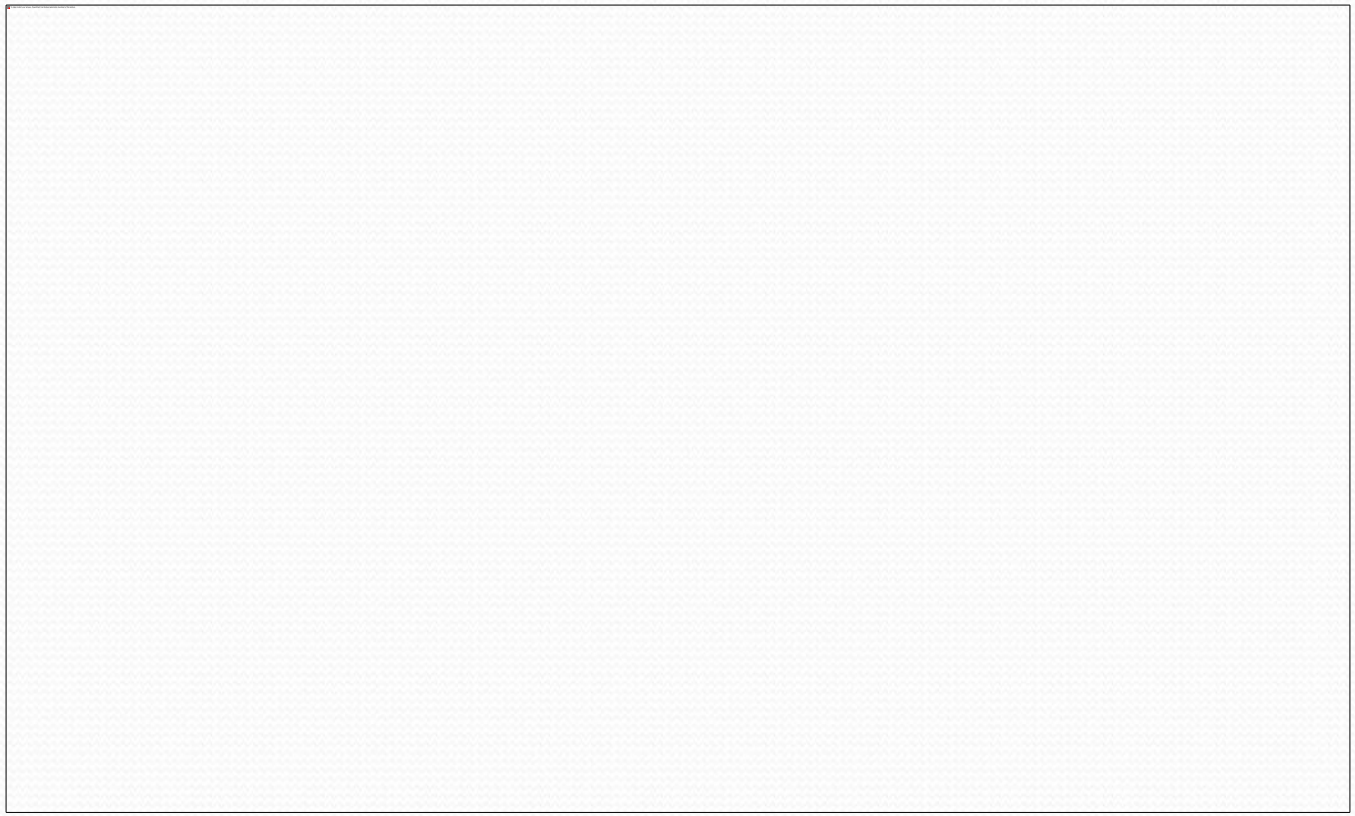


Cloud Adoption

- *SaaS* (63%, up one point) remains the most popular type of cloud computing in use



Achieving Business Goals



What is capacity planning?

"Capacity Planning is the process of predicting when future load levels will saturate the system and determining the most cost- effecting way of delaying system saturation as much as possible"

- by Daniel A. Menasce and Virgilo A.F. Almeida
(Authors of Capacity Planning for Web Services: Metrics, Models & Methods)

Thinking clearly about performance

"If you have performance problems, then you don't need to be spending your time with mathematical models; you need to be spending your time fixing those problems by either ***rescheduling load, eliminating load, or increasing capacity.***"

-Cary Millsap (Thinking clearly about performance)

Question

- How many of you have a dedicated capacity team?
- Or a performance team?
- At Fidelity we have a end-to-end capacity/performance team
  - Database
  - OS
  - SAN
  - Network

Capacity Planning

- Exadata
- Standard Oracle databases

- Do you know:
  - Current resource utilization?
  - If you are hitting the HDD IO limit?
  - If you are hitting the FLASH IO limit or not?
  - Compute Node CPU utilization?
  - Cell Node CPU utilization?

Problem Statement

- Bought x2-2 full rack 6 years ago
- Need to do hardware refresh, should we upgrade to x5-2 full rack or go to Exadata cloud offering?

Common mistakes for Exadata hardware upgrade

- We have seen proposals to upgrade in same class (e.g. x2-2 half rack ->x5-2 half rack or x2-2 Full->x5-2 Full) without looking into current utilization or future needs
- Keep in mind, X5-2 is much more powerful than x2-2

Just compare datasheets

Metric	X2-2 Full Rack (HC)	X5-2 Full Rack (HC)
Flash bandwidth - GB/s	68	140
Flash read IOPS	1,500,000	4,144,000
Flash write IOPS		2,688,000
Flash data capacity (raw) - TB	5.3	89.6
Effective Flash cache capacity - TB[1]		Up to 672 TB
HC Disk bandwidth - GB/s	18	20
HC Disk IOPS	28,000	33,000
HC Disk data capacity (raw) - TB	504	672
HC Disk data capacity (usable) -TB	224	300
HC Data load rate - TB/hr	12	21.5
DB Compute Nodes	8	8
DB CPU Cores	96	288
DB Memory - GB	768	2048
DB Memory Expandable - GB	1152	6144
Storage Cell Nodes	14	14
Cell CPU Cores	168	224

[1] Effective Flash Capacity is larger than the physical flash capacity and takes into account the high flash hit ratios due to Exadata's intelligent flash caching algorithms, and the size of the underlying disk storage. It is the size of the data files that can often be stored in Exadata and be accessed at the speed of flash memory

Source -
http://www.oracle.com/technetwork/database/exadata/dbmachine-x2-2-datasheet-175280.pdf
http://www.oracle.com/technetwork/database/exadata/exadata-x5-2-ds-2406241.pdf

Try to reduce load

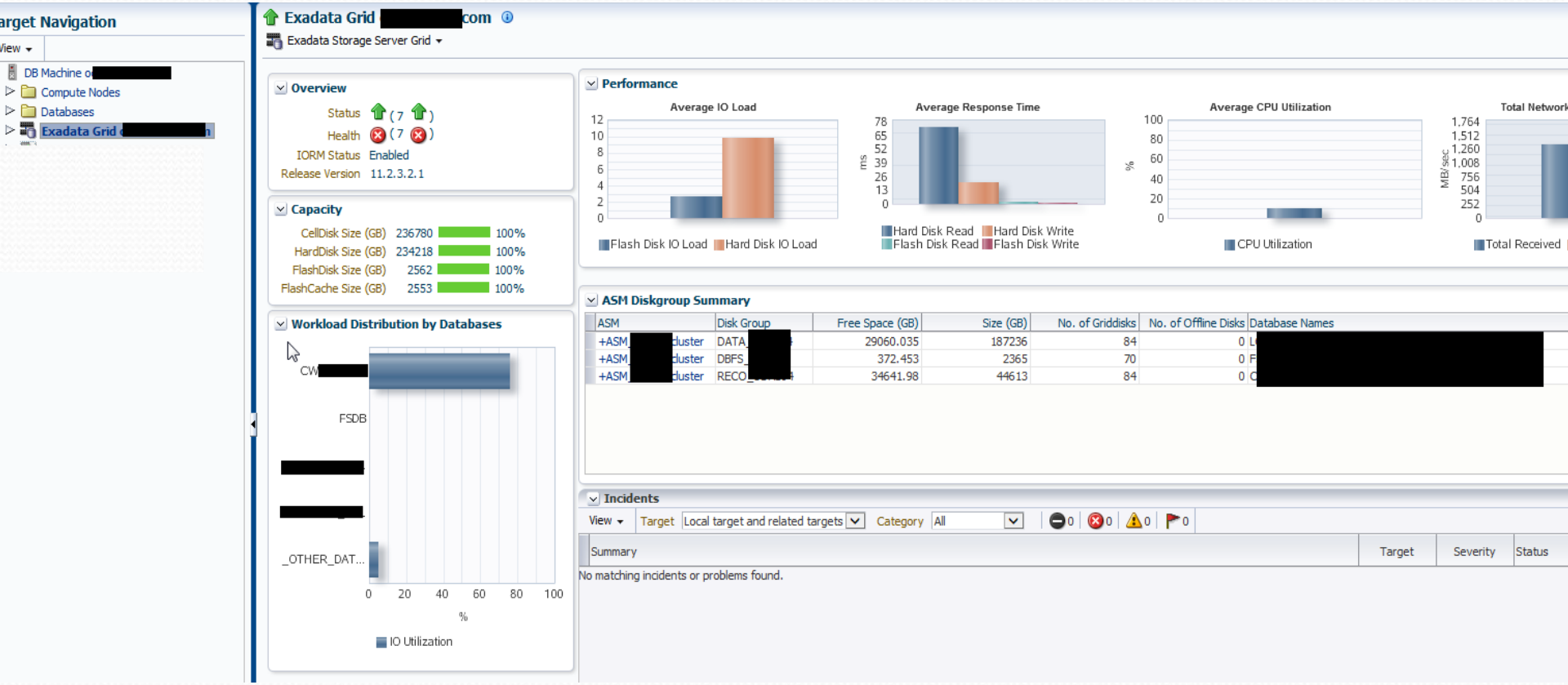
- At Fidelity, we have implemented hugepages and reduced 8-12% compute node CPU (in lab we reduced ~17% for a 2 node RAC)
- Also in some cases pagetable size grew to 60 GB without hugepages and with hugepages under ~1 GB
- *Enabled write back flash cache and reduce HDD disk IOPS
- Evaluate Exachk recommendations
- Control parallelism (pxhcdr.sql from Carlos Sierra)
- Enable IORM

* If database is write intensive

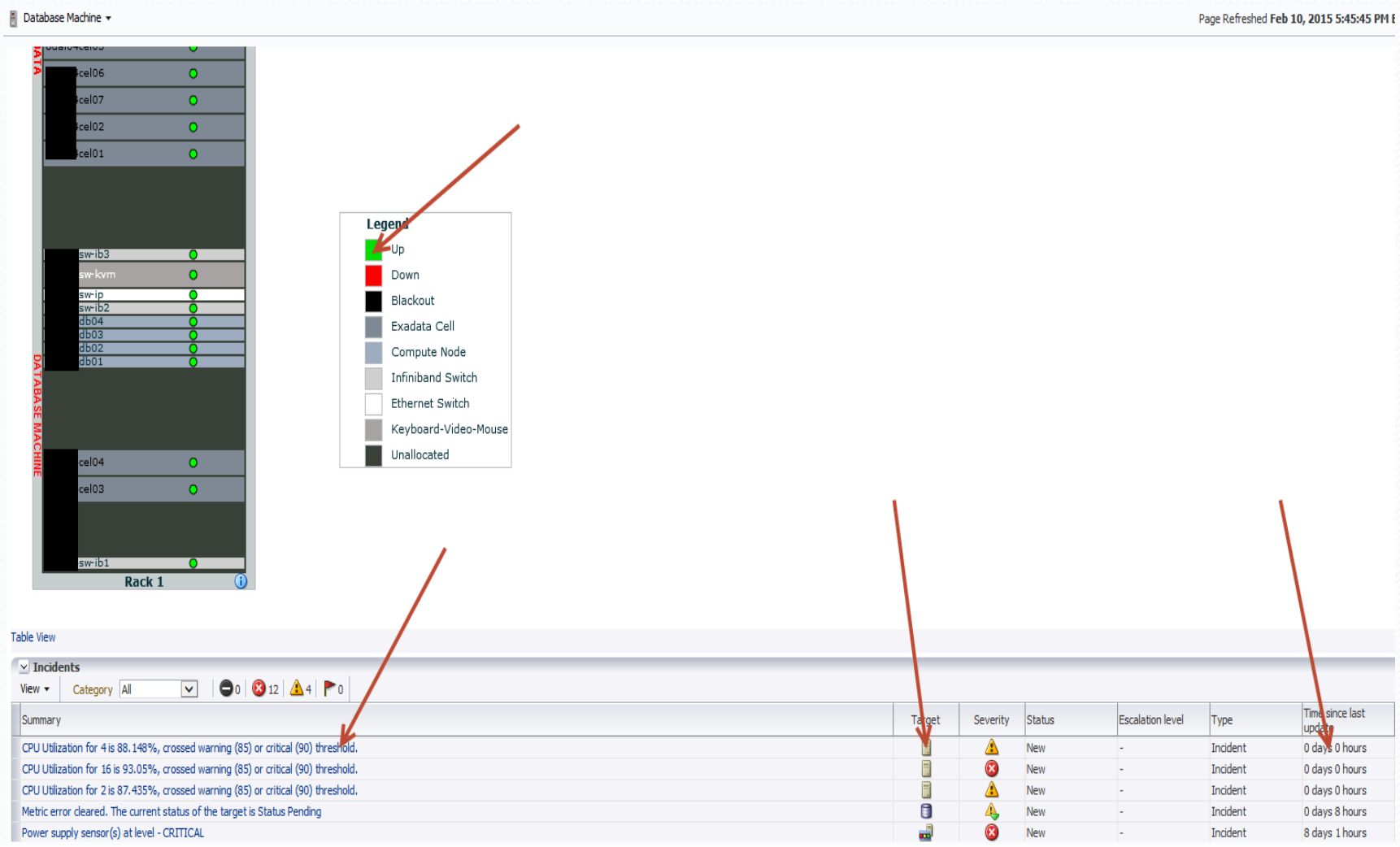
Where to find Exadata resource utilization trend?

- Data is available in EM12c
  - Exadata plugin is required

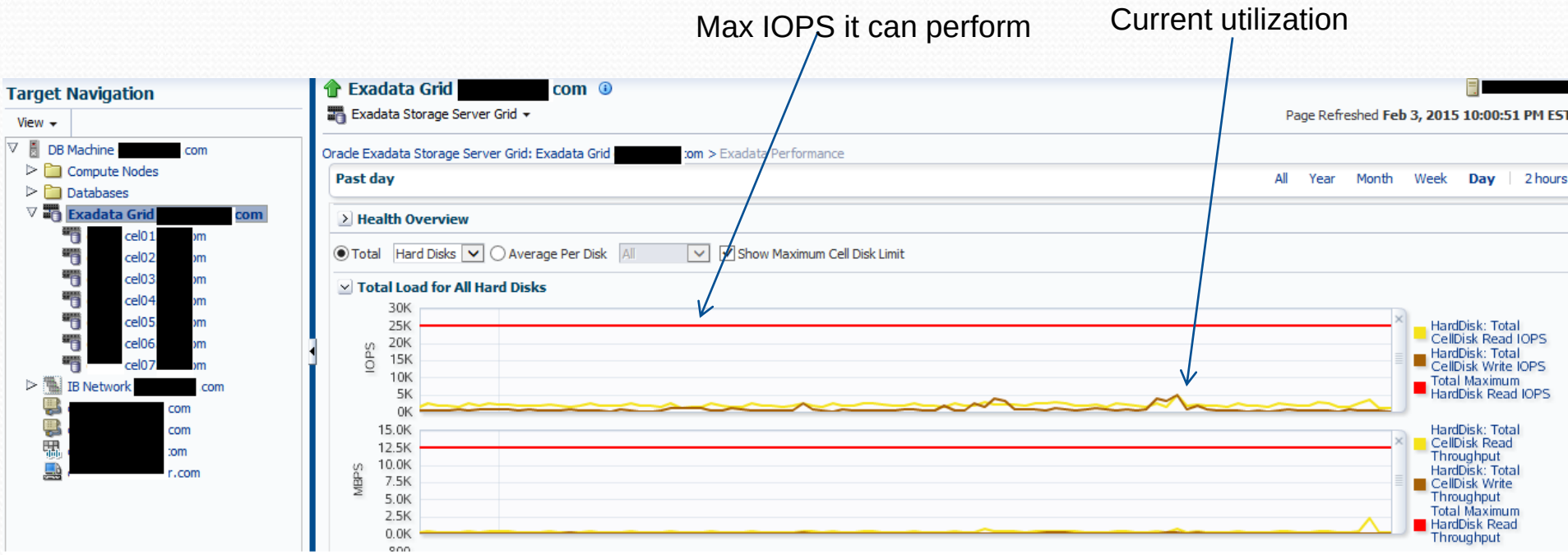
EM12c Snapshot(1)



Snapshot(2)


Database Machine
Page Refreshed Feb 10, 2015 5:45:45 PM E
cel06
cel07
cel02
cel01
sw-ib3
sw-kvm
sw-ip
sw-ib2
db04
db03
db02
db01
cel04
cel03
sw-ib1
Rack 1
Legend
Up
Down
Blackout
Exadata Cell
Compute Node
Infiniband Switch
Ethernet Switch
Keyboard-Video-Mouse
Unallocated
Table View
Incidents
View
Category All
0
12
4
0
Summary
Target
Severity
Status
Escalation level
Type
Time since last update
CPU Utilization for 4 is 88.148%, crossed warning (85) or critical (90) threshold.
New
Incident
0 days 0 hours
CPU Utilization for 16 is 93.05%, crossed warning (85) or critical (90) threshold.
New
Incident
0 days 0 hours
CPU Utilization for 2 is 87.435%, crossed warning (85) or critical (90) threshold.
New
Incident
0 days 0 hours
Metric error cleared. The current status of the target is Status Pending
New
Incident
0 days 8 hours
Power supply sensor(s) at level - CRITICAL
New
Incident
8 days 1 hours

