Nordea Exadata CC Platform

Version 1.4

Making it possible

Nordea Exadata CC. Nordea Inventory

Nordea has used ExaCC for a decade. Exadata has for sure been a very stable and powerful platform and has more or less taken over from the old Solaris platform in Nordea. However, the platform is aging and are bound for replacement.

Nordea has chosen to minimize the footprint and move workload to lesser expensive hardware. Exadata Cloud at Customer is a replacement for the aging Exadata platform if customers has a requirement for that.

For now, 4 major application groups has decided to move to ExaCC.

The ExaCC environment in Nordea consist of:

13 Exadata Infrastructures (Racks)
40 Virtual Clusters
100 Virtual database servers.

A platform this size requires a lot of automation. Both during database creation and maintenance but for sure also during resource allocation.

Nordea Exadata CC. Security Model

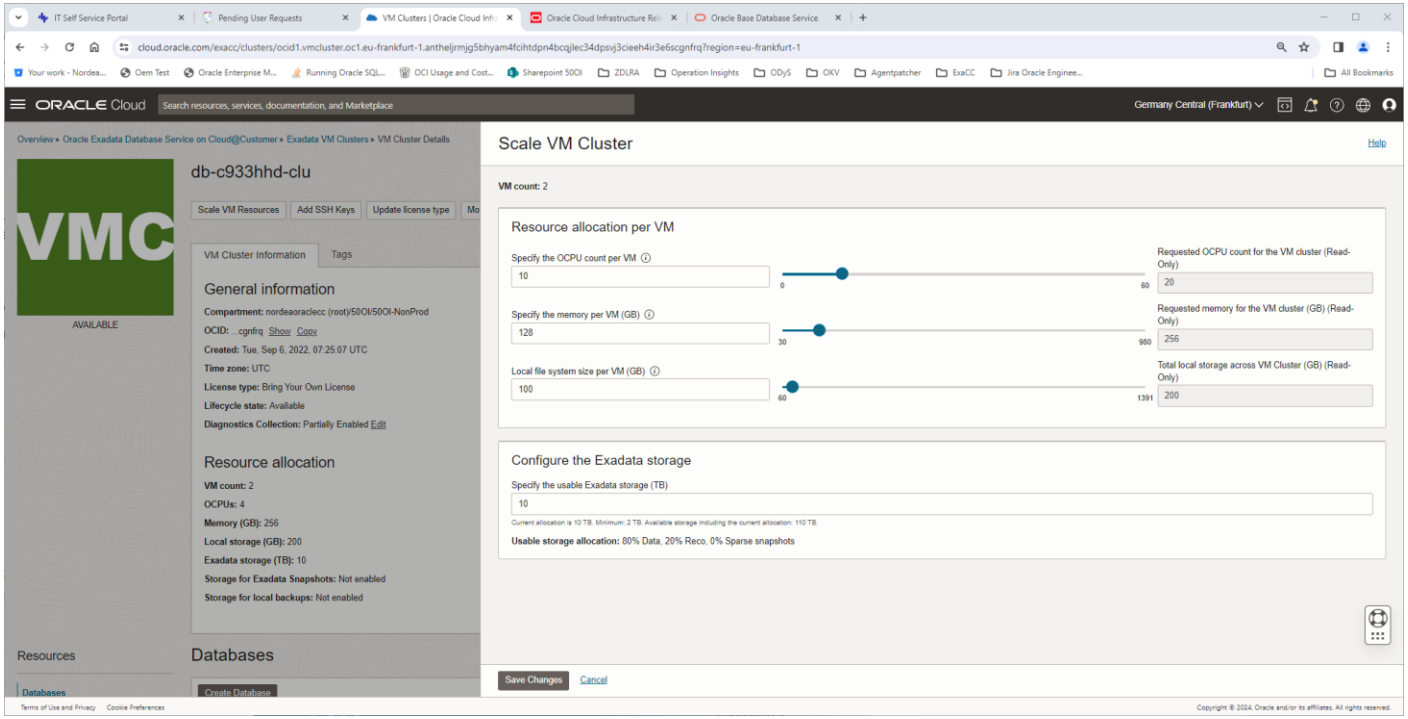
The ExaCC has various ways to connect to the Oracle Cloud. Basically, this can be done using the OCI Cloud console or use a Rest API to connect to the cloud. The virtual machines has tools of their own to connect directly to the Oracle Cloud.

To setup the Rest API You need to:

1. Create a user in the OCI Cloud Console and generate a Pem key for that user.
2. Create a config file include the user OCID, tenancy OCID, fingerprint on the public key.
3. Upload the public key from the key pair
4. Download the Pem key from the console and put it on an application server.

