ORACLE

# Adattárház – Oracle ExaCS, Frankfurt

2022.08.01 – 2023.03.31

Erdősi Zoltán (zoltan.erdosi@erstebank.hu)
Tasi József (jozsef.tasi@oracle.com)

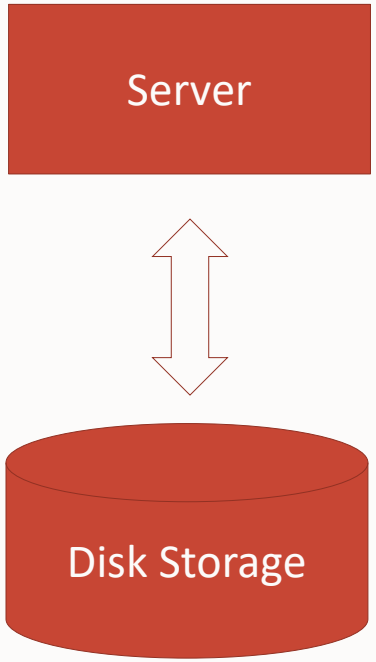
# Safe harbor statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

# POC tartalma: Egy banki adattárház napi és havi töltéseinek vizsgálata EXA környezetben

## Budapest

- 20 db P9 CPU
- 1,4 TB memória




- 4 db 4gbit/csatorna

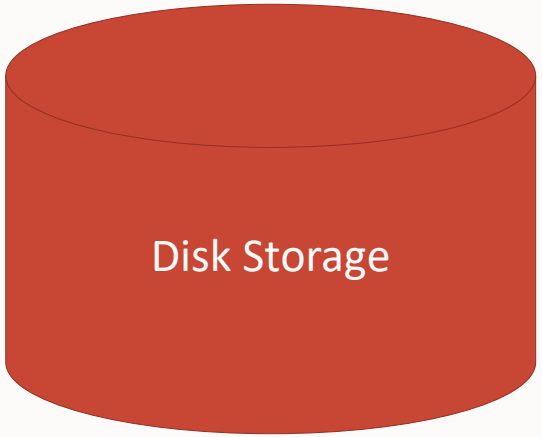
- 350 TB
- No RAC

## EXA CS, Frankfurt

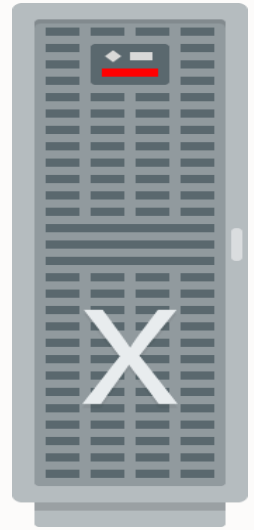
- 2 db VM Cluster, 126 db OCPU/Cluster
- 2 x 1,4TB memória
- 2 db RAC node

- 3 db Storage Servers (Data Cell)
- 60TB

# Kiindulási állapot




- Db méret csökkentés
- Endian forgatás
- Block méret váltás



350 TB adattárház méret

32k-s block szervezésű

AIX 7.2-es operációs rendszeren fut az adatbázis

Adatbázis verziószáma: 19.14

Lokálisan nincs hely a storage-on a mentés végrehajtáshoz

Csonkolt adattárház mérete 60 TB

Linux-ra kell konvertálni az adatbázist (endian forgatás)