Snapshot(3)


Max IOPS it can perform
Current utilization
Target Navigation
View
DB Machine [redacted] com
Compute Nodes
Databases
Exadata Grid [redacted] com
cel01 [redacted]
cel02
cel03
cel04
cel05
cel06
cel07
IB Network [redacted] com
Exadata Grid [redacted] com
Exadata Storage Server Grid
Page Refreshed Feb 3, 2015 10:00:51 PM EST
Oracle Exadata Storage Server Grid: Exadata Grid [redacted] :om > Exadata Performance
Past day
All Year Month Week Day 2 hours
Health Overview
Total Hard Disks Average Per Disk All Show Maximum Cell Disk Limit
Total Load for All Hard Disks
IOPS
30K 25K 20K 15K 10K 5K 0K
HardDisk: Total CellDisk Read IOPS
HardDisk: Total CellDisk Write IOPS
Total Maximum HardDisk Read IOPS
MBPS
15.0K 12.5K 10.0K 7.5K 5.0K 2.5K 0.0K
HardDisk: Total CellDisk Read Throughput
HardDisk: Total CellDisk Write Throughput
Total Maximum HardDisk Read Throughput

But OEM data gets rolled up

EM$METRIC_VALUES	Detailed raw data for 7 days	Most useful but retention is less (highest granularity)
EM$METRIC_VALUES_HOURLY	Aggregated into hourly average for 32 days	Decent useful
EM$METRIC_VALUES_DAILY	Daily average of hourly aggregated – 24 months	sort of meaning less ☺

- Retention can be changed but it gets changed for all the data collection - Doc ID 1405036.1
- We pull Exadata specific data from gc$metric_values to our own repository

Table 13-3 Core EM Metric Data Tables and Default Data Retention in the Management Repository

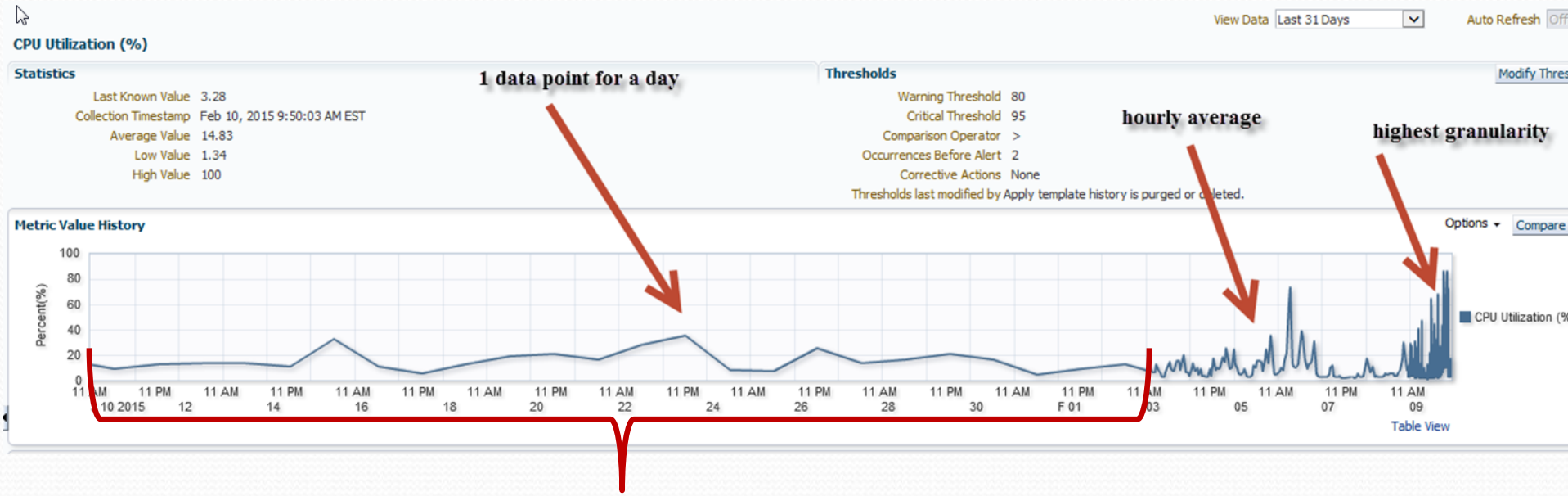
Table Name	Partitions Retained	Partition Size
EM_METRIC _VALUES	7	DAY
EM_METRIC_VALUES_HOURLY	32	DAY
EM_METRIC_VALUES_DAILY	24	MONTH

Check the current value of the retention periods:

```
SQL> select table_name, partitions_retained
from em_int_partitioned_tables
where table_name in ('EM_METRIC_VALUES','EM_METRIC_VALUES_HOURLY','EM_METRIC_VALUES_DAILY');

TABLE_NAME                 PARTITIONS_RETAINED
-------------------------  -------------------
EM_METRIC_VALUES                             7
EM_METRIC_VALUES_HOURLY                     32
EM_METRIC_VALUES_DAILY                      24
```

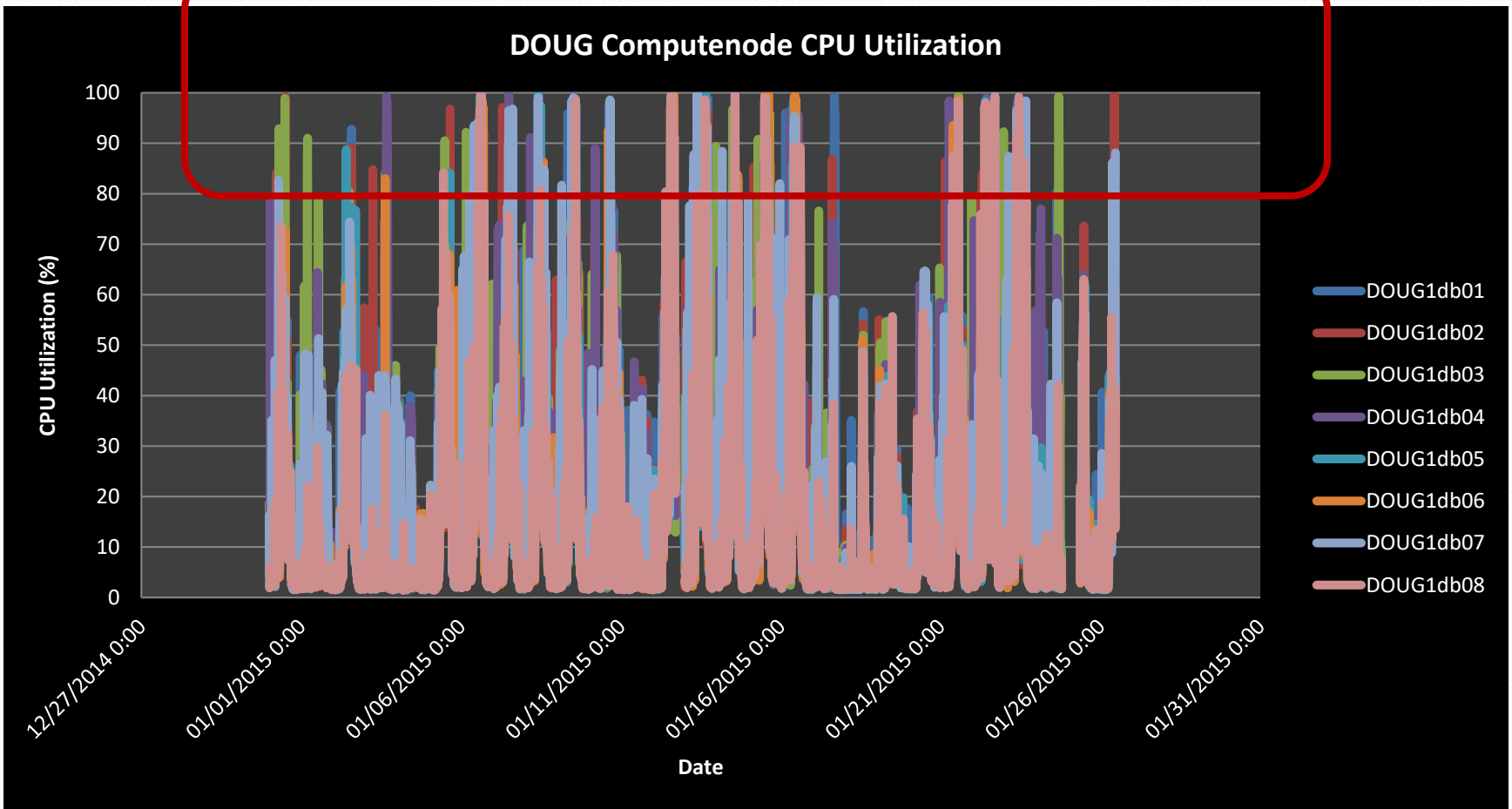
OEM data rolled up



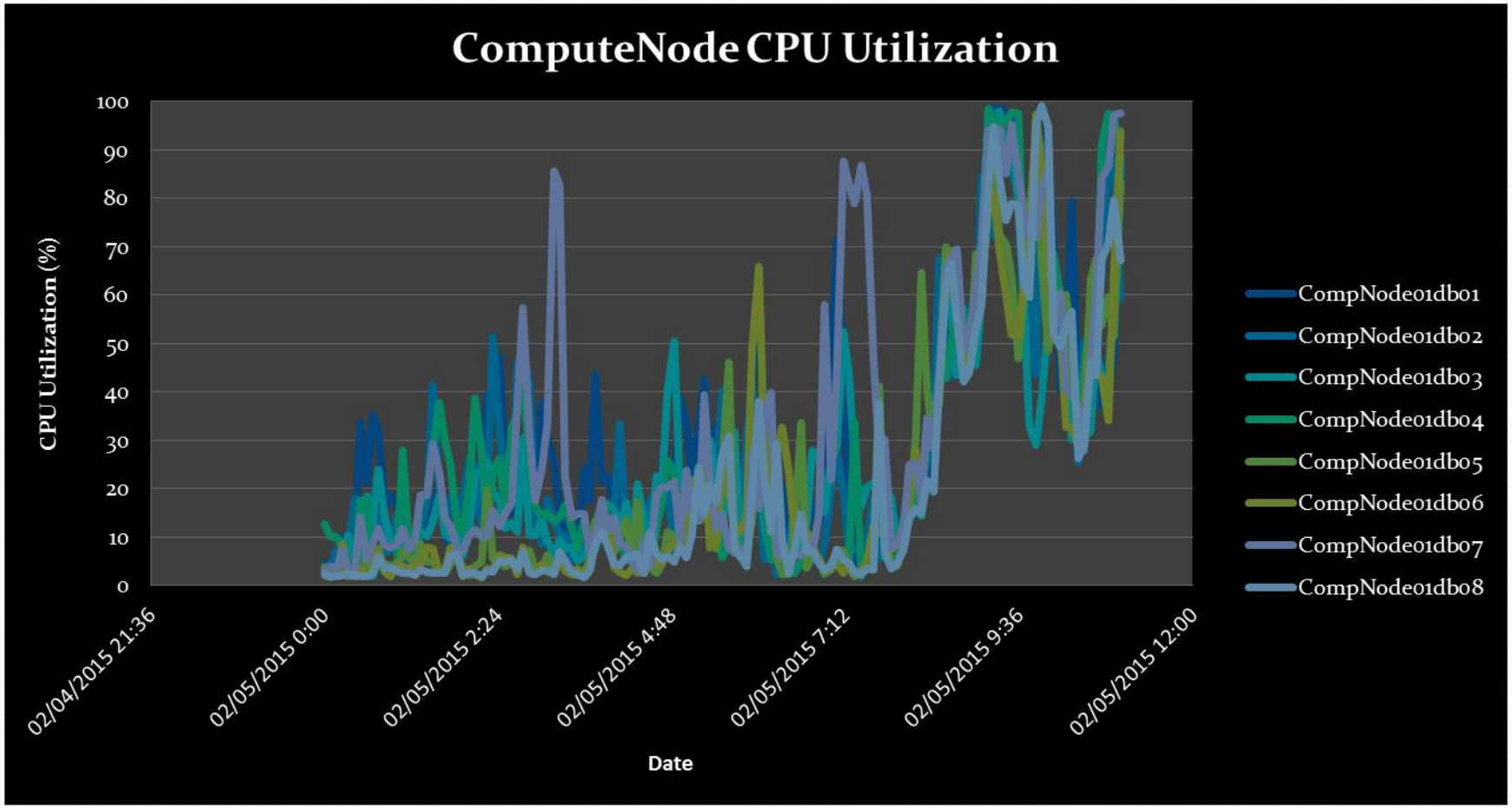
Daily average says CPU utilization% <40

Monthly 5 minutes average CPU

In reality everyday 90+% spike



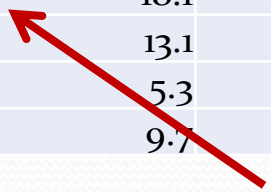
Compute Node CPU Utilization-for one day



How to get trend of any OEM Metric?

Host level CPU utilization% raw data

METRIC_DATE	CompNode01db01	CompNode01db02	CompNode01db03	CompNode01db04
02/05/2015 0:00	4.3	4.2	2.4	12.8
02/05/2015 0:05	4.5	3.8	3.6	10.1
02/05/2015 0:10	4.6	7.6	3.7	9.9
02/05/2015 0:15	9.9	7.9	7.1	8.2
02/05/2015 0:20	5.3	7.4	10.5	2.5
02/05/2015 0:25	7.5	18.1	4.6	3.1
02/05/2015 0:30	33.7	13.1	6.4	17.6
02/05/2015 0:35	21.3	5.3	18.4	17.5
02/05/2015 0:40	35.3	9.7	12.7	2



Raw Data from gc$metric_values, same source as OEM chart

It is same data on following OEM Page:

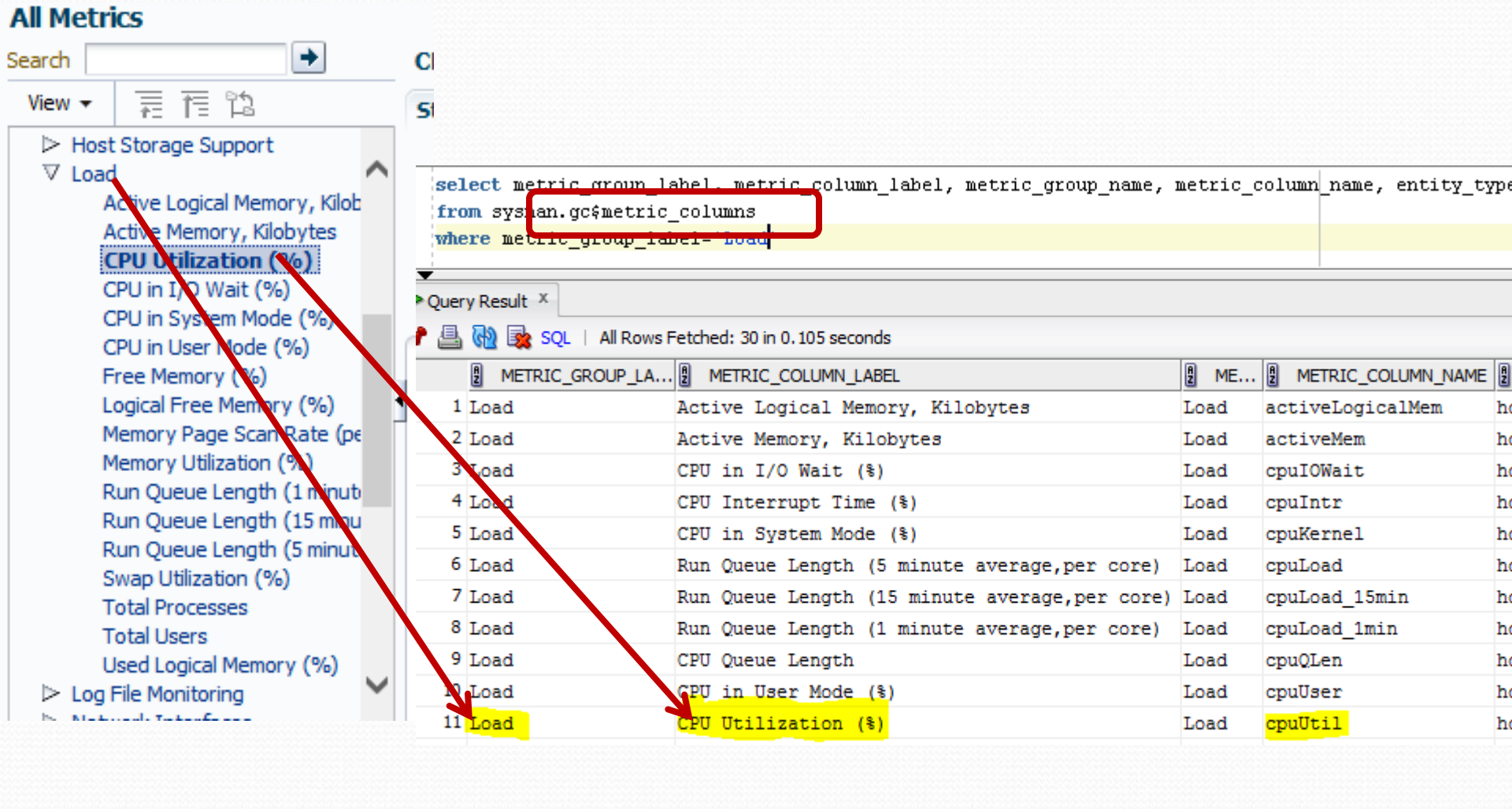
How to write SQL to get any metric's detailed data?