Also, You do need to open the firewall (opening Proxy servers). Installation of ExaCC requires around 15 openings and scaling from an application server requires additionally openings.

Nordea Exadata CC. Scaling Resources using OCI console

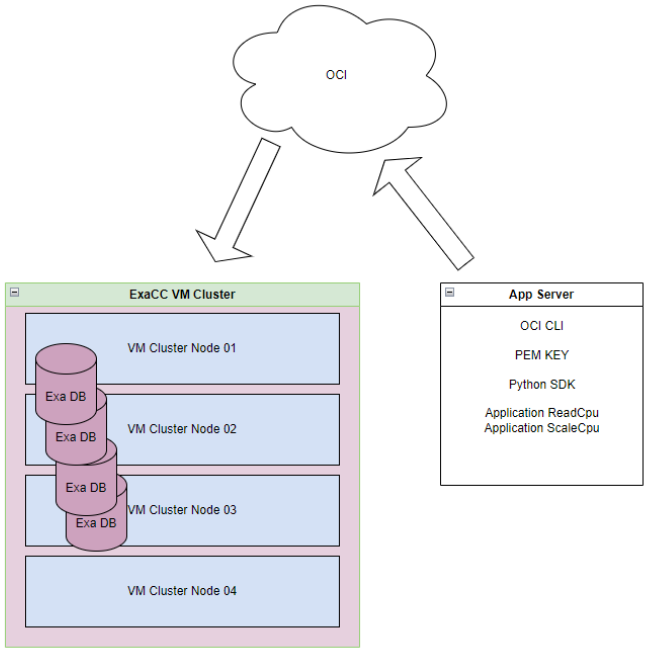


Possibility to scale OCPU, Memory, Local File System and ASM storage.

Solution is adequate for on-demand control of resources. Requires administrator privileges for the user on the Compartment level.

Memory and Local File System will require that the virtual nodes will be booted in a round-robin fashion. Also any scaling operation will take time. The OCPU will typically take 10 minutes. Storage and Memory adjustment can take much more.

Nordea Exadata CC. Scaling OCPU using Rest API



Solution developed using Python SDK and OCI. Can be used by a scheduler to scale up and down using fixed values.

Solution is adequate for handling a few clusters. Communicates with the OCI Cloud using oci-cli. oci-cli communicates with the cloud using Rest API. PEM keys must be created using the OCI console and uploaded to the application server. No direct communication between the App server and the virtual nodes. Solution can (for now) only handle OCPU.

Nordea Exadata CC. Scaling OCPU using Rest API (Continued)

```
$ ./read_cpu.sh db-c933hhd-clu

Scale VM

2024-01-04 14:17:45,974  ReadOcpu. User: 50oisdkusr.#######################################################
2024-01-04 14:17:45,975  ReadOcpu. User: 50oisdkusr. ReadOCPU Started
2024-01-04 14:17:46,756  ReadOcpu. User: 50oisdkusr. Virtual Cluster:              db-c933hhd-clu
2024-01-04 14:17:46,756  ReadOcpu. User: 50oisdkusr. Number of virtual machines:    2
2024-01-04 14:17:46,756  ReadOcpu. User: 50oisdkusr. Actual number of OCPU(Total):  4
2024-01-04 14:17:46,756  ReadOcpu. User: 50oisdkusr. Actual number of OCPU(pr. vm): 2
```