EXADATA adatbázis verziószáma 19.15

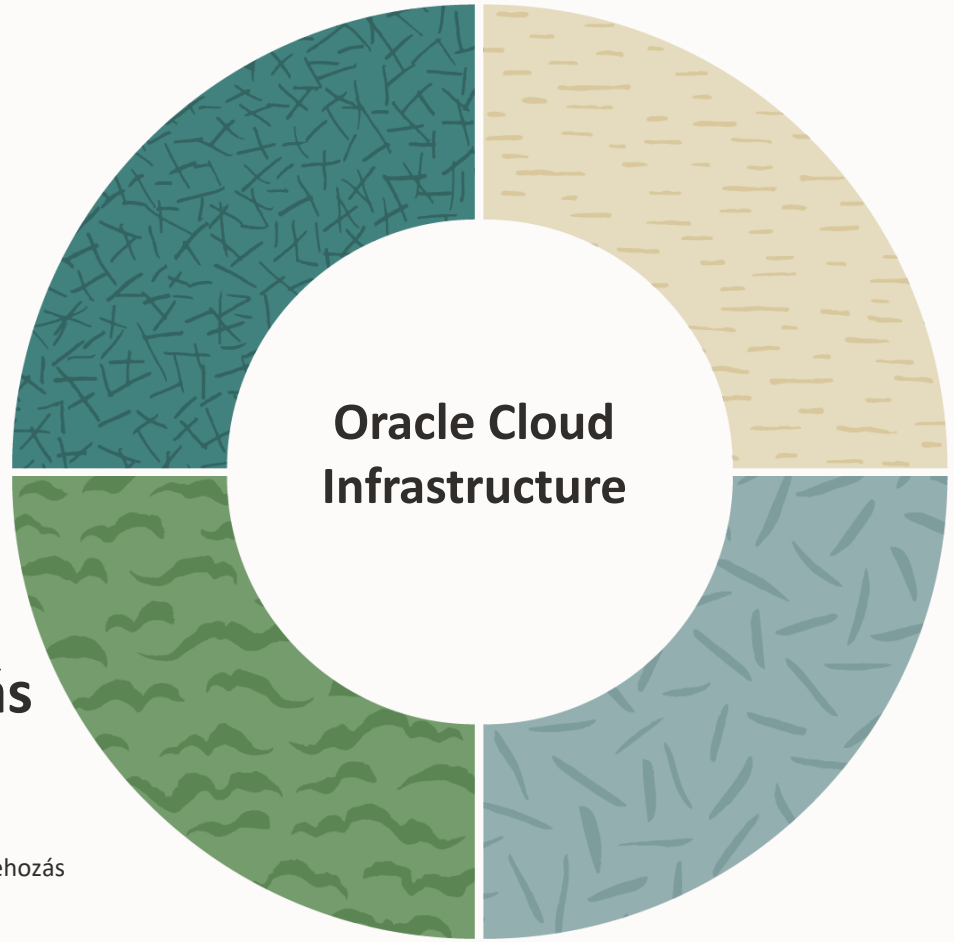
csak 8k-s PDB-t lehet létrehozni cloud GUI segítségével ehhez a 8k-s CDB-hez kellett csatolni egy 32k-s PDB-t

# Kezdeti lépések az EXADATA CS használata előtt

## Hálózat kialakítás

**OCI virtuális hálózat kialakítás**

**Lokális hálózat és OCI virtuális hálózat között VPN kialakítás**




## OCI storage kialakítás

**Egy új bucket létrehozás**

**Kulcspár generálás az eléréshez**

**OCI driver telepítés RMAN-hoz AIX szerveren (opc_installer.zip)**

## Környezet kialakítás

**EXADATA infrán a környezet kialakítása**
Új PDB kialakítás 32k block size

- EXADATA-n kívül egy 32k-s CDB / PDB létrehozás
- 32k-s PDB unplug
- EXADATA-ba 32k-s PDB plug
- PDB_PLUG_IN_VIOLATIONS ellenőrzés és (DB_32k_CACHE_SIZE > 0)

# Adatbázis másolás eltérő ENDIAN esetén – 1. verzió


**Image copy (1)**

- RMAN segítségével, sok szálon (>128) , nagyméretű chunk size (>1 GB) esetén gyors a nagyméretű DB mentés
- OCI storage-ból a visszatöltés egy szálon megy, ami nagyon lassú
- OCI storage-ban a mentést át kellett konvertálni archive tier-be és onnan tudta csak olvasni az RMAN, ami szintén sok időt vett igénybe még átkonvertálta
- Készült egy pár METALINK SR a felmerült problémákról.