SQL:

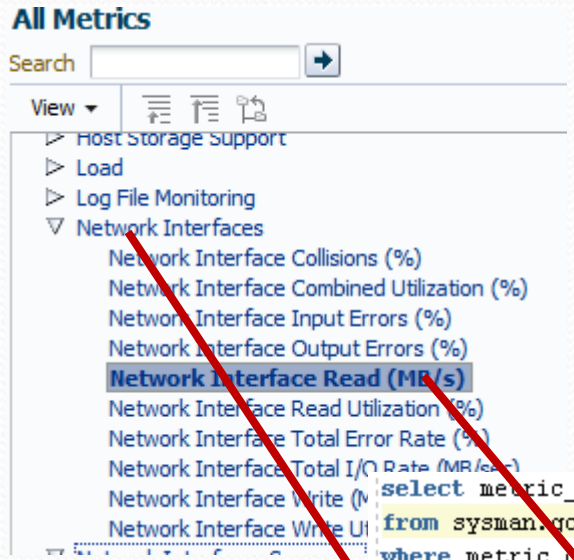
```
select
trunc(collection_time,'hh24') + (trunc(to_char(collection_time,'mi')/5)*5)/24/60 AS METRIC_DATE,
max(case WHEN entity_name ='CompNode01db01.abc.com' then  round(value,1) else null end)    "CompNode01db01",
max(case WHEN entity_name ='CompNode01db02.abc.com' then  round(value,1) else null end)    "CompNode01db02",
max(case WHEN entity_name ='CompNode01db03.abc.com' then  round(value,1) else null end)    "CompNode01db03",
max(case WHEN entity_name ='CompNode01db04.abc.com' then  round(value,1) else null end)    "CompNode01db04"
FROM GC$METRIC_VALUES a
where metric_group_name in ('Load' )
AND  metric_column_name in 'cpuUtil')
AND entity_type= 'host'
and ( entity_name like  'CompNode01db0%')
group by trunc(collection_time,'hh24') + (trunc(to_char(collection_time,'mi')/5)*5)/24/60
ORDER BY 1
```

Query against any metric

Just identify metric_group_name and metric_column_name



- Another Example




```
select metric_group_label, metric_column_label, metric_group_name, metric_column_name, entity_type
from sysman.gc$metric_columns
where metric_group_label='Network Interfaces'
```

Query... x

SQL | All Rows Fetched: 37 in 0.116 seconds

	METRIC_GROUP_LABEL	METRIC_COLUMN_LABEL	METRIC_GROUP_NAME	METRIC_COLUMN_NA
1	Network Interfaces	Network Interface Collisions (%)	Network	collisions
2	Network Interfaces	Network Interface Input Errors (%)	Network	inpErrors
3	Network Interfaces	Network Interface Name	Network	interface
4	Network Interfaces	Network Interface Output Errors (%)	Network	outErrors
5	Network Interfaces	Network Interface Read (MB/s)	Network	readRate

Cell Node Statistics

- HDD IOPS/Throughput
- Flash IOPS/Throughput
- HDD/Flash read/write latency
- CPU Utilization

SQL Script

```
select distinct metric_group_label
from sysman.gc$metric_columns
where metric_group_label like '%Exadata%' or metric_group_label like '%Cell%' order by 1
```

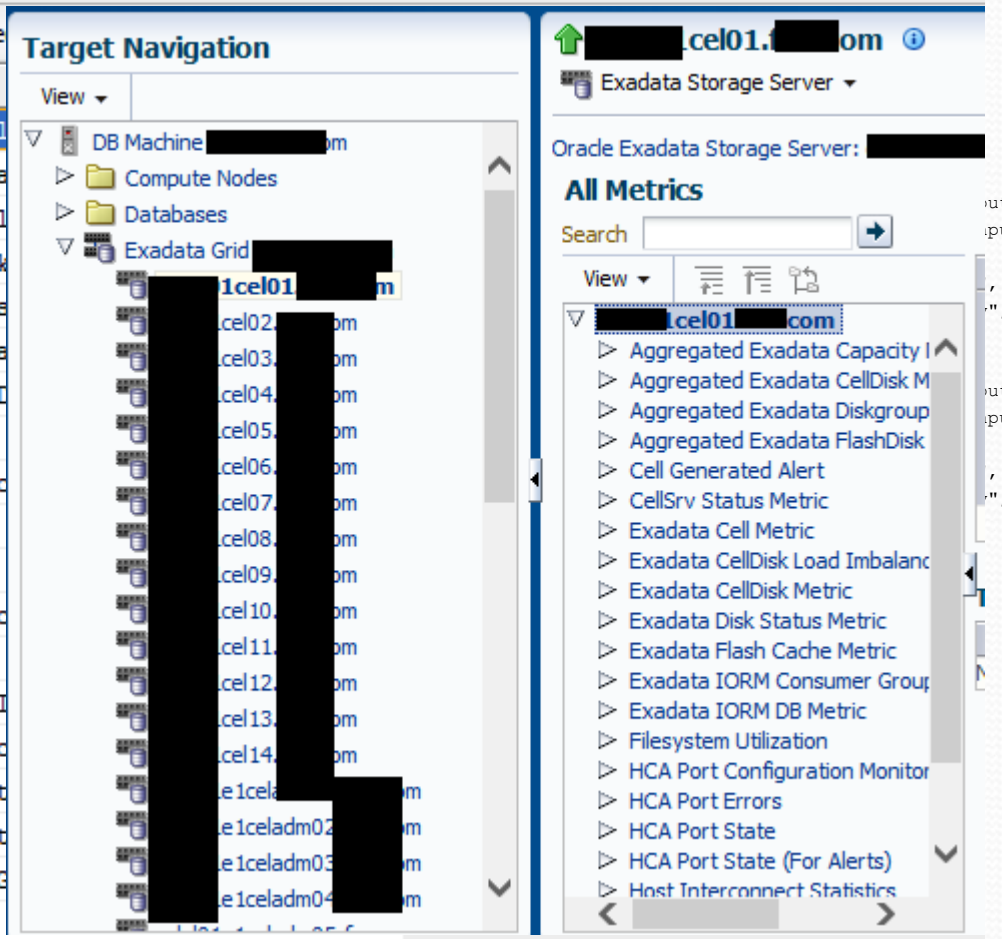
```
alter session set nls_date_form

select
--SUBSTR (entity_name,1, INSTR(
 trunc(collection_time,'hh24')
--metric_group_name,
--entity_name,
sum(case WHEN metric_column_nam
sum(case WHEN metric_column_nam
sum(case WHEN metric_column_nam
sum(case WHEN metric_column_nam
sum(case WHEN metric_column_nam
round(avg(case WHEN metric_colu
round(avg(case WHEN metric_colu
sum(case WHEN metric_column_nam
sum(case WHEN metric_column_nam
sum(case WHEN metric_column_nam
sum(case WHEN metric_column_nam
sum(case WHEN metric_column_nam
round(avg(case WHEN metric_colu
round(avg(case WHEN metric_colu
round(avg(case WHEN metric_colu
FROM sysman.gc$metric_VALUES
where metric_group_name in
( 'Aggregated_HardNFlashDisk_Me
AND  metric_column_name in(
      'sum_cd_read_iops',
      'sum_cd_write_iops',
      'sum_cd_read_throughput',
      'sum_cd_write_throughput'
      'sum_cd_io_load',
      'avg_cd_read_latency',
      'avg_cd_write_latency'
      ,'cpu_utilization'
      )
AND entity_type= 'oracle_exadat
and ( entity_name like  'DOUGce
group by trunc(collection_time,'hh24') + (trunc(to_char(collection_time,'mi')/15)*15)/24/60
ORDER BY 1,2
```

Query...

SQL | All Rows Fetche

	METRIC_GROUP_LABEL
1	Agg_Exadata_System_Cell
2	Aggregated Exadata Capa
3	Aggregated Exadata Cell
4	Aggregated Exadata Disk
5	Aggregated Exadata Flas
6	CELL CellDisk Configura
7	CELL Flash Cache Cell D
8	Cell Configuration
9	Cell Configuration Patc
10	Cell Generated Alert
11	CellSrv Status Metric
12	Exadata Capacity Metric
13	Exadata Cell Metric
14	Exadata CellDisk Load I
15	Exadata CellDisk Metric
16	Exadata Disk Status Met
17	Exadata Flash Cache Met
18	Exadata IORM Consumer G
19	Exadata IORM DB Metric

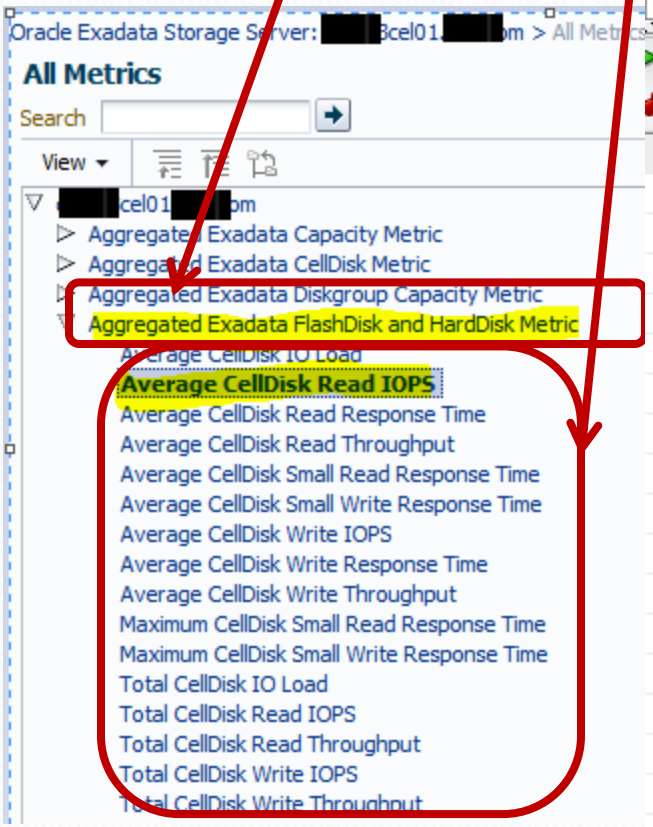


Base Script - Kapil Goyal (Fidelity Investments), Greg O'Hearn (Fidelity Investments) and Kalyan Maddali (Oracle)

In OEM, look for group_label and column_label then map with gc$metric_column's group_name and column_name

metric_column_label

metric_group_label




```
select metric_group_label, metric_column_label,metric_group_name,metric_column_name
from sysman.gc$metric_columns
where metric_group_label='Aggregated Exadata FlashDisk and HardDisk Metric' order by 1
```

Query Result

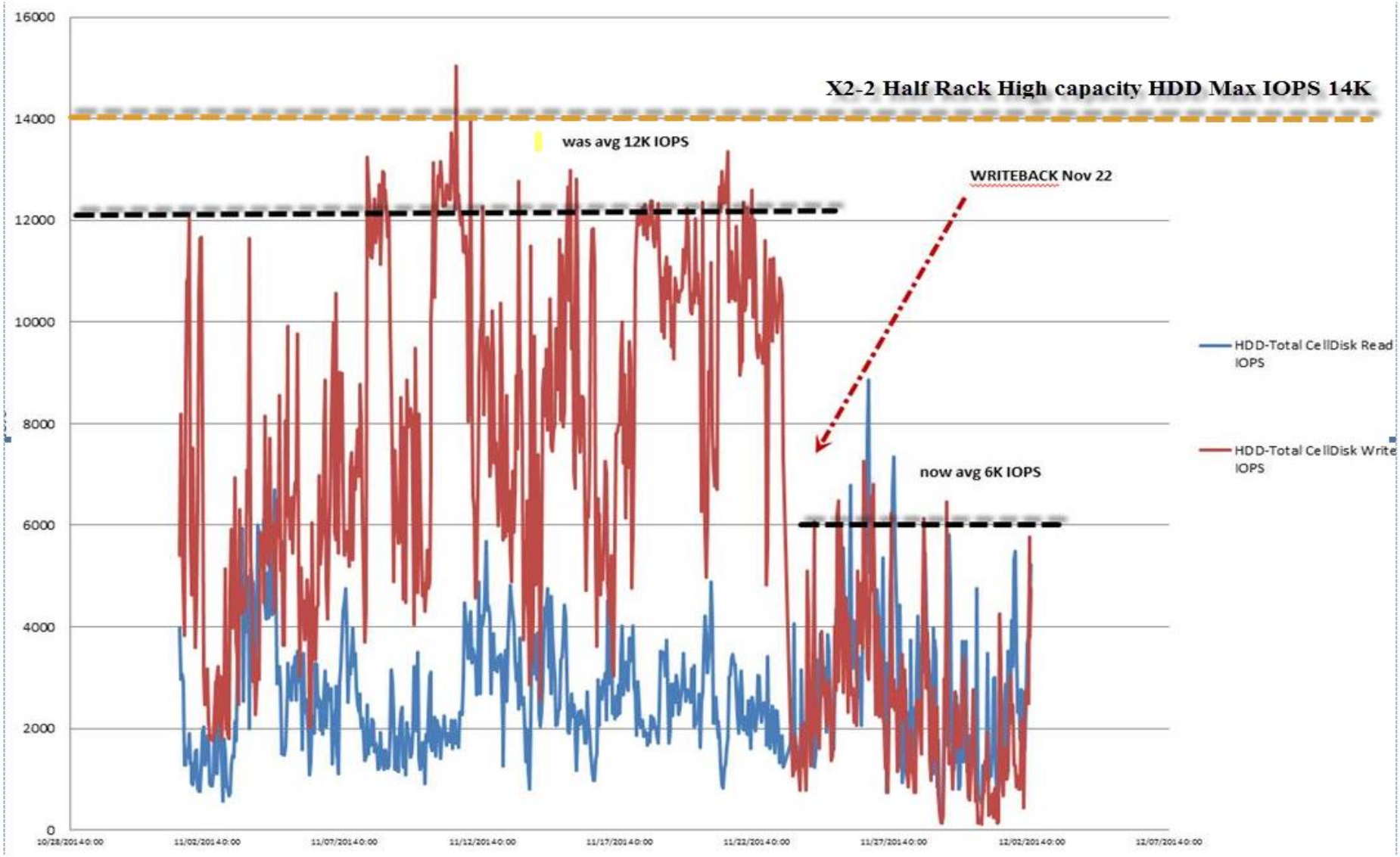
SQL | All Rows Fetched: 17 in 0.106 seconds

	METRIC_GROUP_LABEL	METRIC_COLUMN_LABEL
1	Aggregated Exadata FlashDisk and HardDisk Metric	Average CellDisk IO Load
2	Aggregated Exadata FlashDisk and HardDisk Metric	Average CellDisk Read IOPS
3	Aggregated Exadata FlashDisk and HardDisk Metric	Average CellDisk Read Response Time
4	Aggregated Exadata FlashDisk and HardDisk Metric	Average CellDisk Read Throughput
5	Aggregated Exadata FlashDisk and HardDisk Metric	Average CellDisk Small Read Response
6	Aggregated Exadata FlashDisk and HardDisk Metric	Average CellDisk Small Write Response
7	Aggregated Exadata FlashDisk and HardDisk Metric	Average CellDisk Write IOPS
8	Aggregated Exadata FlashDisk and HardDisk Metric	Average CellDisk Write Response Time
9	Aggregated Exadata FlashDisk and HardDisk Metric	Average CellDisk Write Throughput
10	Aggregated Exadata FlashDisk and HardDisk Metric	CellDisk Type
11	Aggregated Exadata FlashDisk and HardDisk Metric	Maximum CellDisk Small Read Response
12	Aggregated Exadata FlashDisk and HardDisk Metric	Maximum CellDisk Small Write Response
13	Aggregated Exadata FlashDisk and HardDisk Metric	Total CellDisk IO Load
14	Aggregated Exadata FlashDisk and HardDisk Metric	Total CellDisk Read IOPS
15	Aggregated Exadata FlashDisk and HardDisk Metric	Total CellDisk Read Throughput
16	Aggregated Exadata FlashDisk and HardDisk Metric	Total CellDisk Write IOPS
17	Aggregated Exadata FlashDisk and HardDisk Metric	Total CellDisk Write Throughput

Why Cell Statistics are important?