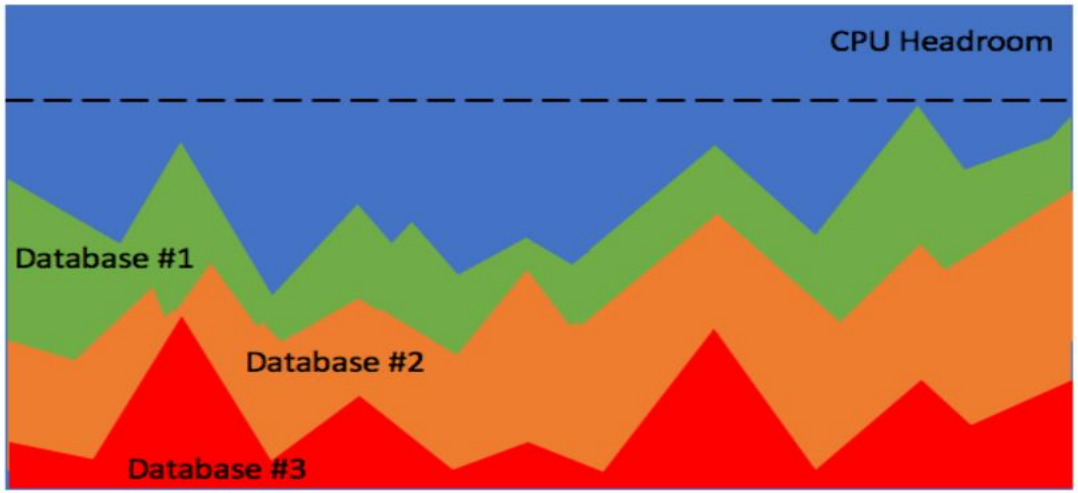
```
$ ./scale_cpu.sh db-c933hhd-clu 6

Scale VM

2024-01-04 14:24:37,619  ScaleOcpu. User: 50oisdkusr. #######################################################
2024-01-04 14:24:37,619  ScaleOcpu. User: 50oisdkusr. Started
2024-01-04 14:24:37,619  ScaleOcpu. User: 50oisdkusr. Virtual Cluster:           db-c933hhd-clu
2024-01-04 14:24:37,619  ScaleOcpu. User: 50oisdkusr. Requested OCPU(Total):      6
2024-01-04 14:24:38,072  ScaleOcpu. User: 50oisdkusr. Actual number of OCPU:      4
2024-01-04 14:24:38,409  ScaleOcpu. User: 50oisdkusr. Number of virtual machines: 2
2024-01-04 14:24:38,410  ScaleOcpu. User: 50oisdkusr. Scaling cluster db-c933hhd-clu up.
New number of OCPU: 6. Please wait....

2024-01-04 14:31:15,683  ScaleOcpu. User: 50oisdkusr. Cluster db-c933hhd-clu has been updated.
2024-01-04 14:31:15,836  ScaleOcpu. User: 50oisdkusr. New number of CPU:       6
```

Nordea Exadata CC. Dynamic Scaling

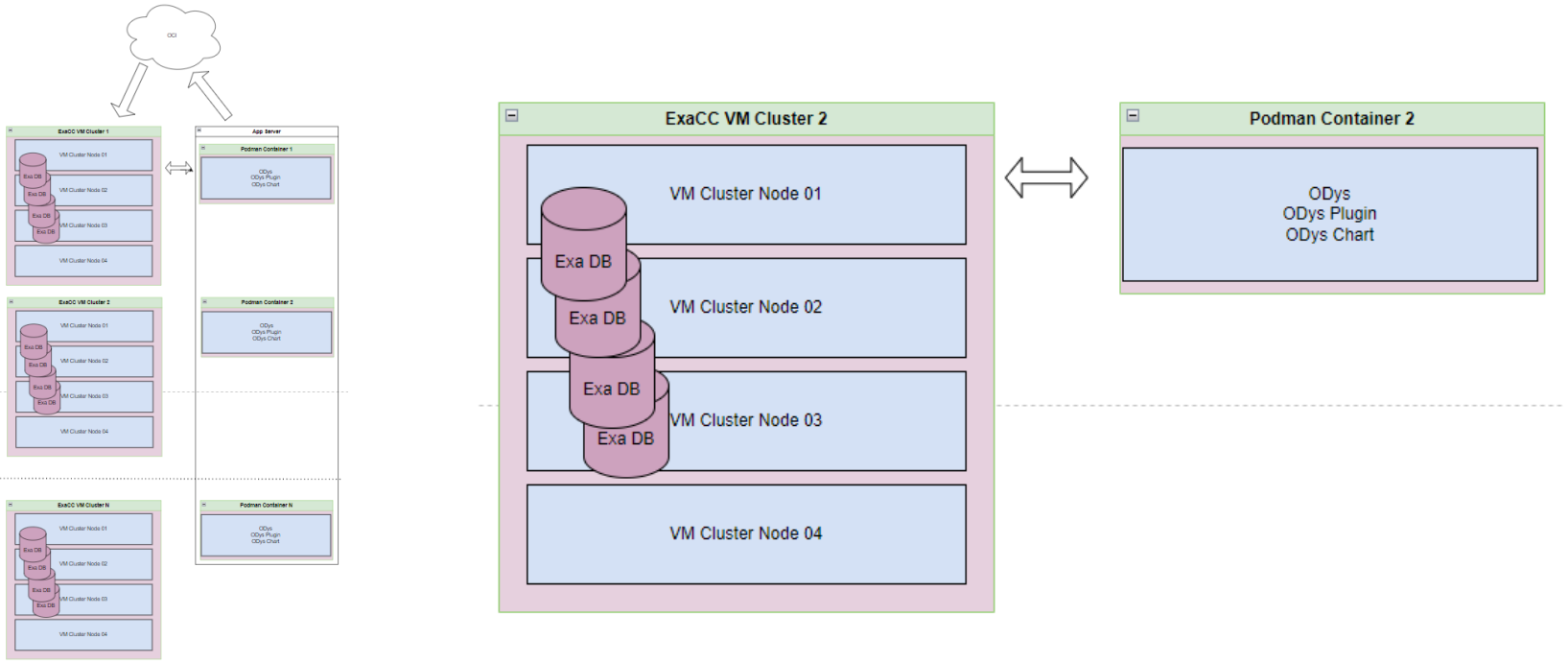
One of the strong benefits of Oracle Clout@Cstomer is '*Dynamic Scaling',* where a considerable part of the invoice is based on the actual number of *OCPUs* configured for all clusters on a Cloud@Customer machine. It is possible manually to adjust the OCPUs configures for any cluster to have effect online without having to restart the cluster or any database running in the cluster.