# Adatbázis másolás eltérő ENDIAN esetén – 2. verzió


**Táblaterenként XTTS (2)**

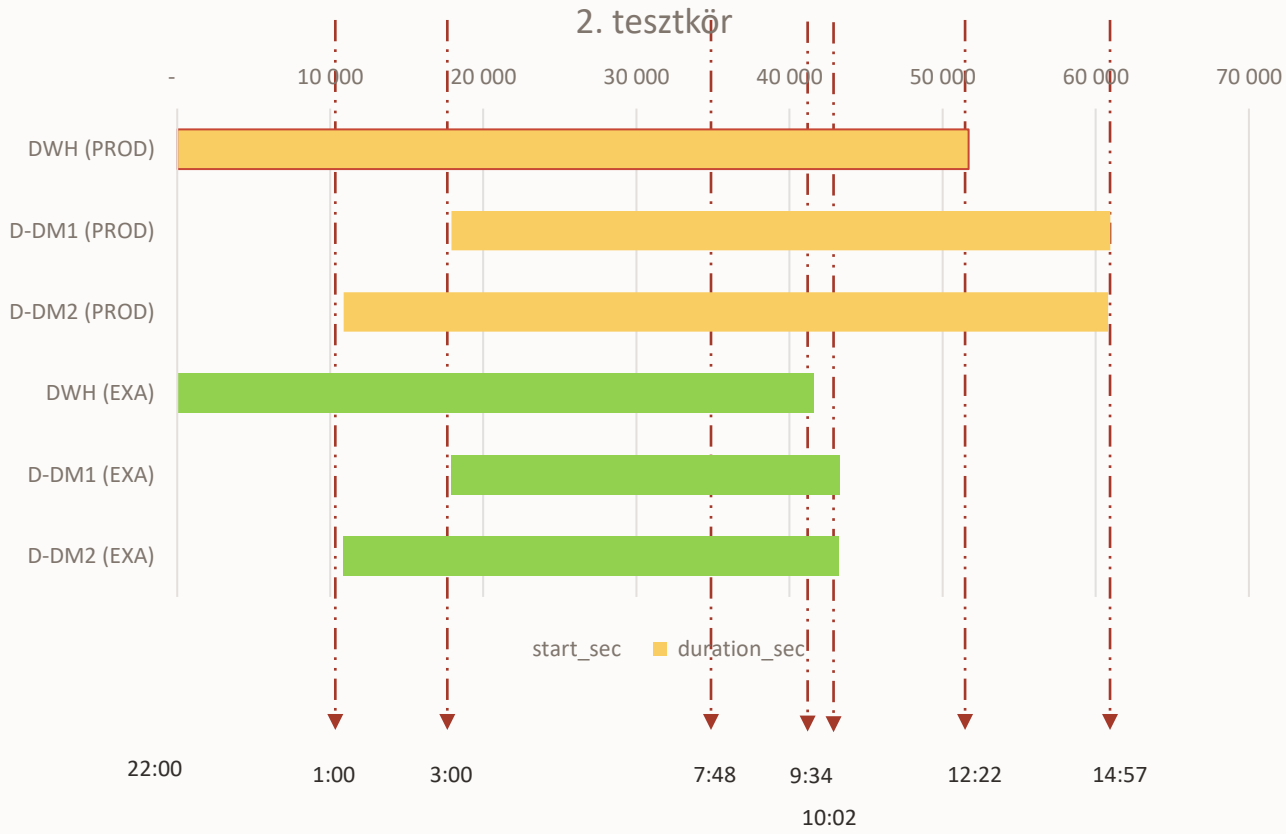
- Ellenőrizni kell az összes táblateret, hogy menthető-e önállóan (nem lehet több táblateret használni partíciónált tábla esetén. Mindennek egy táblatérbe kell lennie)
- Read only-ba kell lenni a táblatereknek még az RMAN mentés + TTS export elkészül
- AIX-on készült TTS .dmp nem kompatibilis a linux-os verzióval. Minden RMAN mentés után kézzel kell készíteni egy külön TTS exportot
- Itt is lehet több szálon menteni, ahány adatfile van annyi RMAN channel (138 db 128 GB datafile, 128 channel, 3,5 óra)
- Lokálisan fut a bitkonverzió
- Visszatöltésnél pedig ahány backup piece keletkezett annyi RMAN channel (60 db backup piece, 2 óra)

# Tesztelési módszertan

| | Szinkron működés (SYNC)<br>A forrásrendszerek érkezési idejéhez ütemezett teszttöltések | Nem szinkron működés (NO SYNC)<br>Feltételezük, hogy minden forrásadat meg van, ezért egyszerre indítjuk el a töltéseket |
|---|---|---|
| Napi és havi adattárház töltések végrehajtása | • Napi DWH töltés, 100+ forrásrendszer<br>• Napi adatszolgáltatások: 4 adatpiac (D-DM1, D-DM2, D-DM3, D-DM4)<br>• Havi adatszolgáltatások: 2 adatpiac (M-DM1, M-DM2) | |
| Terhelhetőség/Skálázhatóság tesztelése | **Fix konfiguráció**<br>• PROD: 20 processzor 160 szál (4 szál/CPU)<br>• EXA: 20+20 OCPU 80 szál (2 szál/OCPU)<br>**Dinamikus skálázás**<br>• dynamic scaling a gép terhelése alapján állítja be a VM clusterben az OCPU számokat (le- és felskálázás) | |