- If you enable/change IORM, enable/disable write back flash cache etc. you want to how it is impacting
- IO is not distributed properly across all cells (very unlikely)
- Workload is not distributed across different racks/expansions

Example(1) – HDD IOPS reduction after enabling writeback flash cache



Write path from Database perspective

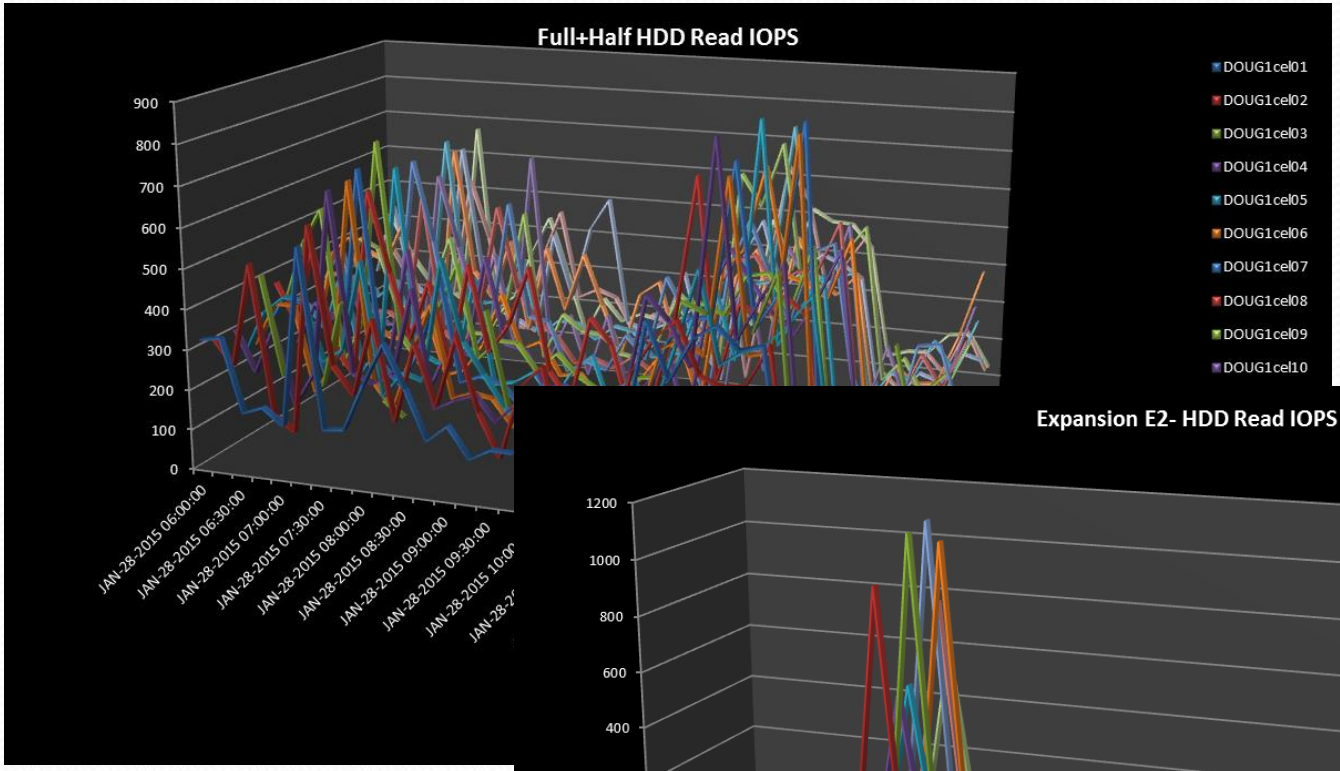
- Log writes- (LGWR)
  - Upon every transaction “commit”
  - Latency is #1 priority
- Exadata cell behavior: Flash logging
  - Write to DISK and Flash .. Ack on first completion
- Flushing the Buffer Cache – (DBWR)
  - Clean dirty buffers in SGA
  - IOPS bandwidth is key
- Exadata cell behavior
  - Writes to Flash
  - Write to Disk - > 128k
- Direct path writes -- (Oracle Shadow and PQ)
- Bandwidth is typically #1 priority
- Typically LARGE IO sizes
- Exadata cell behavior:
  - cell_flash_cache determines the path
    - Keep -> writes to FLASH
    - Default -> writes to DISK
- TEMP writes - spills out of PGA
  - Sort, Hash Join, etc.
  - Exadata cell behavior
    - Writes go to DISK

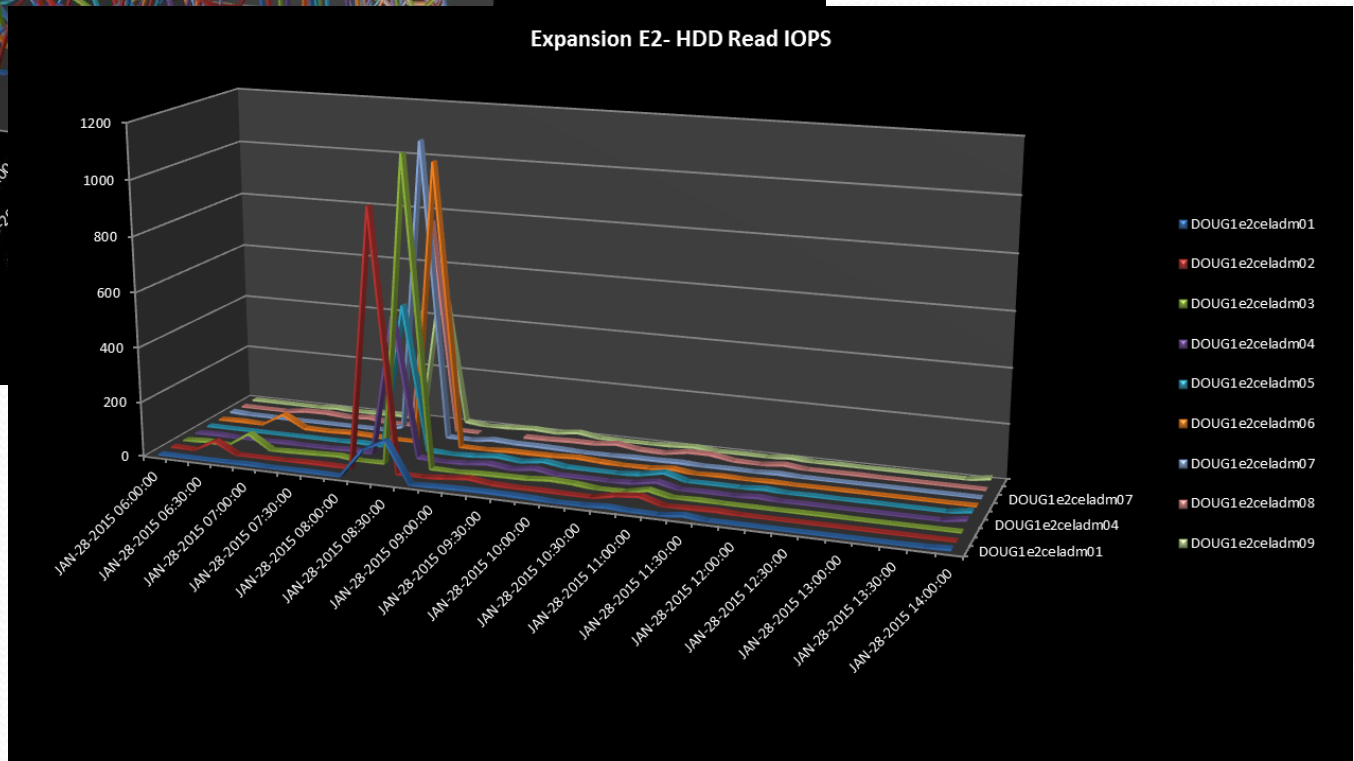
Write path from Database perspective… cont

- Archive Log writes -- (ARCH)
- – After REDO logs are switched, they
- are written to the “RECO” space by
- ARCH
- These are “LARGE” writes
- Exadata cell behavior:
  - Write to DISK only

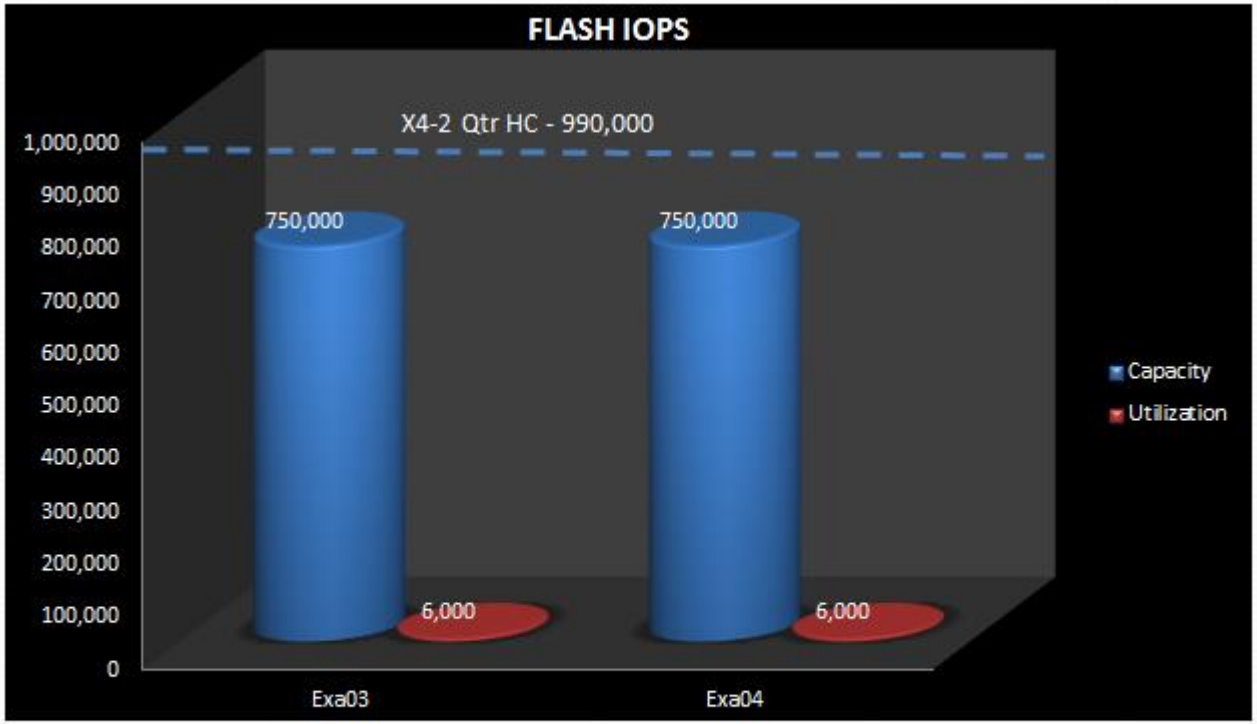
- FLASHBACK database – (RVWR)
- – Uses it’s own format, but volume is similar to “redo” writes
- Typically written to RECO
- Exadata cell behavior:
  - Write to DISK only

Example(2) –abnormal HDD IOPS across Full/Half/Expansion





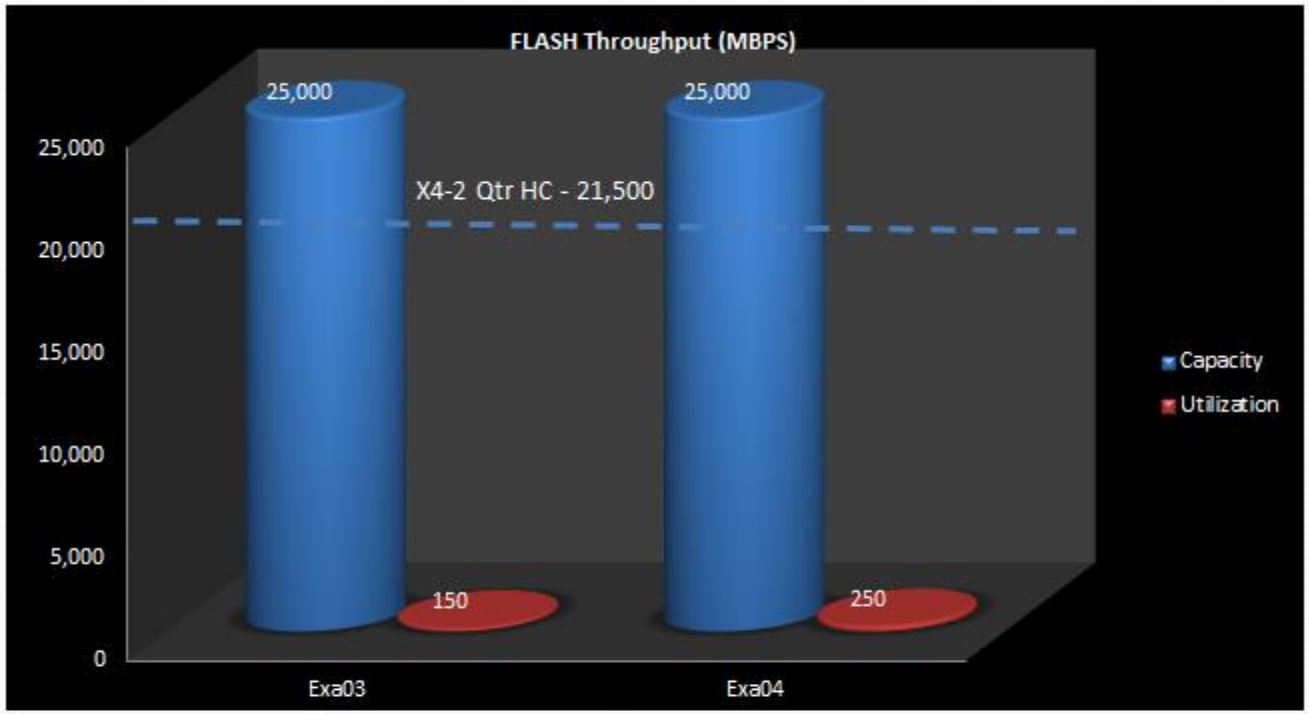
Sample Charts of Exadata Review-snapshot(1)



- The x2-2 half rack can perform 750K Flash IOPS
- XYZ prod and standby database uses 6K Flash IOPS (~.8%) – 98 percentile
- X4-2 Qtr HC can perform 990K Flash IOPS

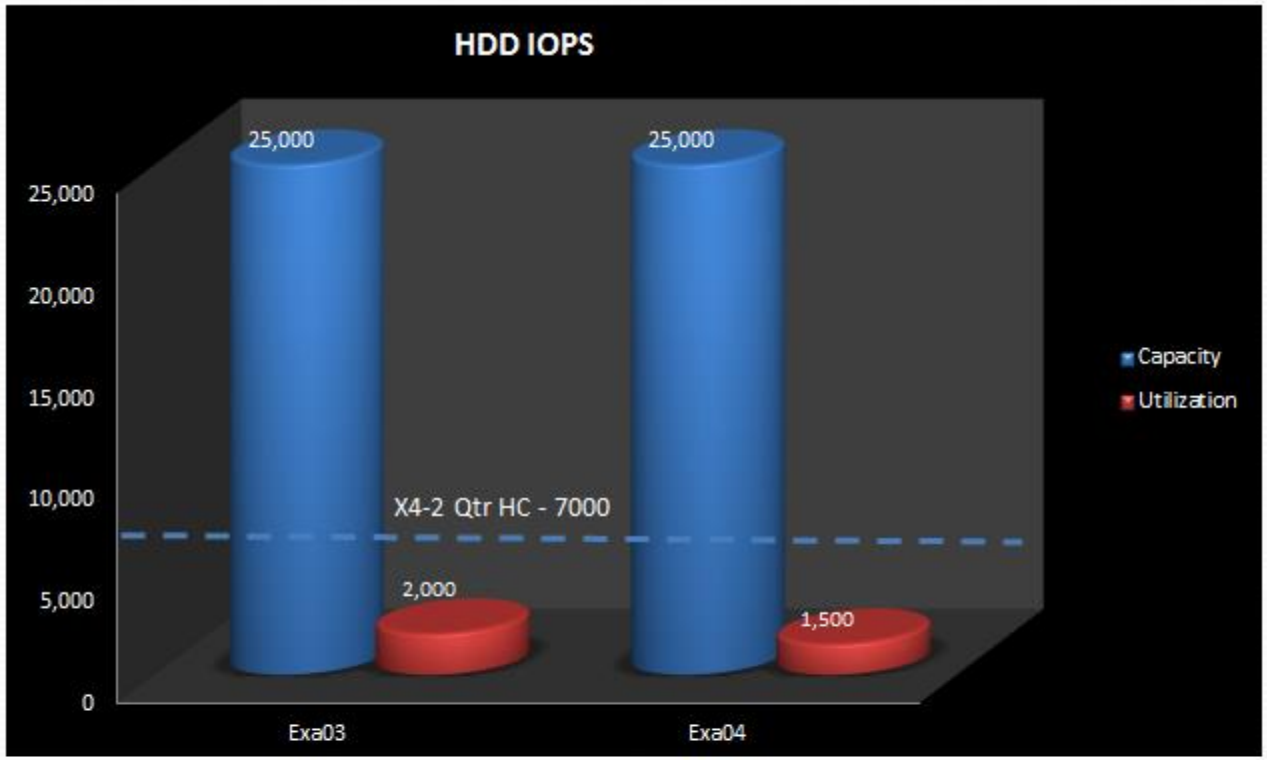
- Current – x2-2 Half Rack HP
- Utilization
- Proposed Exadata x4-2 qtr Rack HC

Snapshot(2)



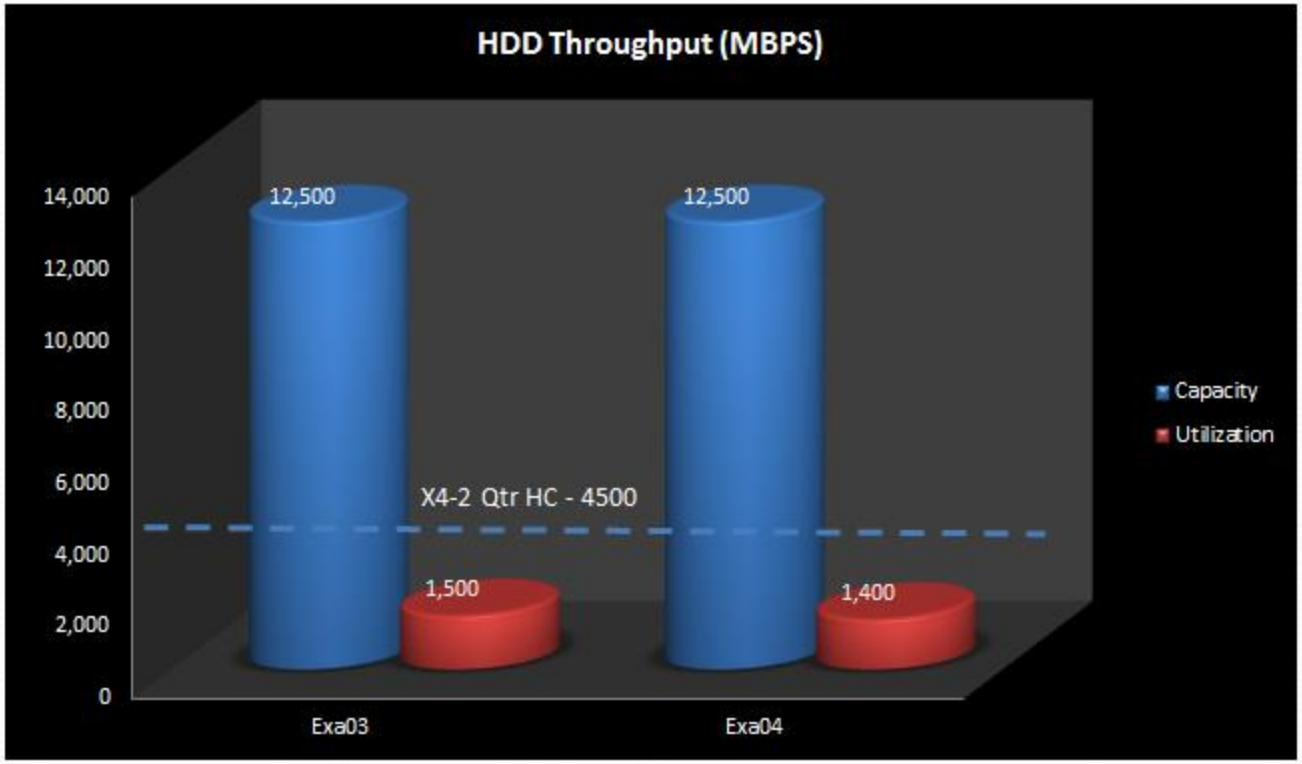
- The x2-2 half rack can perform 25K Flash MBPS Throughput
- XYZ prod database uses 150 Flash MBPS (~.6%) and standby uses 250 Flash MBPS (1%)- 98 percentile
- X4-2 Qtr HC can perform 21.5K Flash MBPS throughput