With the introduction of *Dynamic Scaling* the automatic online adjustment of OCPUs for any cluster is based on a configuration and with an outgoing message to the appropriate oracle datacenter.

Dynamic Scaling is Documented here:
(ODyS) Oracle Dynamic Scaling Suite Main Index Page (Doc ID 2774779.1)



Nordea Exadata CC. Scaling OCPU using Podman Containers



Podman wrapper delivered by Oracle. Configuration is using a text file and are reasonable easy to automate. One Podman container handles one virtual cluster.

ODyS plugin connects to the cluster to measure the Load. Solution is adequate for handling a large number of clusters. Solution communicates with the OCI Cloud using oci-cli using pem keys, but that complexity are controlled in the Podman container. Solution can (for now) only handle OCPU.

Documentation: (ODyS) Oracle Dynamic Scaling on Podman (Doc ID 2828699.1)

Nordea Exadata CC. Creating a DS4podman container

- ds4podman.bin is the generic tool to maintain the podman containers
- ds4podman.bin make creates a container
- Other interesting options:
  - ds4podman.bin status
  - ds4podman.bin start
  - d4podman.bin stop
  - ds4podman.bin showconf
  - ds4podman.bin chart

All ds4podman options has the required option:

--container-name <container>

Since we do have one podman container for each cluster it makes sense to use the cluster name (defined in the ExaCC) as the name of the container.

```
/opt/ds4podman/ds4podman.bin make \
      --tenancy-id ocid1.tenancy.oc1…… \
      --ociregion eu-frankfurt-1 \
      --shape Exadata.X9M \
      --network-type client \
      --noregionyum \
      --proxyhost <proxyhost> \
      --proxyport <proxyport> \
      --nohttps \
      --user-id ocid1.user.oc1……. \
      --keyfingerprint 17:82:77:31…… \
      --privatekey /opt/scale_cpu/ds4/cloud-keys/db-c911oed.pem \
      --ssh-user opc \
      --logvolume /var/log/dynamicscale/cbp/logs \
      --chartvolume /var/log/dynamicscale/cbp/charts \
      --container-name db-c911oed-clu \
      --vm-cluster-id ocid1.vmcluster.oc1.eu-frankfurt-1….. \
      --opcsshkey /opt/scale_cpu/ds4/cloud-keys/db-c911oed \
      --skip-node xx.xx.xx.xa, xx.xx.xx.xx.xb,xx.xx.xx.xc,xx.xx.xx.xd \
      --nodecount 4 \
      --loadtype max \
      --maxthreshold 80 \
      --minthreshold 20 \
      --maxocpu 84 \
      --minocpu 4 \
      --ocpu 16 \
      --interval 60 \
      --scheduling “Friday:18-23:4;Saturday:0-23:4;Sunday:0-23:4”
```

Nordea Exadata CC. Ocpu scaling parameters

The following parameters control the OCPU allocation.

Loadtype: Can be max or Avg.
Maxthreshold: Max acceptable load (in %)
Minthreshold: Min acceptable load (in %)
Ocpu: Amount of CPU to add/remove during scale

The plugin returns the load and compares it to the max and min threshold. If load is over or under threshold a scale command will be send to the OCI.

Maxocpu: Upper limit
Minocpu: Lower limit
Scheduling: Hard coded values in specific time periods

The parameter Scheduling will always take preference.

Notes:

Oracle does not guaranty any timing of when a request for scaling will have effect – typically however we see this to be between 6 and 10 minutes – Through private discussions we have learned that an internal timeout is set to be 15 min, executed twice.

So, for a fixed schedule we need to issue the needed scaleup request about 30 minute before needed, to be on the safe side, and a ramp up interval for a regression test would also require some time depending on how many OCPUs have been configured for each the scale-up / scale-down operation

Nordea Exadata CC. ODyS Charts

Charts is a great tool to give an overview how a cluster has been allocated over a given period of time.