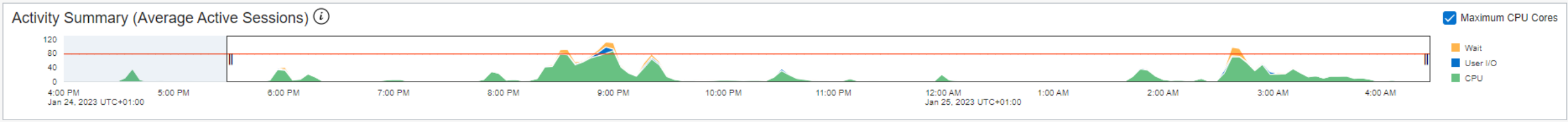
# 2. Tesztkör
## SYNC

| Befejezési idő | PROD | EXA |
|---|---|---|
| DWH | 12:22 | 9:34 |
| D-DM1 | 14:57 | 10:03 |
| D-DM2 | 14:54 | 10:02 |




Parameters:

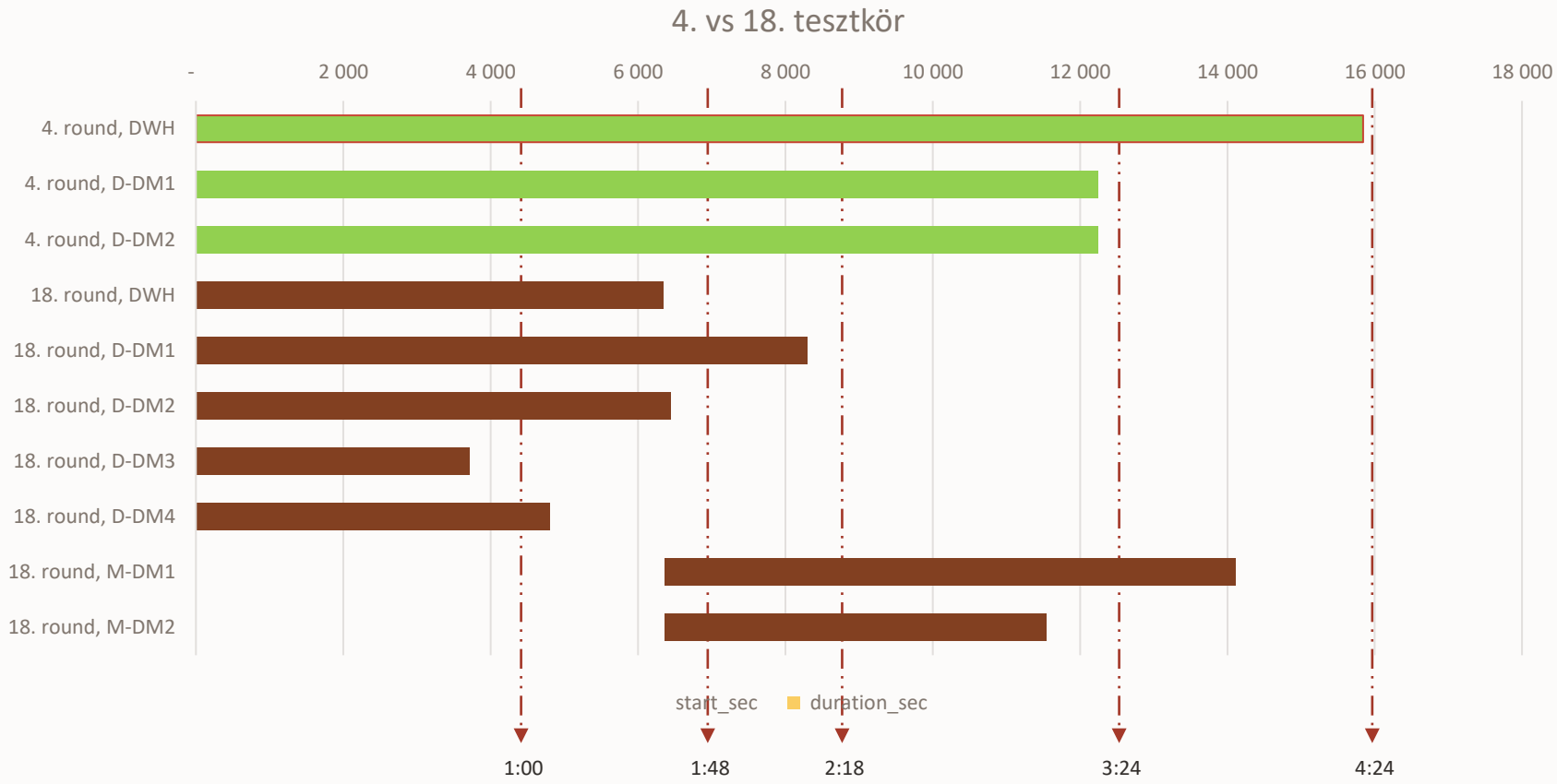
- CPU 20+20 (80)
- Degree of Parallelism (DOP) 2x48
- parallel_max_servers 2x800
- pga_aggregate_target 2x400G
- sga_max_size 2x200G