Snapshot(3)



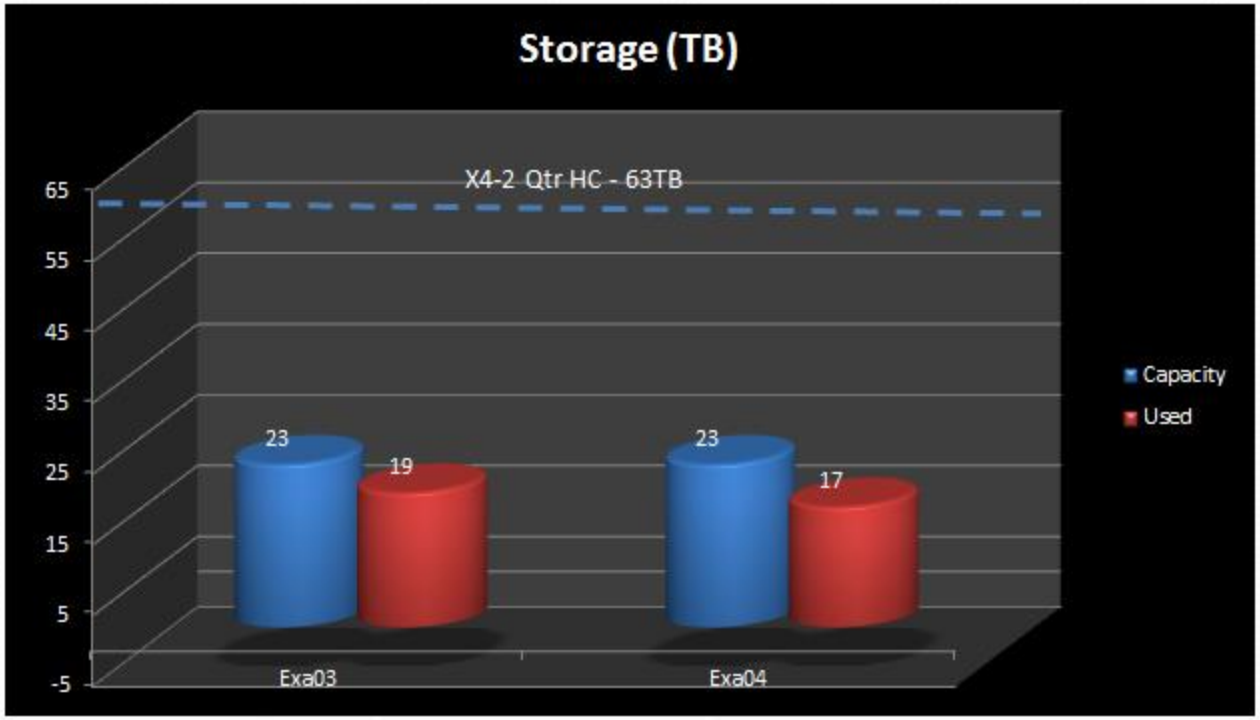
- The x2-2 HP half rack can perform 25K HDD IOPS
- XYZ prod database uses 2K (~8%) and standby uses1.5K (~6%) – 98 percentile
- This can be further reduced by leveraging more Flash
- X4-2 Qtr HC performs 7K HDD IOPS

Snapshot(4)



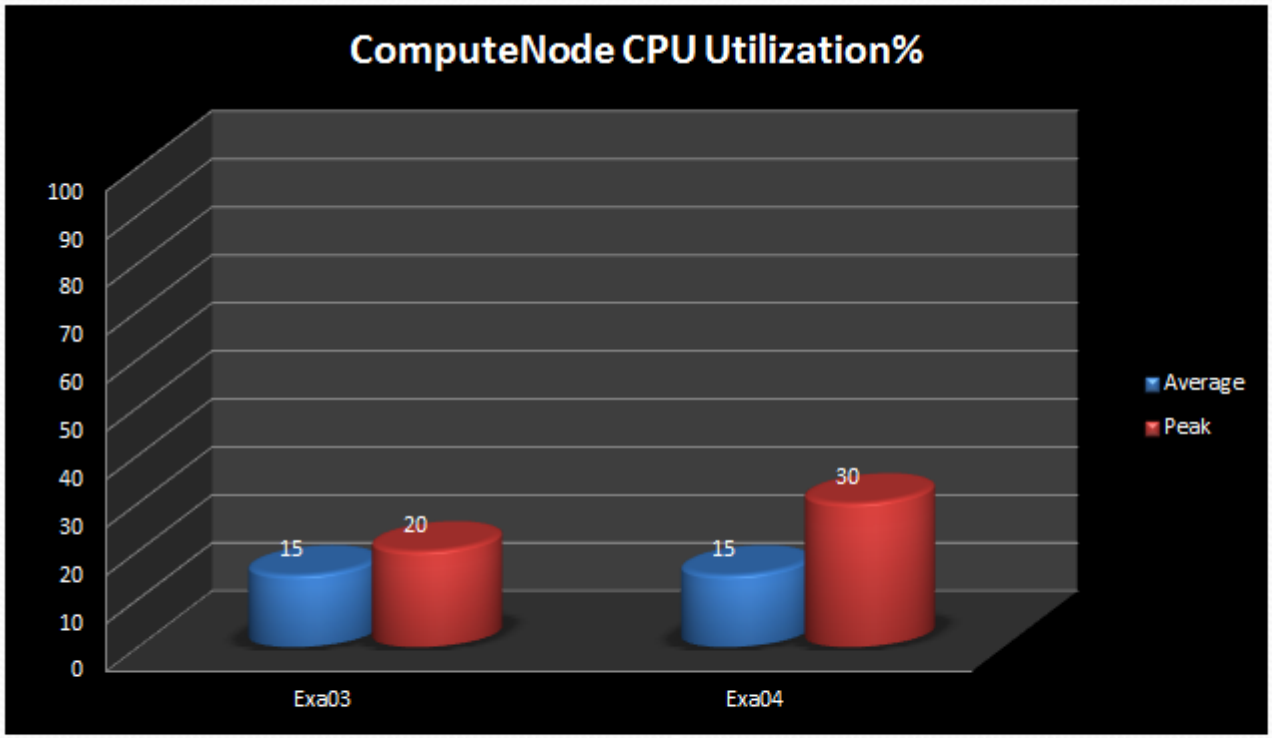
- The x2-2 half rack can perform 12.5K Hard disk MBPS Throughput
- XYZ prod database uses 1.5K HDD MBPS (~12%) and standby uses 1.4K HDD MBPS (~11%) – 98 percentile
- X4-2 Qtr HC performs 4.5K HDD

Snapshot(5)



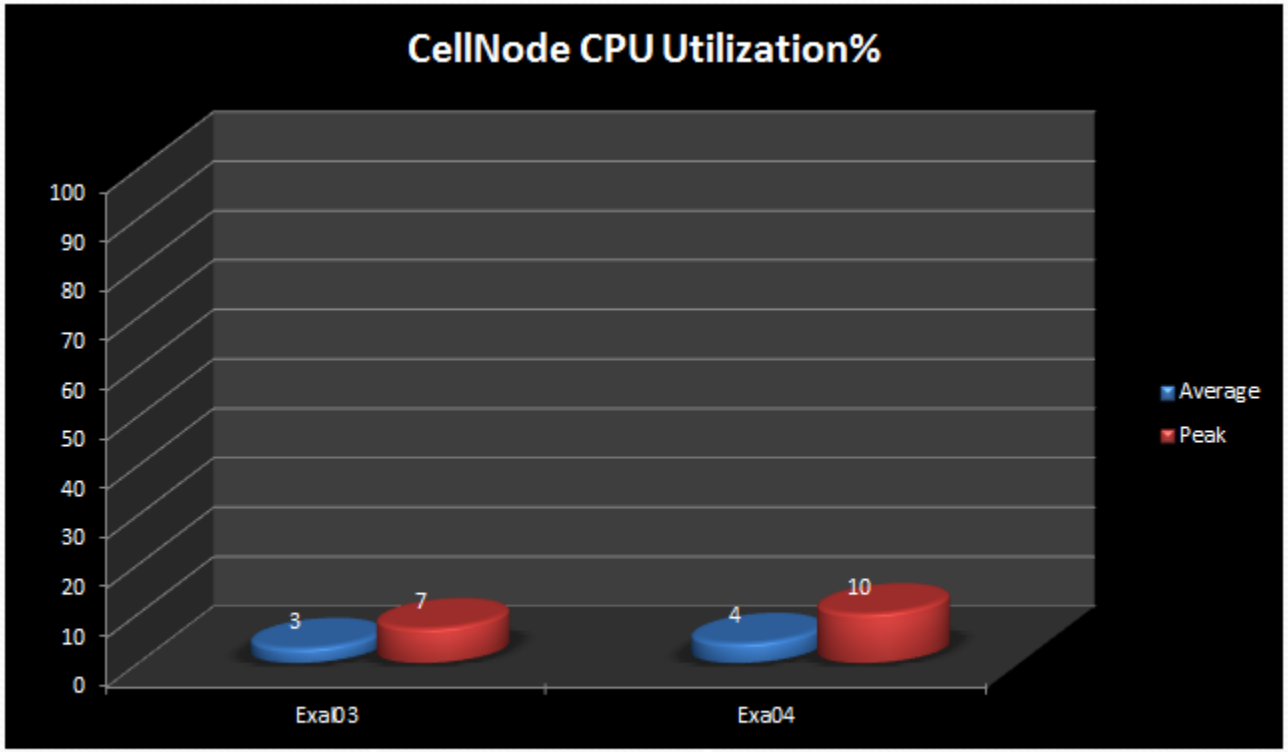
- The x2-2 half rack (High Performance disks) has 23 TB available storage. 19G is used (~83%)
- X4-2 Qtr HC has usable 63 TB available

Snapshot(6)



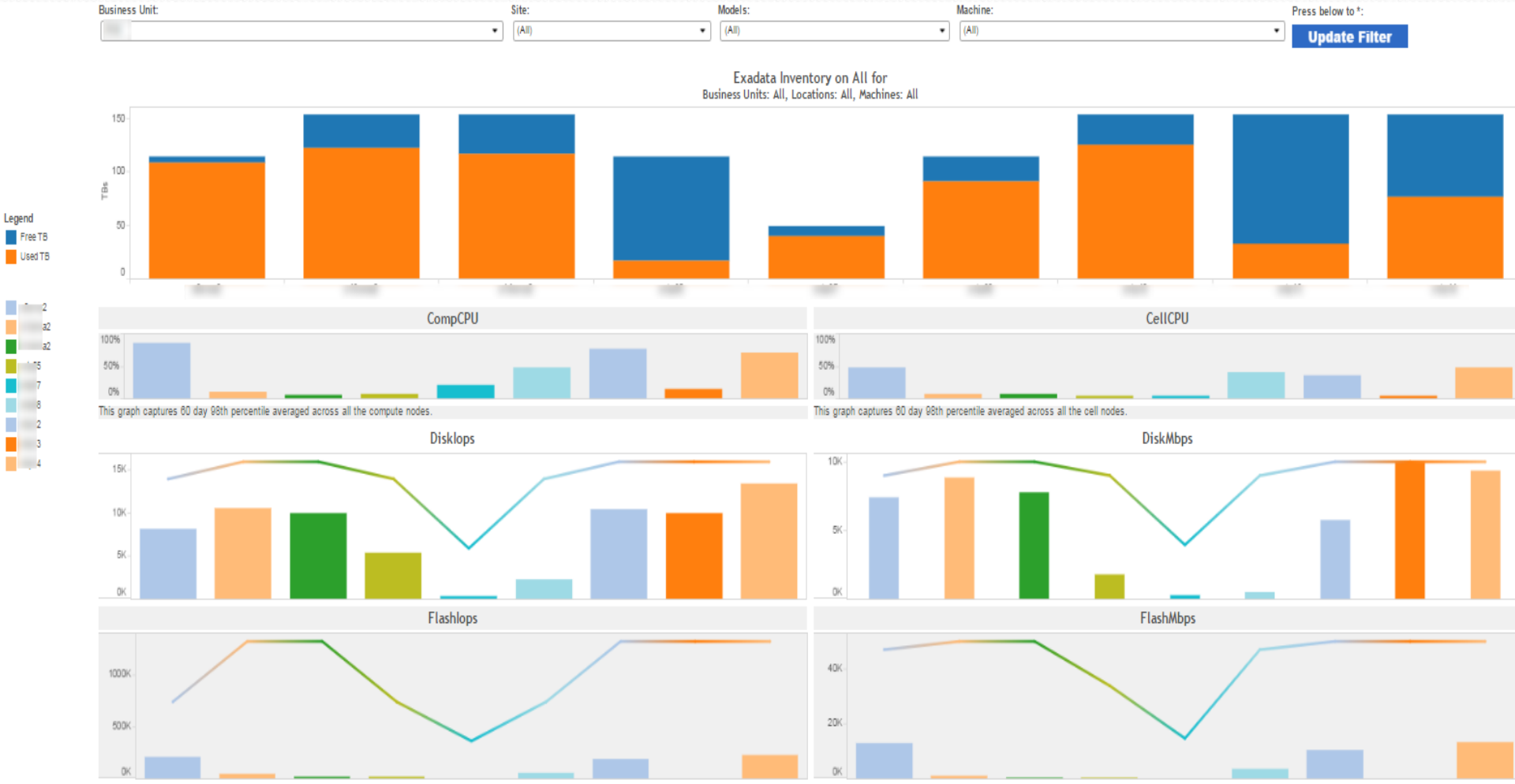
- CPU Utilization across all the 4 nodes is ~15% and at times peaks ~20%
- # of cores in x4-2 Qtr is 48 (Each compute node - 2 x Twelve-Core Intel® Xeon® E5-2697 v2 Processors -2.7 GHz)

Snapshot(7)



- CPU Utilization across all the 7 cell nodes is ~3% and at times peaks ~7%
- # of cores in x4-2 Qtr is 36