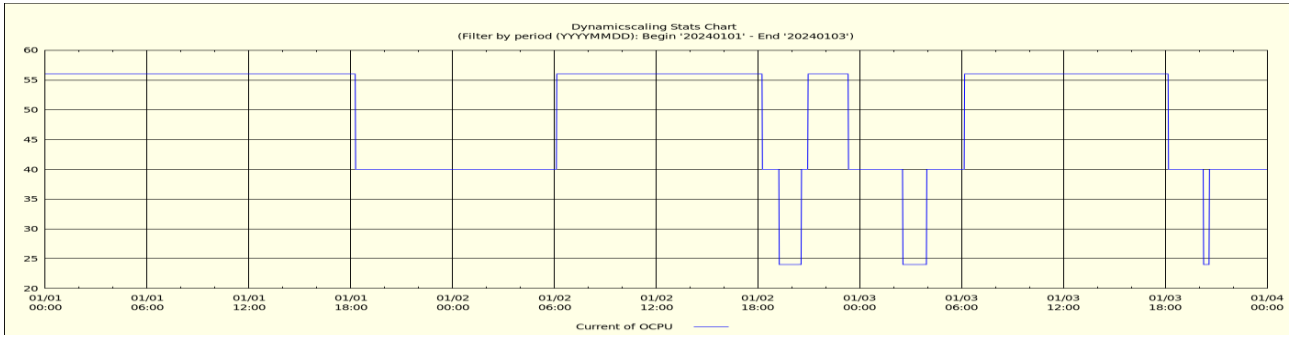
Arguments:

```
  --cocpu        Current OCPU
  --pload        Plugin load
  --datefilter   Filter by date (valid format YYYYMMDD)
  --begin/--end  Filter by period (valid format YYYYMMDD)
  --png          Make a png image
  --csv          Make a csv file
  --label        String to be used in title and file name
  --out          Path where to save the html/png file inside the container
  --logfile      Log file name (Default: ' dynamicscaling-chart.log')
  --logpath      Log file path (Default: '/tmp')
```

Nordea Exadata CC. ODyS Charts

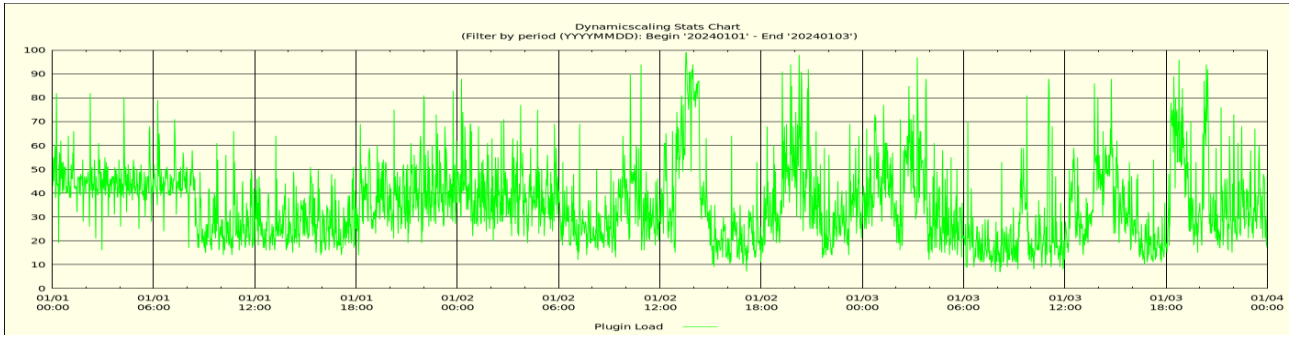
Example: Show the allocated ocpu on cluster between 1/1-2024 and 3/1-2024

ds4podman.bin chart --container-name db-c911hhd-clu --cocpu --begin 20240101 --end 20240103 --png



Example: Show the calculated load on cluster between 1/1-2024 and 3/1-2024

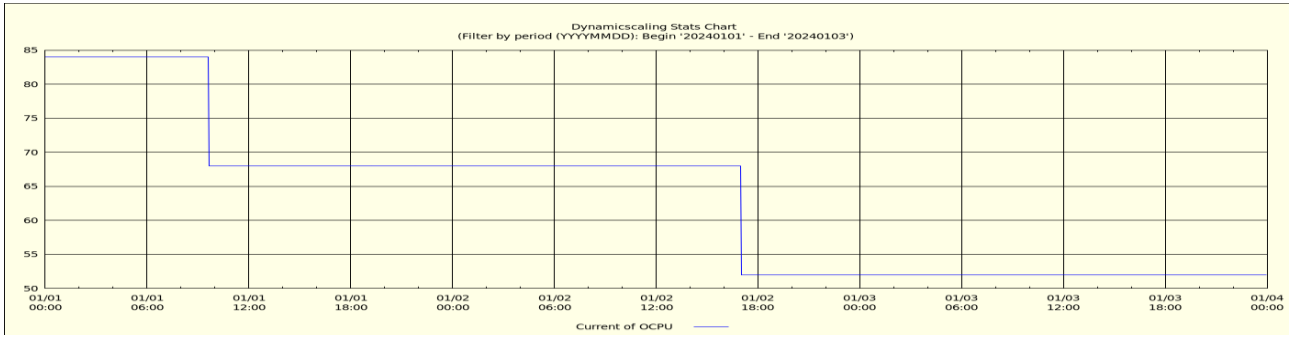
ds4podman.bin chart --container-name db-c911hhd-clu --pload --begin 20240101 --end 20240103 --png