# 4. vs 18. Tesztkör

## NO SYNC

| Átfutási idő | EXA 4. | EXA 18. |
|---|---|---|
| DWH | 4:24 | 1:48 |
| D-DM1 | 3:24 | 2:18 |
| D-DM2 | 3:24 | 1:48 |
| D-DM3 | - | 1:00 |
| D-DM4 | - | 1:18 |
| M-DM1 | - | 2:12 |
| M-DM2 | - | 1:24 |



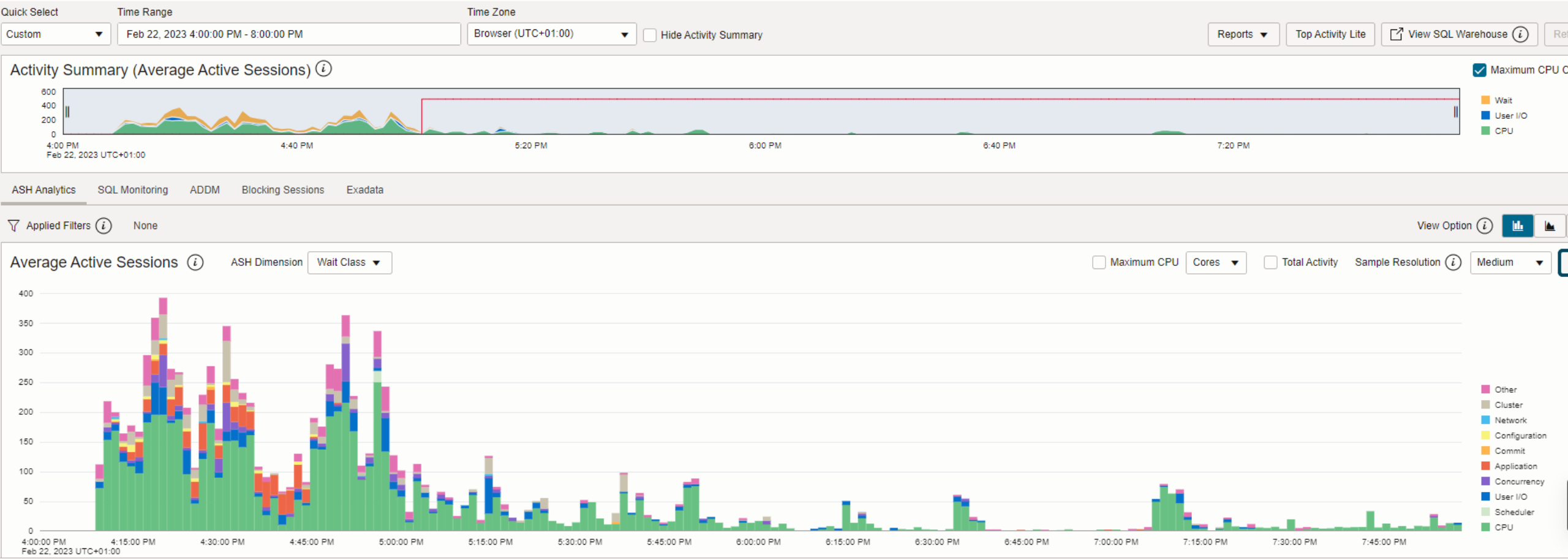

Parameters (4):