Single View across all Exadata



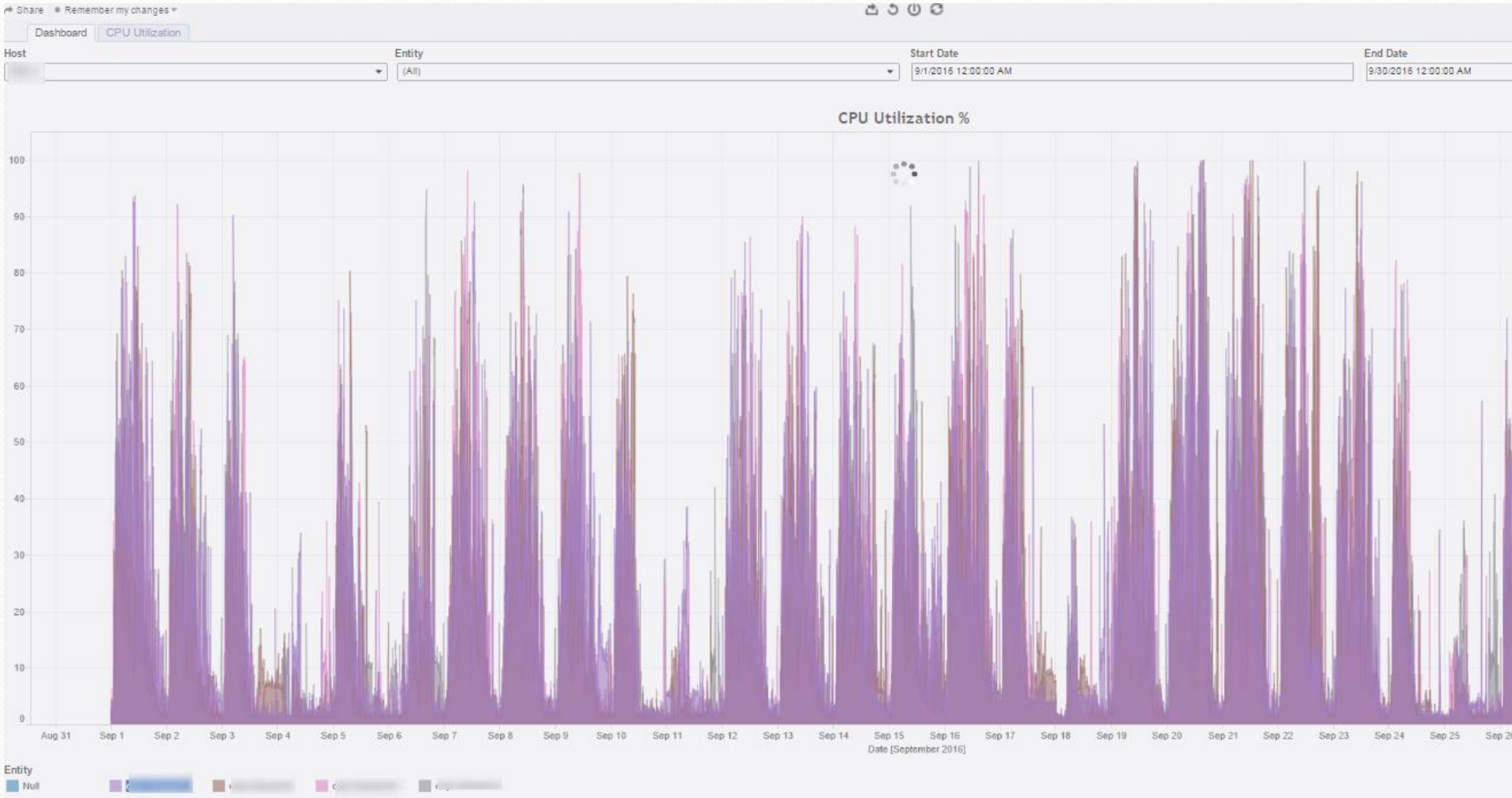
60 days 98th percentile data

Single View across all Exadata

Exadata Data Table

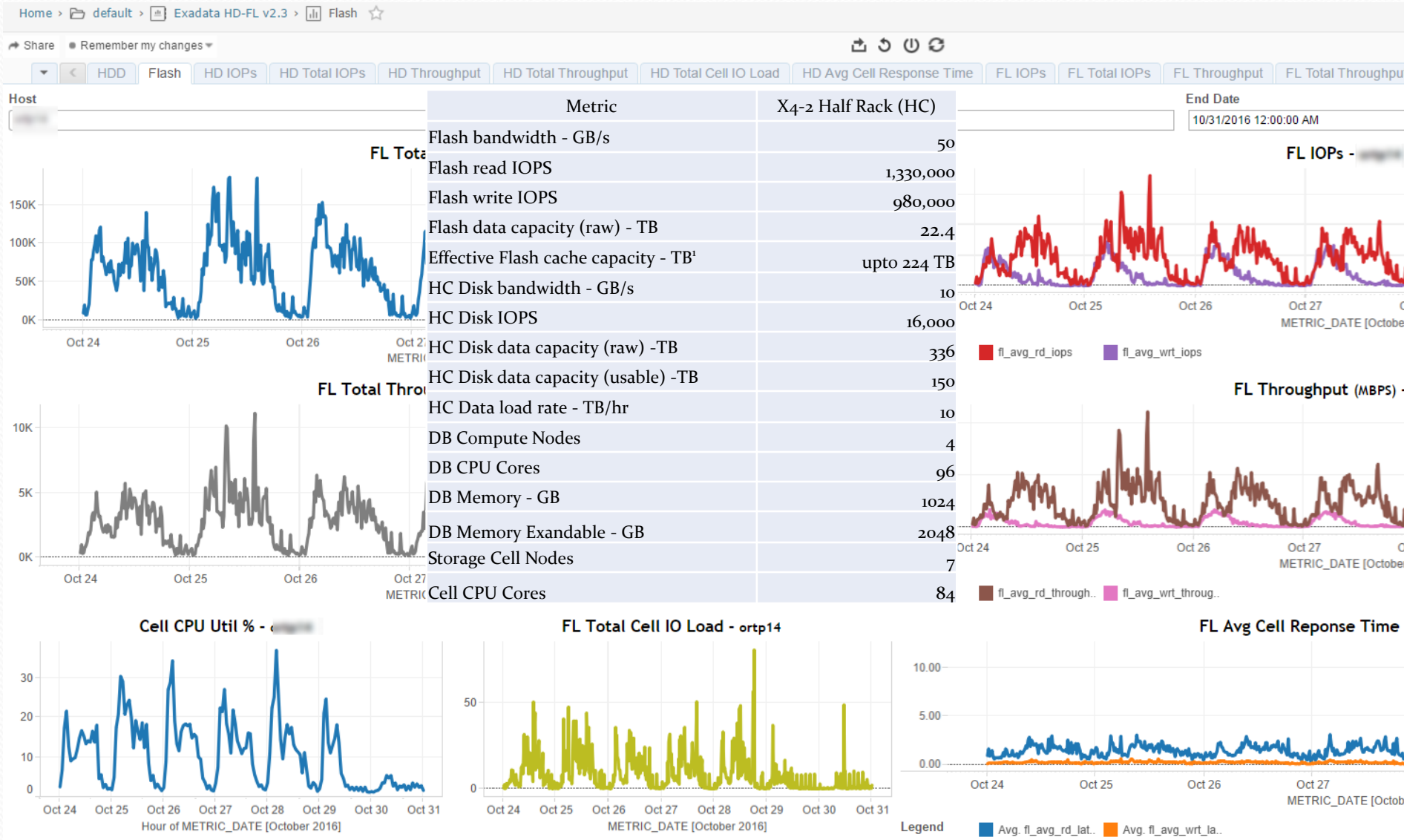
Business Unit	Machine	Location	Env Type	Rack Model	MX Replacement	Average Cell CPU 98%tile	Average CompCPU 98%tile	Disk IOPs	Disk IOPs 98%tile	Disk MBPS	Disk MBPS 98%tile	Flash IOPs	Flash IOPs 98%tile	Flash MBPS
[illegible]	[illegible]	[illegible]	DR	X3-2 HC HALF	20[illegible]-Q4	47	85	14,000	8,168	9,000	7,433	750,000	219,599	47,000
[illegible]	[illegible]	[illegible]	DR	X4-2 HC HALF	20[illegible]-Q2	8	10	16,000	10,619	10,000	8,800	1,330,000	52,866	50,000
[illegible]	[illegible]	[illegible]	DR	X4-2 HC HALF	20[illegible]-Q4	8	6	16,000	9,977	10,000	7,774	1,330,000	29,014	50,000
[illegible]	[illegible]	[illegible]	PROD	X2-2 HC HALF	20[illegible]-Q3	5	8	14,000	5,439	9,000	1,823	750,000	24,899	34,000
[illegible]	[illegible]	[illegible]	NONPROD	X2-2 HC QTR	20[illegible]-Q3	5	21	6,000	484	4,000	374	375,000	6,450	15,000
[illegible]	[illegible]	[illegible]	NONPROD	X3-2 HC HALF	20[illegible]-Q4	41	48	14,000	2,363	9,000	572	750,000	62,879	47,000
[illegible]	[illegible]	[illegible]	PROD	X4-2 HC HALF	20[illegible]-Q4	35	76	16,000	10,483	10,000	5,749	1,330,000	194,547	50,000
[illegible]	[illegible]	[illegible]	NONPROD	X4-2 HC HALF	20[illegible]-Q4	4	15	16,000	9,982	10,000	9,899	1,330,000	11,148	50,000
[illegible]	[illegible]	[illegible]	PROD	X4-2 HC HALF	20[illegible]-Q4	48	70	16,000	13,391	10,000	9,356	1,330,000	233,884	50,000
Grand Total						200	339	128,000	70,906	81,000	51,779	9,275,000	835,284	393,000

Detail level compute node CPU Utilization



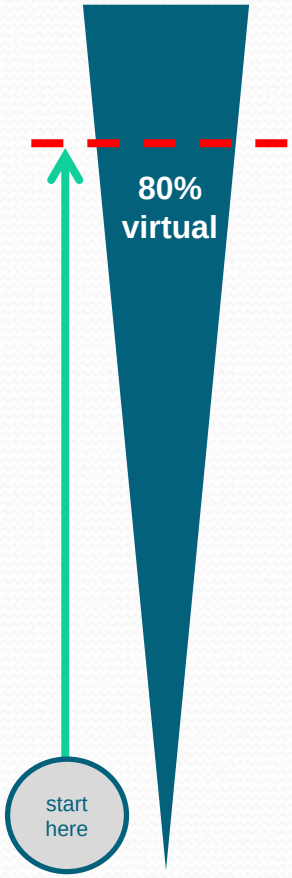
Detail level HDD data

Detail level Flash Data



Metric	X4-2 Half Rack (HC)
Flash bandwidth - GB/s	50
Flash read IOPS	1,330,000
Flash write IOPS	980,000
Flash data capacity (raw) - TB	22.4
Effective Flash cache capacity - TB[1]	upto 224 TB
HC Disk bandwidth - GB/s	10
HC Disk IOPS	16,000
HC Disk data capacity (raw) -TB	336
HC Disk data capacity (usable) -TB	150
HC Data load rate - TB/hr	10
DB Compute Nodes	4
DB CPU Cores	96
DB Memory - GB	1024
DB Memory Exandable - GB	2048
Storage Cell Nodes	7
Cell CPU Cores	84

Stack rank your databases

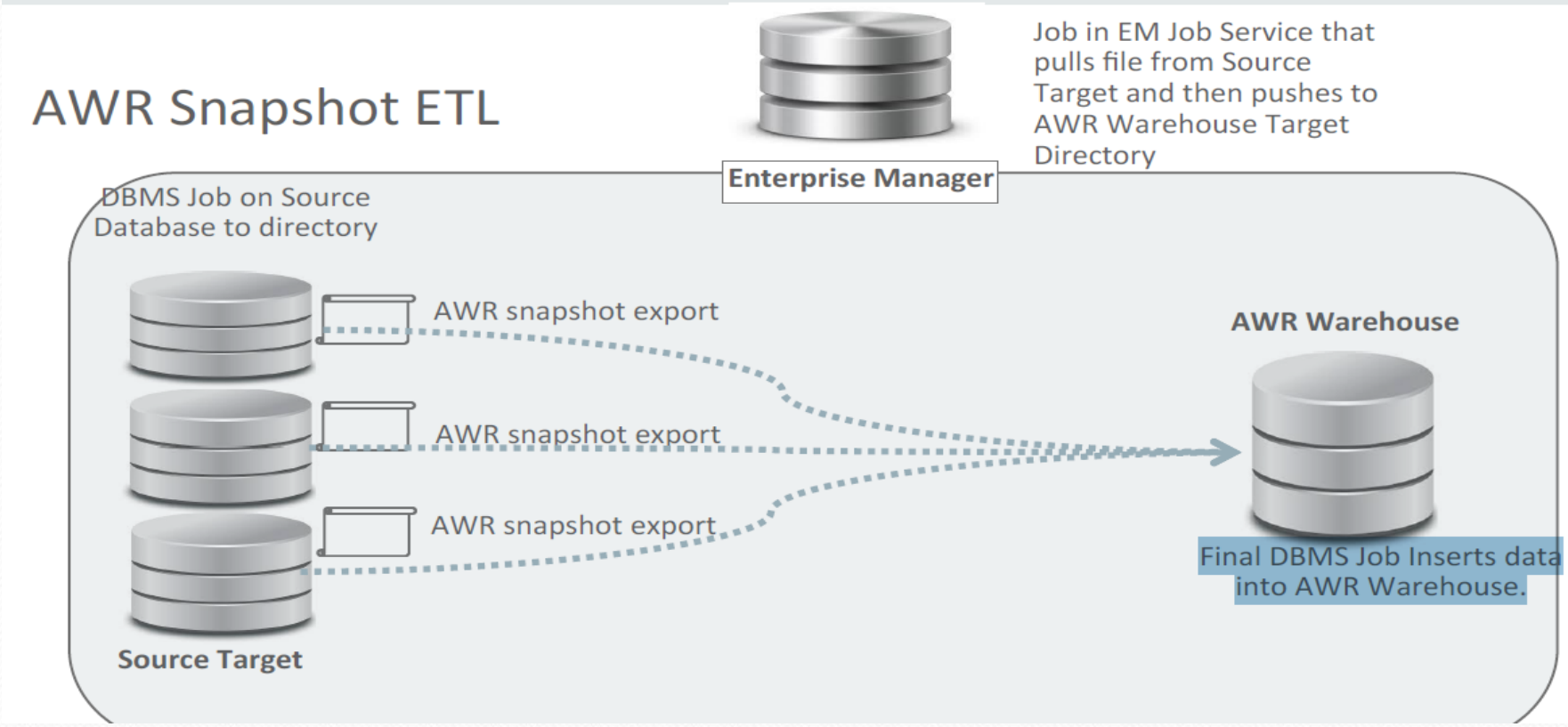


	Database	Rank	IOPS	GB/s	Latency	Cores	Storage	
High End Physical	ORAprod	1	1.5M	100	<1ms	128	100TB	
	ORAprod	2	1.4M	47	<1ms	128	75TB	
	DB2prod	3	1.3M	46	<1ms	128	75TB	
High End Virtual	ORAprod	.	.	.	.	.	.	
	DB2prod	.	.	.	.	.	.	
	SQLprod	.	.	.	.	.	.	
Virtual	ORAprod	.	.	.	.	.	.	
	DB2prod	.	.	.	.	.	.	
	SQLprod	.	.	.	.	.	.	
On-Prem As-A-Service	ORAtest	.	.	.	.	.	.	
	DB2test	.	.	.	.	.	.	
	SQLtest	.	.	.	.	.	.	
	ORAdev	.	.	.	.	.	.	
Off-Prem As-A-Service	SQLprod	.	.	.	.	.	.	
	SQLtest	.	.	.	.	.	.	
	SQLdev	3000+	100	1	>10ms	2	20GB	

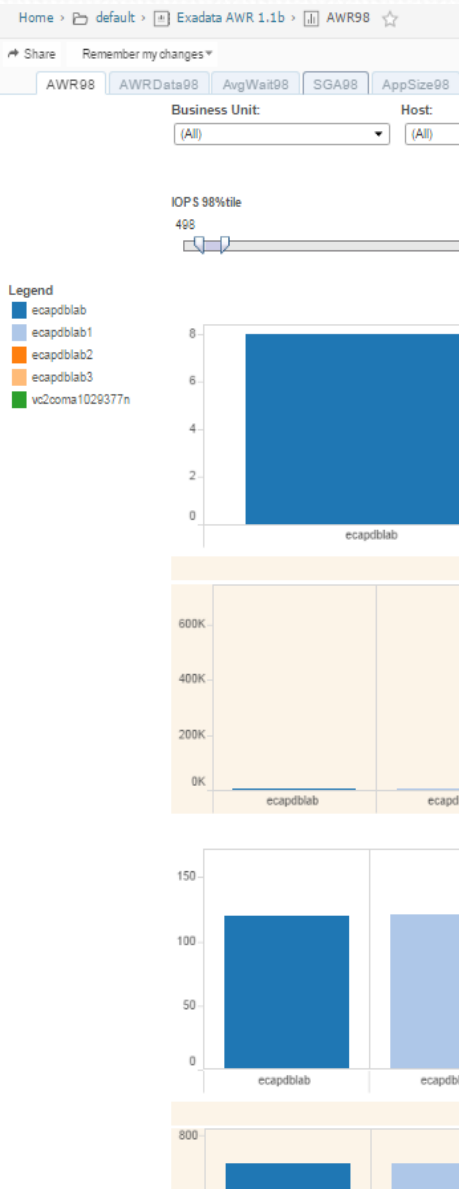
AWR Warehouse

- Centralized warehouse for long term AWR data retention
- All AWR features available on long term AWR data
- Seamless integration with OEM
- Comparative Analysis of workload, no impact on target database
- Performance Analysis of DBs on Shared Infrastructure

Data Warehouse



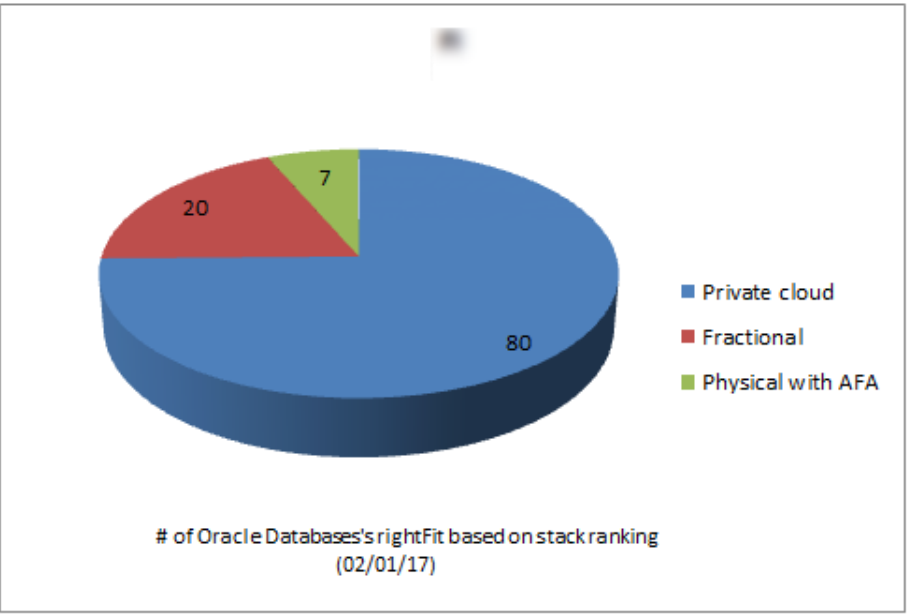
Enterprise wide 95th, 98th and 99th database level stats

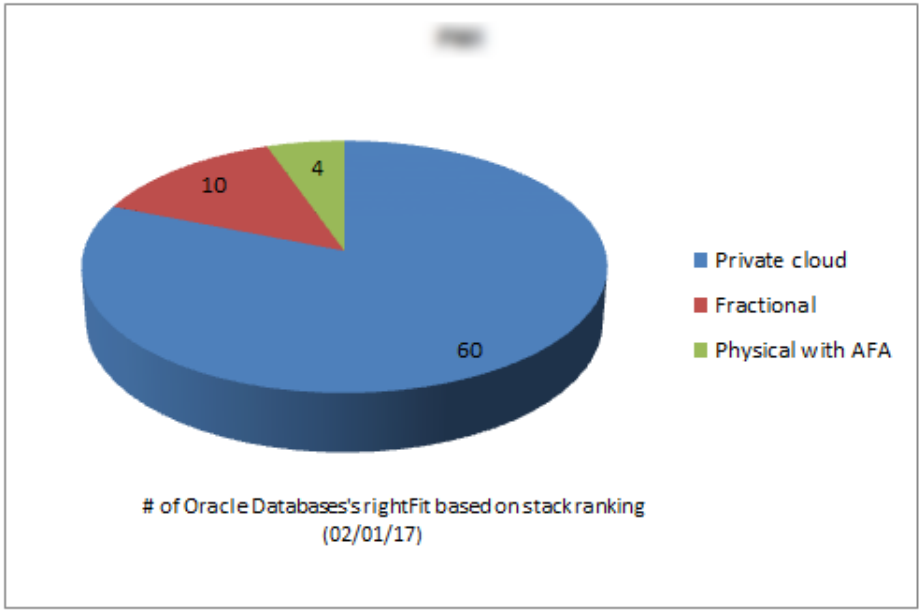



```
create or replace PROCEDURE capacity_awr_metrics
AS
startdt date;
enddt date;
count number(20);
BEGIN
select sysdate into startdt from dual;
 insert into capacity_awr_tab a
WITH PGA AS
  (SELECT target_name,
    SYSDATE collection_date,
    ROUND(percentile_cont(.95) within GROUP (
  ORDER BY PGA_MB),0) PGA_MB_95Ptile,
    ROUND(percentile_cont(.98) within GROUP (
  ORDER BY PGA_MB),0) PGA_MB_98Ptile,
    ROUND(percentile_cont(.99) within GROUP (
  ORDER BY PGA_MB),0) PGA_MB_99Ptile
  FROM
    (SELECT target_name,
      etime,
      snap_id,
      SUM(PGA_MB) PGA_MB
    FROM
      (SELECT m.target_name target_name,
        TO_CHAR(s.end_interval_time,'DD-MON-YY HH24:MI') etime ,
        h.snap_id snap_id ,
        h.instance_number inst_number ,
        ROUND(SUM(h.value)/1024/1024,0) PGA_MB
      FROM dba_hist_pgastat h,
        dba_hist_snapshot s,
        dbsnmp.caw_dbid_mapping m
      WHERE name                       ='total PGA allocated'
      AND TRUNC(s.begin_interval_time) >=TRUNC(sysdate)- 35
      AND h.dbid            = s.dbid
      AND s.dbid            =m.new_dbid
      AND h.instance_number = s.instance_number
      AND h.snap_id         = s.snap_id
      GROUP BY m.target_name,
        TO_CHAR(s.end_interval_time,'DD-MON-YY HH24:MI') ,
        h.snap_id ,
        h.instance_number
      ORDER BY h.instance_number
      )
    GROUP BY target_name,
      etime,
      snap_id
    ORDER BY etime,
```