Nordea Exadata CC. ODyS Charts

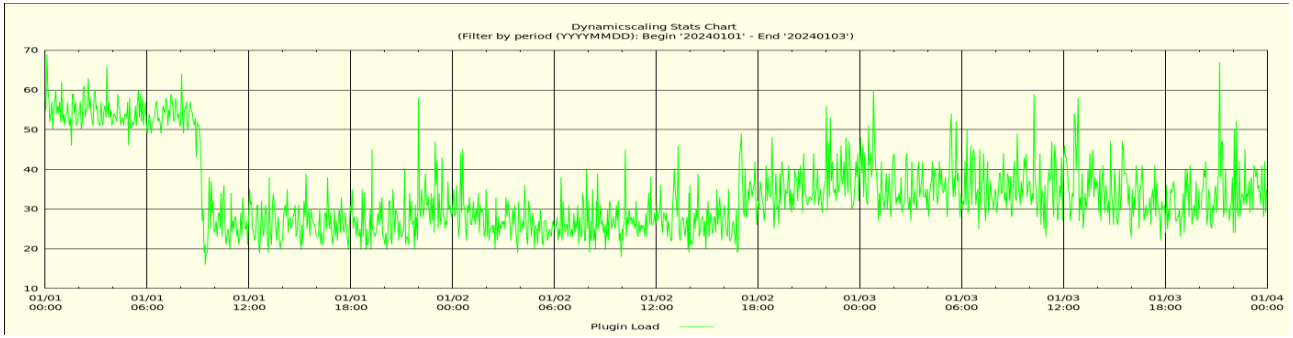
Example: Show the allocated ocpu on cluster between 1/1-2024 and 3/1-2024

ds4podman.bin chart --container-name db-c912oed-clu --cocpu --begin 20240101 --end 20240103 --png



Example: Show the calculated load on cluster between 1/1-2024 and 3/1-2024

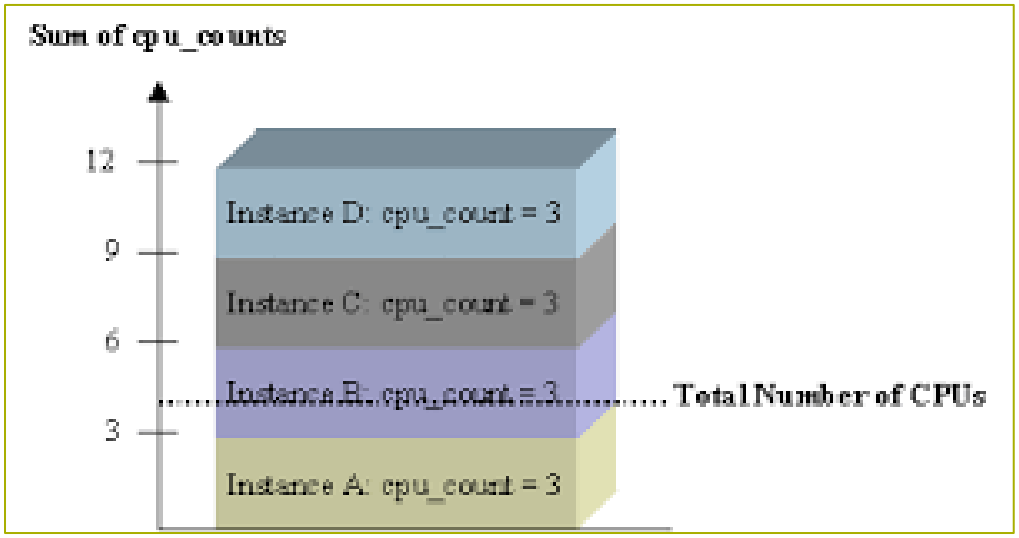
ds4podman.bin chart --container-name db-c912oed-clu --pload --begin 20240101 --end 20240103 --png



Nordea Exadata CC. Short discussion regarding Cpu_Count

On each cluster a number of databases are running – in our case in the order of 20 – 40 container databases serving a number of pluggable databases.