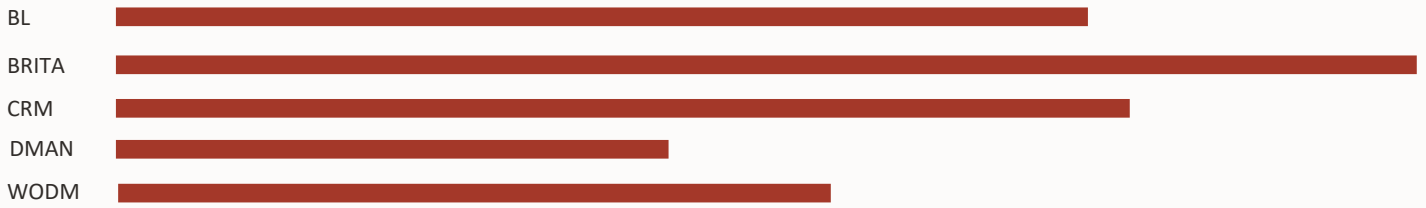
- CPU 20+20 (80)
- Degree of Parallelism (DOP) 2x48
- parallel_max_servers 2x800
- pga_aggregate_target 2x400G
- sga_max_size 2x200G

Parameters (18):

- CPU 126+126 (504) max 200 threads were used
- Degree of Parallelism (DOP) 2x48
- parallel_max_servers 2x1800
- pga_aggregate_target 2x600G
- sga_max_size 2x400G