Gets executed daily

Sr. Management level view





Offering	IOPS	Throughput(MB)	Latency (ms)	DB Size (GB)	Max # of Cores/CPUs
Public Cloud	3000	50	30+	<50	4
Private Cloud	5000	100	20+	<100	8
High end VM	?	?	?	<6000	14
Physical+AFA	Any	>5000	<2	Any	Any
Appliance					

Capacity analysis at DB level

- A big thanks to Carlos and Mauro for great help in google charts

Mauro's SQLD360 Code ☺

Host name(s):	EXAECO01, EXAECO02, EXAECO03, EXAECO04
Database name:	ORCL1
Oracle Database version:	11.2.0.4.0
Database block size:	8 KB
Database size:	57.198 TB
Datafiles:	71 (on 44 tablespaces)
Database configuration:	4-node RAC cluster
Database memory:	SGA 64.0 GB, PGA 144.0 GB, ASMM
Hardware:	Engineered System with 7 storage servers
Processor:	
Physical CPUs:	96 cores, on 4-node RAC cluster
Oracle CPUs:	192 CPUs (threads), on 4-node RAC cluster
Physical RAM:	1,008.0 GB, on 4-node RAC cluster
Operating system:	Linux x86 64-bit

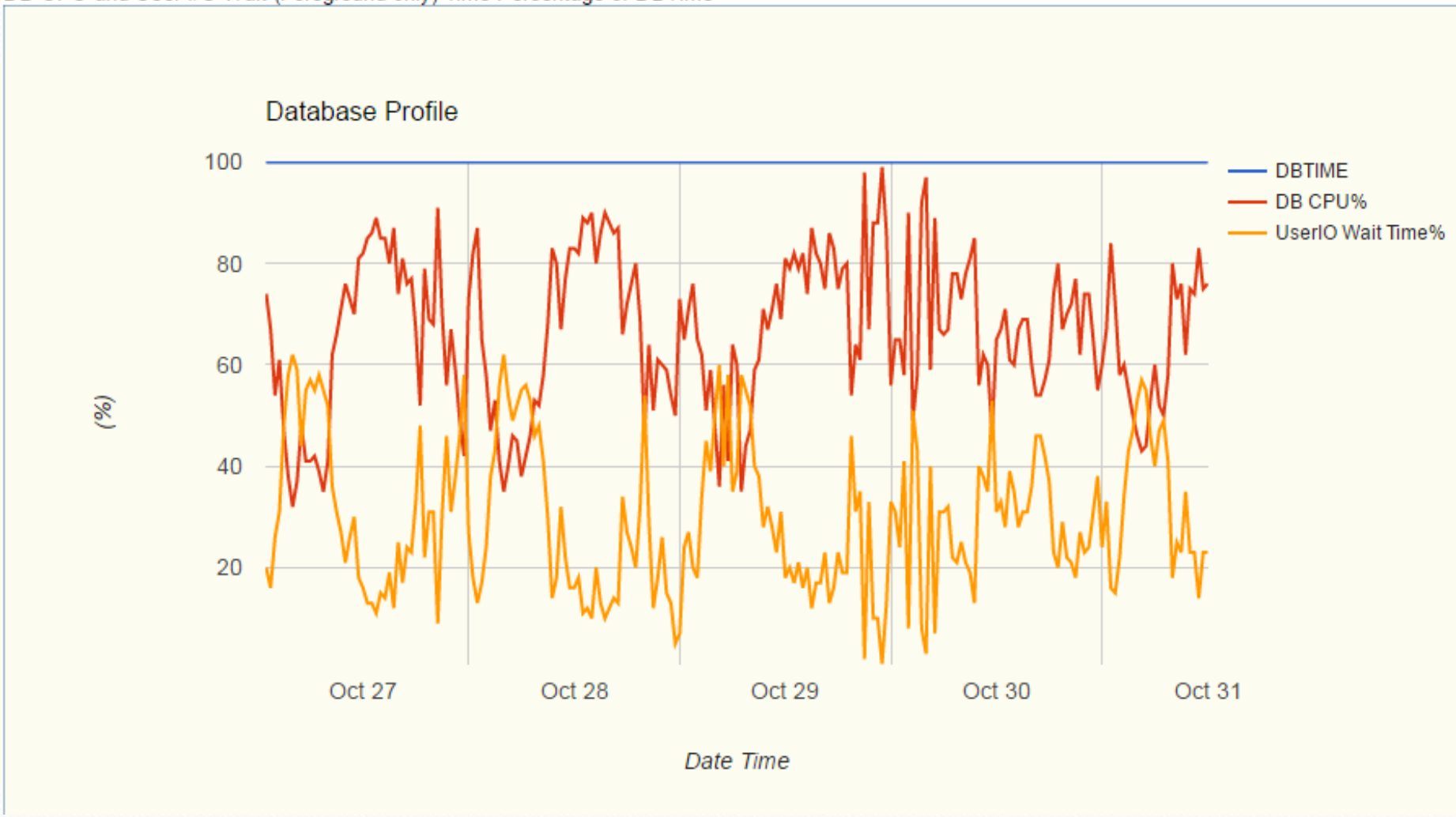
Tanel's script

Show Exadata cell versions from V$CELL_CONFIG....

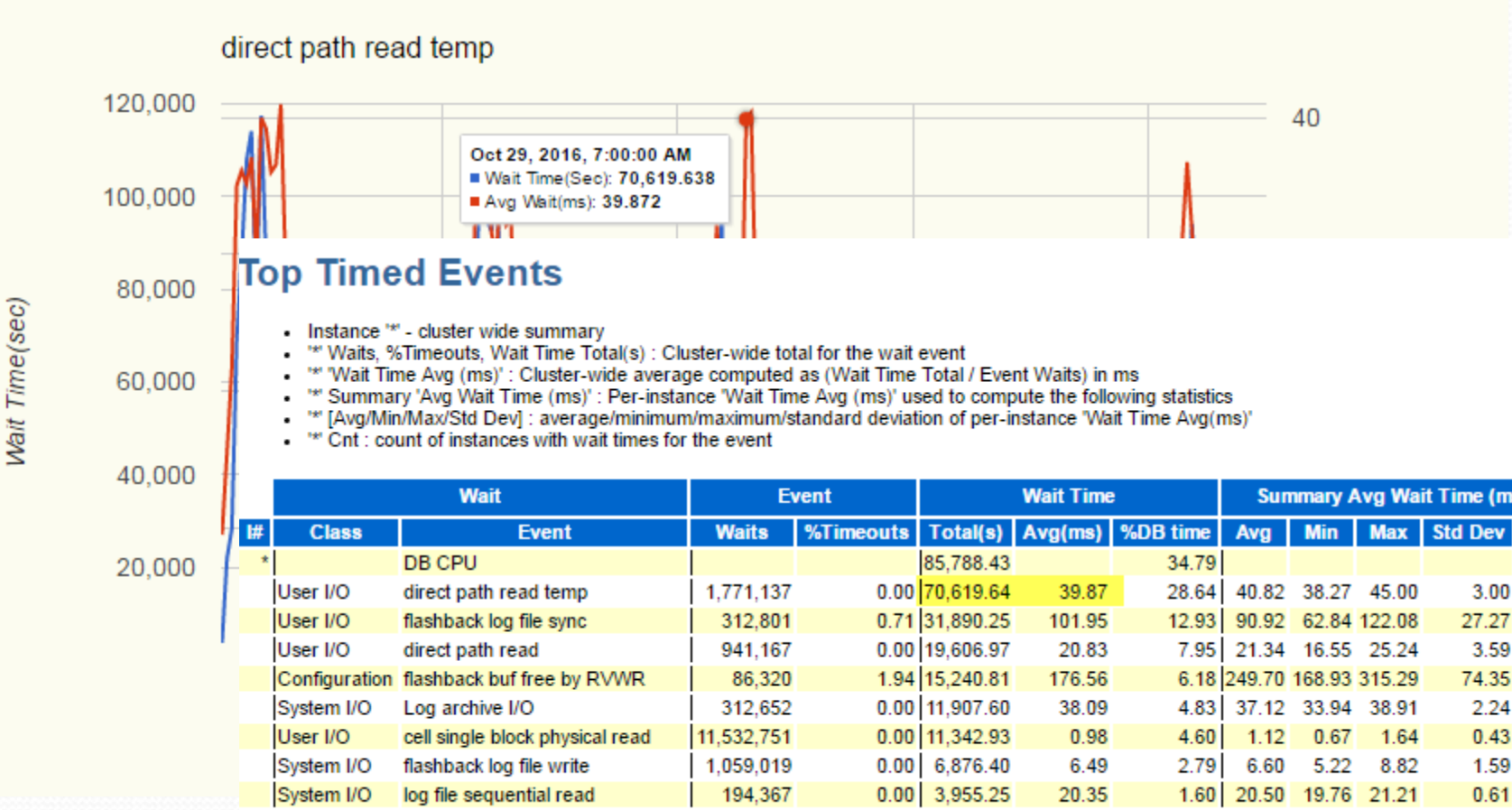
CELL_NAME	CELL_PATH	CELLSRV_VERSION	FLASH_CACHE_MODE	CPU_COUNT	UPTIME	KERNEL_VER	MAKE_MODEL
EXAECOceladm01	192.168.14.9;192.168.14.10	12.1.1.1.1	WriteBack	24	274 days, 8:18	2.6.39-400.128.17.el5uek	Oracle Corporation SUN SERVER X4-2L
EXAECOceladm02	192.168.14.11;192.168.14.12	12.1.1.1.1	WriteBack	24	274 days, 8:34	2.6.39-400.128.17.el5uek	Oracle Corporation SUN SERVER X4-2L
EXAECOceladm03	192.168.14.13;192.168.14.14	12.1.1.1.1	WriteBack	24	274 days, 8:33	2.6.39-400.128.17.el5uek	Oracle Corporation SUN SERVER X4-2L
EXAECOceladm04	192.168.14.15;192.168.14.16	12.1.1.1.1	WriteBack	24	274 days, 8:27	2.6.39-400.128.17.el5uek	Oracle Corporation SUN SERVER X4-2L
EXAECOceladm05	192.168.14.17;192.168.14.18	12.1.1.1.1	WriteBack	24	274 days, 7:51	2.6.39-400.128.17.el5uek	Oracle Corporation SUN SERVER X4-2L
EXAECOceladm06	192.168.14.19;192.168.14.20	12.1.1.1.1	WriteBack	24	274 days, 8:13	2.6.39-400.128.17.el5uek	Oracle Corporation SUN SERVER X4-2L
EXAECOceladm07	192.168.14.21;192.168.14.22	12.1.1.1.1	WriteBack	24	274 days, 7:42	2.6.39-400.128.17.el5uek	Oracle Corporation SUN SERVER X4-2L

Database Profile

DB CPU and User I/O Wait (Foreground only) Time Percentage of DBTime



Wait event – Direct path read temp

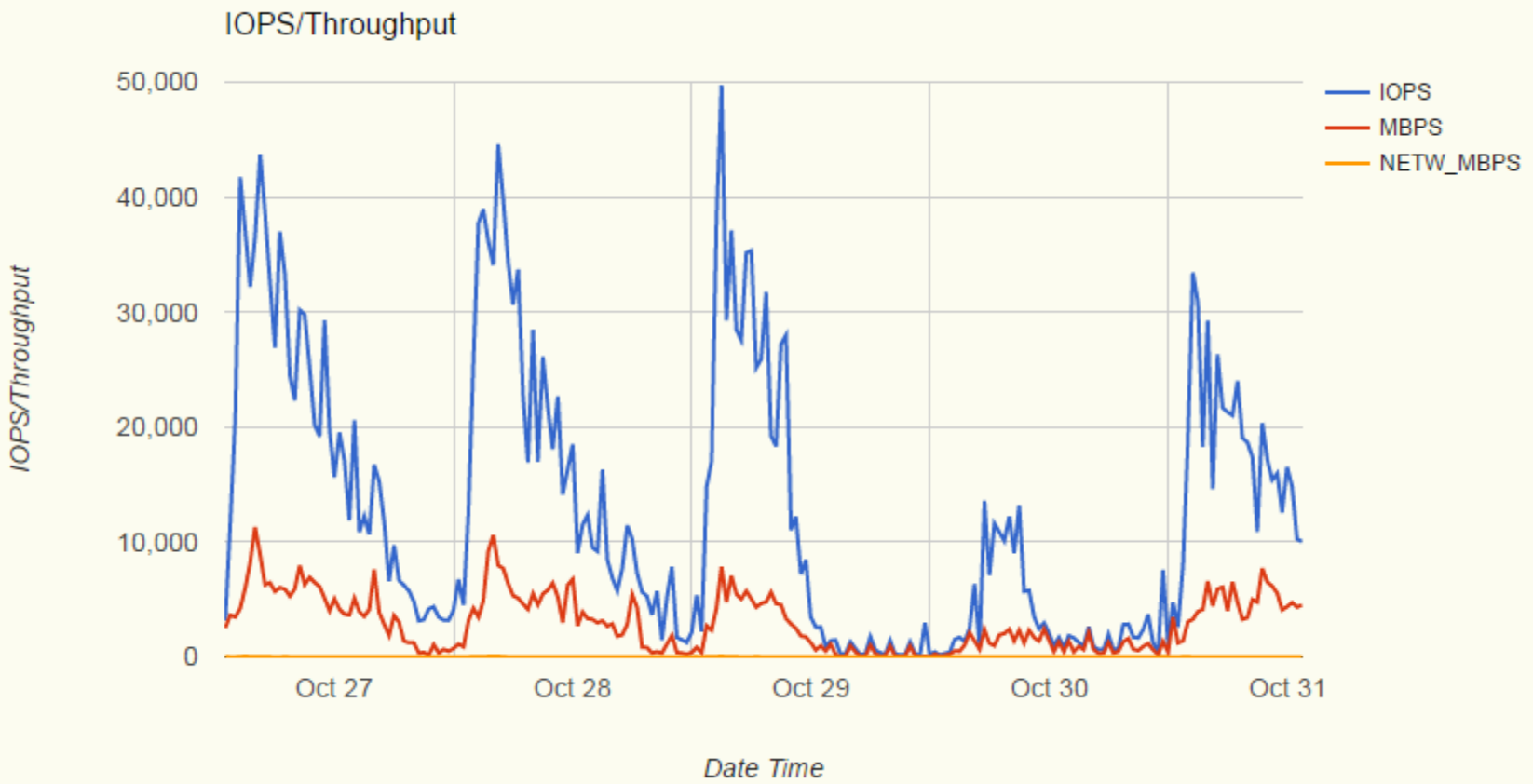


Top Timed Events

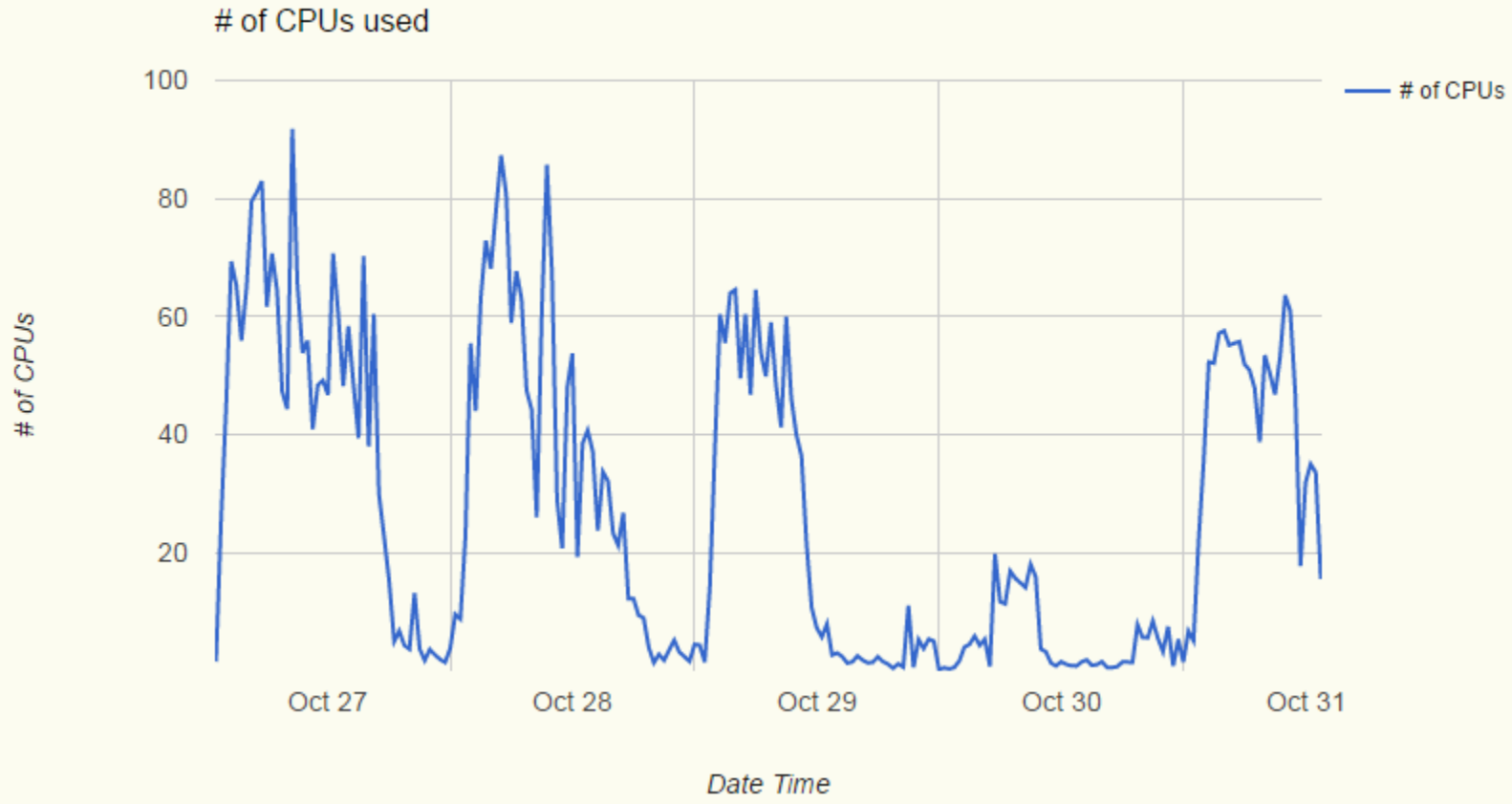
- Instance '*' - cluster wide summary
- '*' Waits, %Timeouts, Wait Time Total(s) : Cluster-wide total for the wait event
- '*' 'Wait Time Avg (ms)' : Cluster-wide average computed as (Wait Time Total / Event Waits) in ms
- '*' Summary 'Avg Wait Time (ms)' : Per-instance 'Wait Time Avg (ms)' used to compute the following statistics
- '*' [Avg/Min/Max/Std Dev] : average/minimum/maximum/standard deviation of per-instance 'Wait Time Avg(ms)'
- '*' Cnt : count of instances with wait times for the event