The database kernel have for quite some time been able to adjust to a change of the CPU_COUNT parameter on the fly – so we will need to discuss how an up- or downscale of OCPUs at the cluster level should have effect on the individual database level.



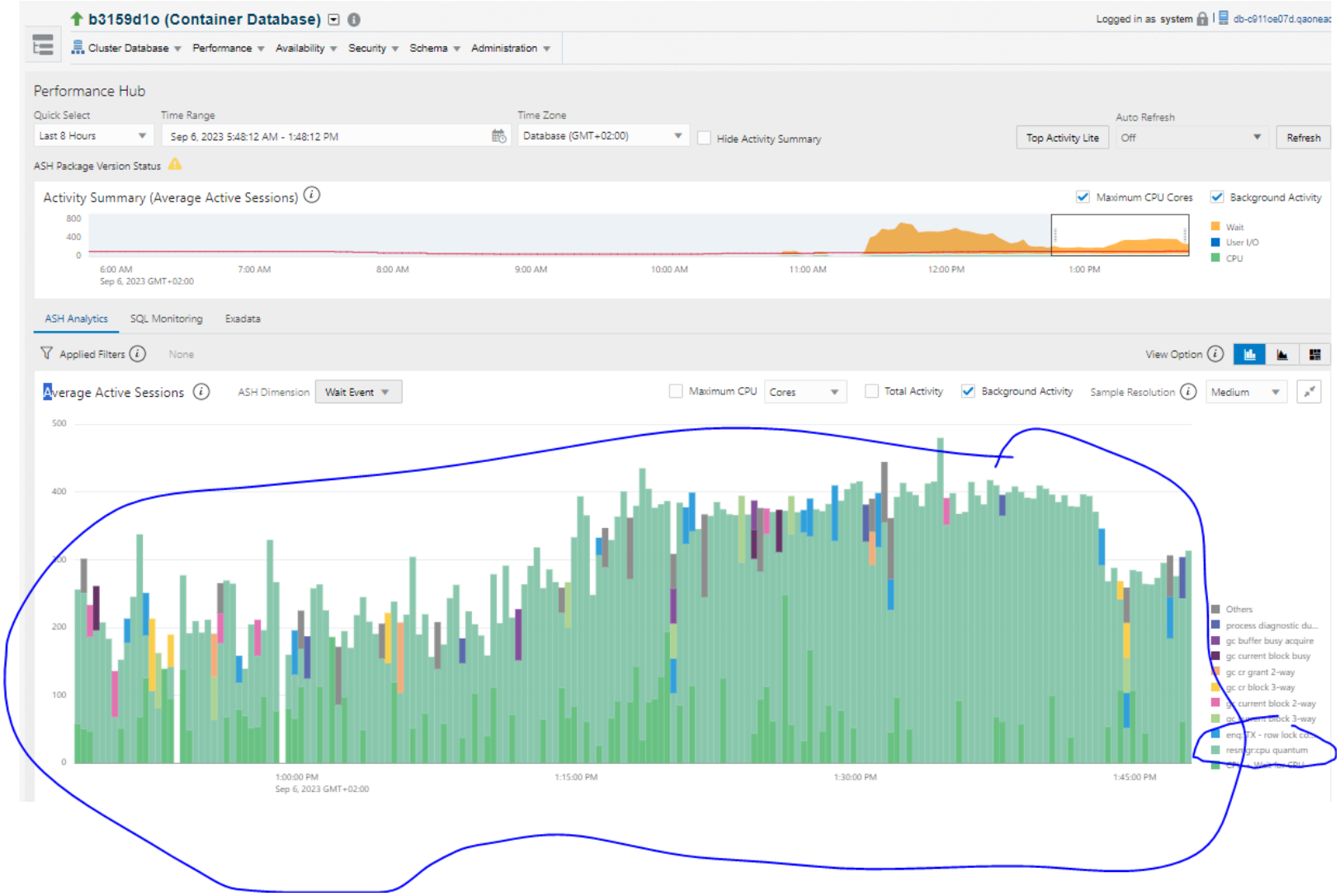
Nordea Exadata CC. Introduction to DcpuCount

Oracle does deliver a so-called *dcpucount* utility, which will be triggered when the OCPUs for a cluster is changed (manually or automated), and will calculate a new CPU_COUNT for the databases (CDBs / PDBs) on the cluster – either to be fixed or to be a percentage of the cpu capacity of the cluster – here is an example:

```
CDBA:::6       # --> Fixed CPU_COUNT=6
CDBA:PDBA:50:  # --> PDBA 50% of container CPU_COUNT
CDBA:PDBB::2   # --> PDBB Fixed CPU_COUNT=2
CDBB::15:      # --> 15% of host's OCPUs
CDBC::15:      # --> 15% of host's OCPUs
CDBD::25:      # --> 25% of host's OCPUs
CDBE::50:      # --> 50% of host's OCPUs
```

(ODCC) Dynamic CPU Count - automation utility to change cpu_count dynamically (ExaDB-D/ExaDB-C@C) (Doc ID 2915837.1)

Nordea Exadata CC. Issues with DcpuCount



Nordea Exadata CC. Issues with DcpuCount (continued)

What seems to be the issue for us (seen above) happens when we are using both OCPU scaling on the cluster level as well as dcpucount on the database level. Here we see that scaling down on the cluster will indeed follow the DB cpu distribution and downscale the databases in reflecting the DB cpu distribution configuration file. But It seems to be hard automatically to scale up again, as increasing pressure on some of the databases, will not always translate to increase of the load of the cluster as the usage in '*Instance Caging*' automatically cap the cpu usage of the database where the increase of cpu is demanded from the applications.

This issue have been acknowledge by the Oracle development engineer who have done the development. We have thus disabled the dcpucount utility and set the CPU_COUNT to the values for the databases on a fully scale cluster, with the drawback that databases will fight for resources when the cluster is having few OCPUs

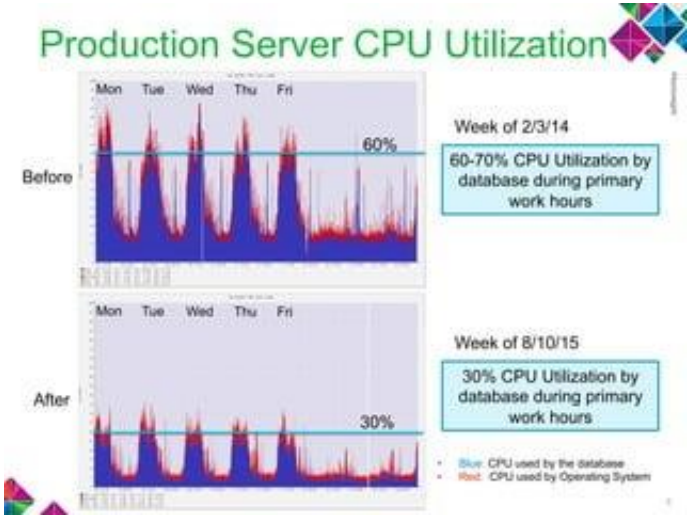
Configuring and Monitoring Instance Caging (Doc ID 1362445.1)

Nordea Exadata CC. Discussion regarding ODyS

Even if the combination of cluster dynamic scaling and database CPU count adjustment do not work elegantly, we still envision that periods with low user activity should theoretically use less CPU capacity and should thus be able to have only a few OCPUs – and be less costly. But can we assume that seemingly quiet weekends to have almost quiet databases?

Point is that typically in weekends the databases have an open maintenance window, where internal house-keeping is executed, which could be asking the cluster for more CPU resources than expected.

Let's look closer at statistics gathering, automatic advisors and OEM jobs checking audit information for not-successful connects.



Nordea Exadata CC. Discussion regarding ODyS (Continued)

Typically statistics gathering will be running every nights and weekends, looking for stale database objects. If f.ex a huge database table has a larger *Degree* more slaves would be spanned by the automatic statistics gathering procedure, potentially driving up demand for OCPUs for the cluster.

Have a look for such database tables and consider if the degree Can be reset to 1, especially if such a table have frequent DMLs

Other automatic advisors may be active such as the tuning and space advisors.

Sometimes we see the space advisor to be active during weekends for half or full hours, potentially driving up the demand for OCPUs, as this advisor will typically start at the same time for all databases in the cluster.

You may consider to disable this advisor if you are not using advises from this advisor for a longer period.

We have kept the Tuning advisor enabled as we need the advises and as this advisor do not spawn many slaves

Nordea Exadata CC. Discussion regarding ODyS (Continued)

We have seen OEM jobs checking the unified audit information for not-successful connects – even if unified auditing is now selected, default will still ask unified audit to collect (among other information) not successful connects. And we have seen such jobs to be very CPU hungry, even if the unified audit trail have been purged. If you are checking such events by other means, you could ask OEM to bypass such check, to avoid scaling up the OCPUs for the cluster for this reason.

EM 13c: EM processes using high cpu utilization on target DB (Doc ID 2785707.1)

ORA-600 [15803] Reported For Unified Audit Query From OEM (Doc ID 2953532.1)

Exadata CC Platform

The End

Be careful out there
Best wishes Thorbjørn / Martin