	Wait		Event		Wait Time			Summary Avg Wait Time (ms)				
I#	Class	Event	Waits	%Timeouts	Total(s)	Avg(ms)	%DB time	Avg	Min	Max	Std Dev	Cnt
*		DB CPU			85,788.43		34.79					4
	User I/O	direct path read temp	1,771,137	0.00	70,619.64	39.87	28.64	40.82	38.27	45.00	3.00	4
	User I/O	flashback log file sync	312,801	0.71	31,890.25	101.95	12.93	90.92	62.84	122.08	27.27	4
	User I/O	direct path read	941,167	0.00	19,606.97	20.83	7.95	21.34	16.55	25.24	3.59	4
	Configuration	flashback buf free by RVWR	86,320	1.94	15,240.81	176.56	6.18	249.70	168.93	315.29	74.35	4
	System I/O	Log archive I/O	312,652	0.00	11,907.60	38.09	4.83	37.12	33.94	38.91	2.24	4
	User I/O	cell single block physical read	11,532,751	0.00	11,342.93	0.98	4.60	1.12	0.67	1.64	0.43	4
	System I/O	flashback log file write	1,059,019	0.00	6,876.40	6.49	2.79	6.60	5.22	8.82	1.59	4
	System I/O	log file sequential read	194,367	0.00	3,955.25	20.35	1.60	20.50	19.76	21.21	0.61	4
	User I/O	direct path write temp	170,547	0.00	3,256.19	19.09	1.32	22.72	15.03	34.16	8.17	4

IOPS/Throughput

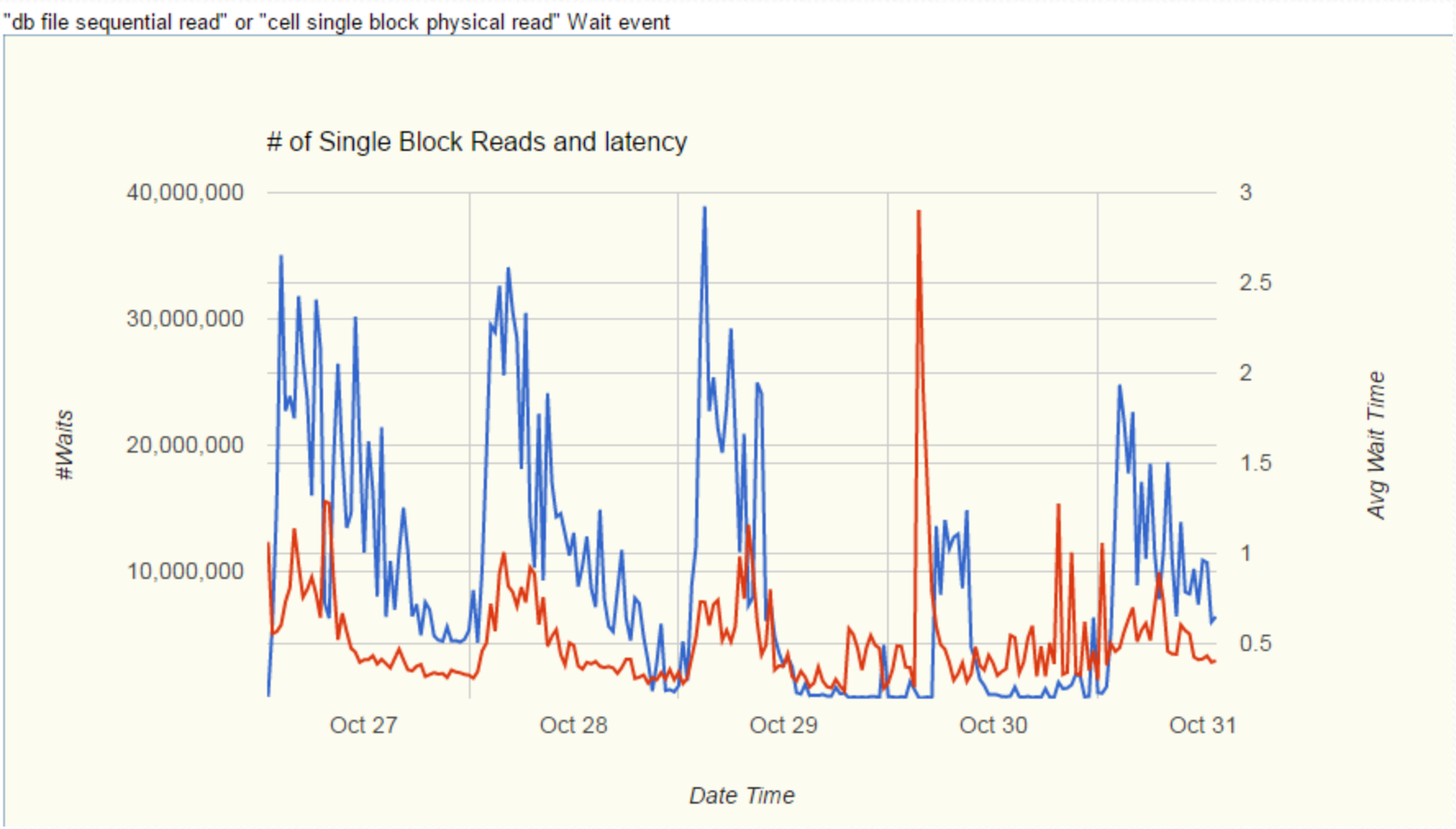


of CPU/core used

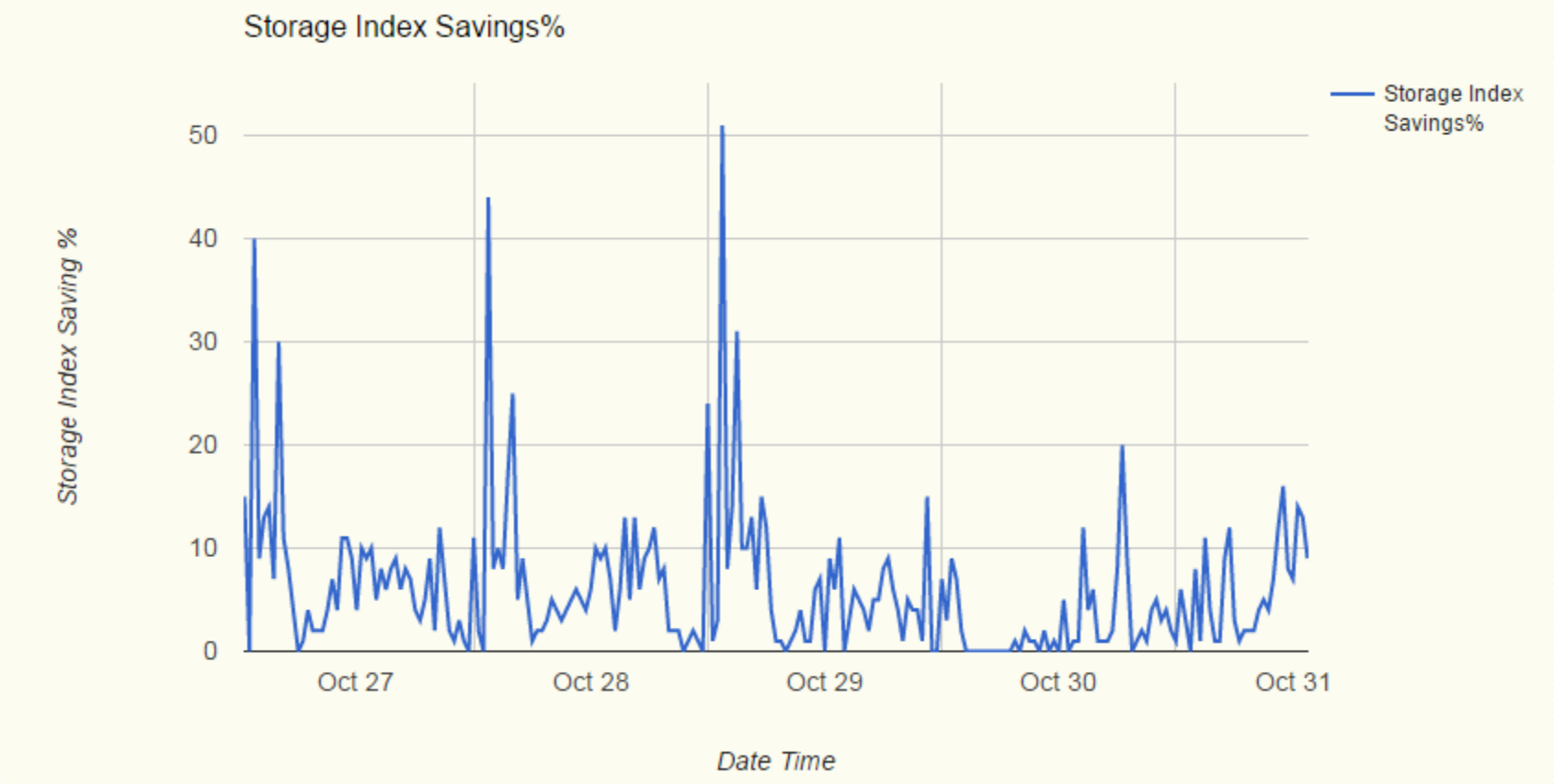


Single block read latency

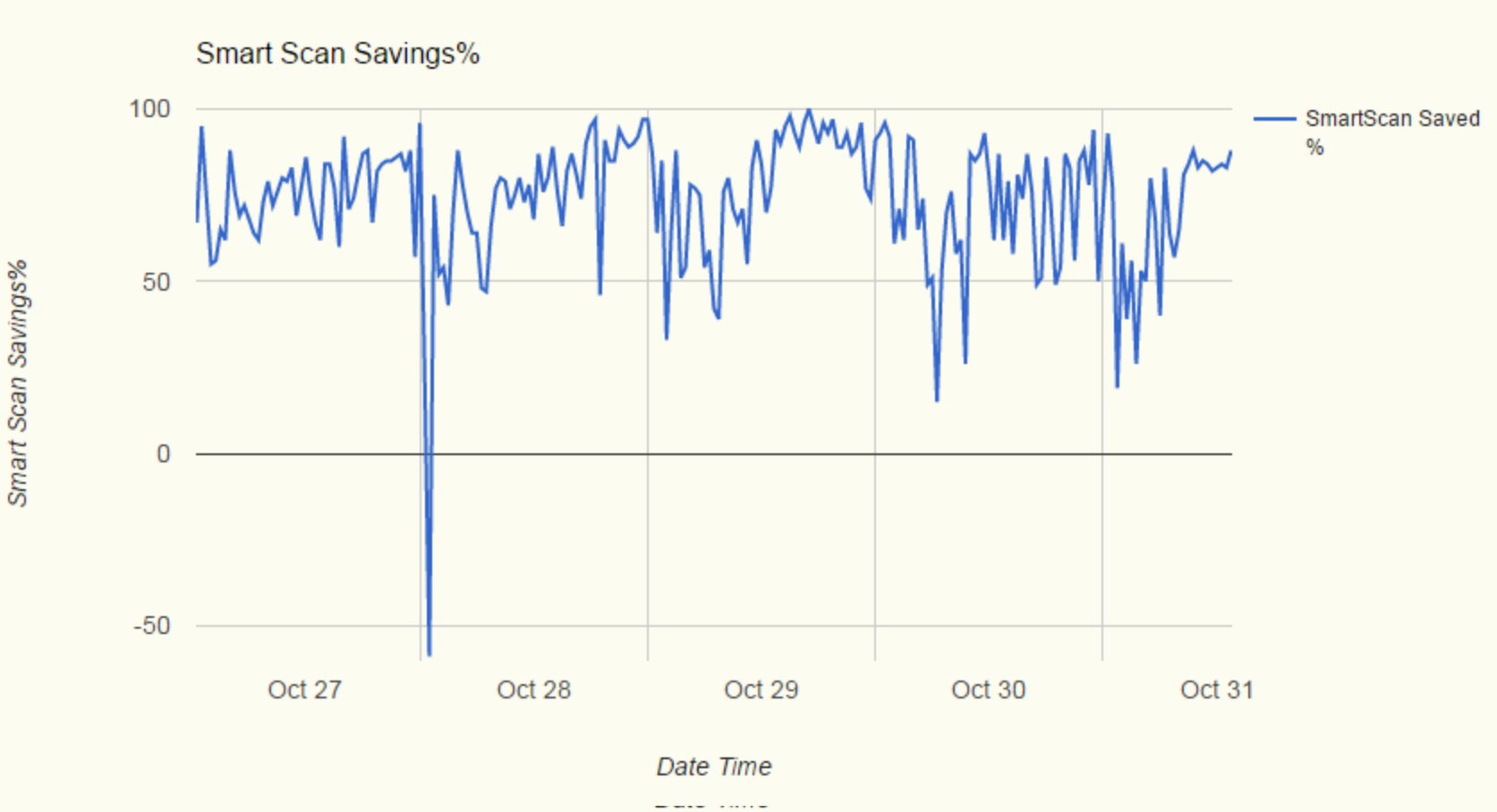
"db file sequential read" or "cell single block physical read" Wait event



Storage Index Savings



Smart Scan Savings



Next Session

FEBRUARY 2, 2017	12:00 pm-12:30 pm	**Capacity Planning at Scale [838821]**
FEBRUARY 2, 2017	2:30 pm-3:20 pm	**Trace File Analyzer – Faster Time To Resolution Through Intelligent First Failure Diagnostics [834300]** Room 102

OEM Plugin Reference manual

8.1 Oracle Exadata Storage Server

The Oracle Exadata target monitors the software and hardware performance of an individual Oracle Exadata Storage Server in the database.

8.1.1 Aggregated Exadata Capacity

This metric category contains the aggregated metrics of the Exadata Capacity metric category and it collects every 60 minutes.

8.1.1.1 Disk Size (GB)

This metric gives an indication of the size of the status in GB.

Target Version	Collection Frequency
All Versions	Every 60 Minutes

8.1.2.2 Average CellDisk Read IOPS

This metric gives an indication of the average number of read input/output operations per second.

Target Version	Collection Frequency
All Versions	Every 15 Minutes

8.1.2.3 Average CellDisk Read Response Time

This metric gives an indication of the average read response time to the cell disk.

Target Version	Collection Frequency
All Versions	Every 15 Minutes

Appendix - A

Db_capacity.sql will generate a .html file.

DEF conf_days=7	Report will run SYSDATE-conf_days, running it for longer period may take some time.
DEF conf_day_ptile=7	This is to call capacity.sql for 95th, 98th and 99th Ptile calculation

References:

- https://carlos-sierra.net/
- https://mauro-pagano.com/
- http://dbakevlar.com/category/awr-warehouse/
- Thinking clearly about performance
  - http://method-r.com/papers/file/44-thinking-clearly-about-performance
- Maintaining and Troubleshooting the Management Repository
  - http://docs.oracle.com/cd/E24628_01/doc.121/e24473/repository.htm#EMADM12672
- Enterprise Manager Oracle Database Plug-in Metric Reference Manual
  - https://docs.oracle.com/cd/E24628_01/em.121/e25160/oracle_exadata.htm#CIACDFFI
- Enterprise Manager Framework, Host, and Services Metric Reference Manual- Host
  - https://docs.oracle.com/cd/B19306_01/em.102/b16230/host.htm
- Oracle® Enterprise Manager Oracle Database and Database-Related Metric Reference Manual
- http://www.slideshare.net/MaazAnjum/maaz-anjum-ioug-em12c-capacity-planning-with-oem-metrics
- 12c Cloud Control Repository: How to Modify the Default Retention and Purging Policies for Metric Data? (Doc ID 1405036.1)