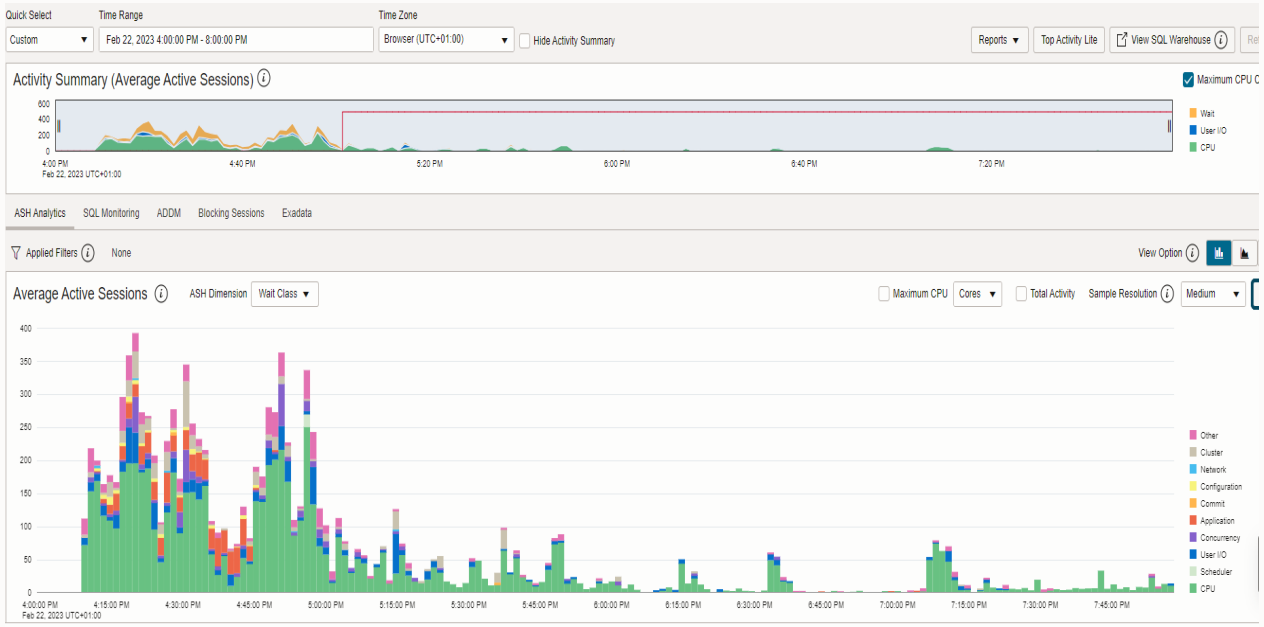
# 18. tesztkör

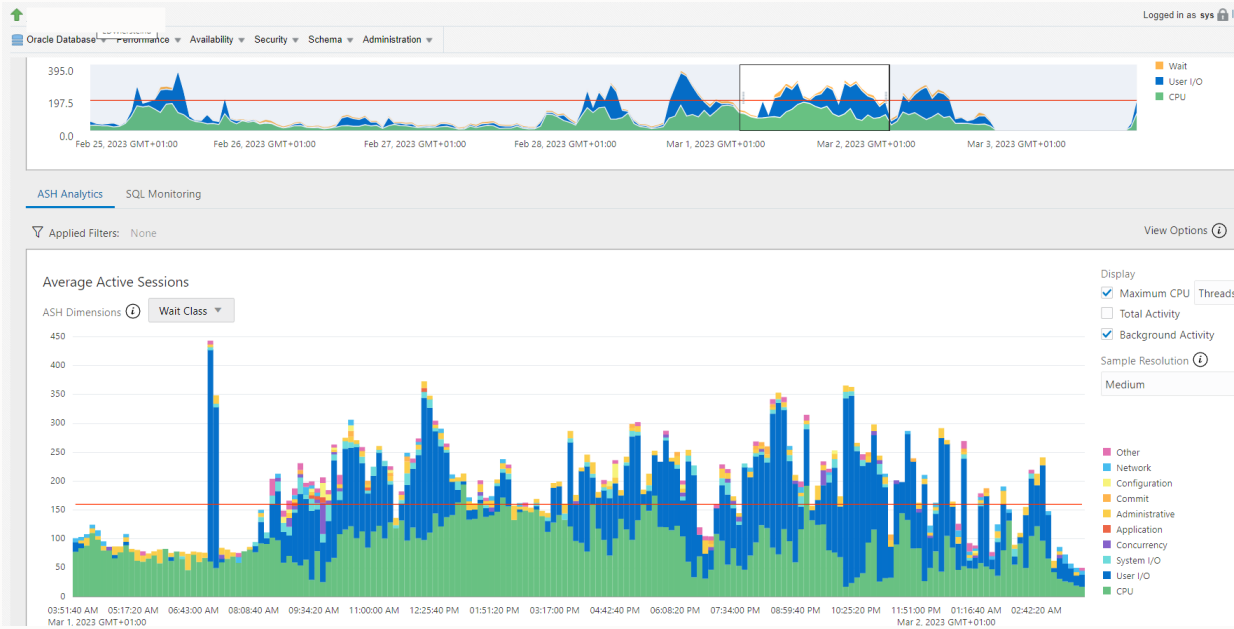

Quick Select
Custom
Time Range
Feb 22, 2023 4:00:00 PM - 8:00:00 PM
Time Zone
Browser (UTC+01:00)
Hide Activity Summary
Reports
Top Activity Lite
View SQL Warehouse
Activity Summary (Average Active Sessions)
Maximum CPU C
600
400
200
0
4:00 PM
Feb 22, 2023 UTC+01:00
4:40 PM
5:20 PM
6:00 PM
6:40 PM
7:20 PM
Wait
User I/O
CPU
ASH Analytics
SQL Monitoring
ADDM
Blocking Sessions
Exadata
Applied Filters
None
View Option
Average Active Sessions
ASH Dimension
Wait Class
Maximum CPU
Cores
Total Activity
Sample Resolution
Medium
400
350
300
250
200
150
100
50
0
4:00:00 PM
Feb 22, 2023 UTC+01:00
4:15:00 PM
4:30:00 PM
4:45:00 PM
5:00:00 PM
5:15:00 PM
5:30:00 PM
5:45:00 PM
6:00:00 PM
6:15:00 PM
6:30:00 PM
6:45:00 PM
7:00:00 PM
7:15:00 PM
7:30:00 PM
7:45:00 PM
Other
Cluster
Network
Configuration
Commit
Application
Concurrency
User I/O
Scheduler
CPU


BL
BRITA
CRM
DMAN
WODM


BITA
CASIR

# IO: EXA vs PROD



# AWR elemzés alapján

| | EXA | AIX |
|---|---|---|
| User I/O wait (%) | 8,0% | 47,0% |
| CPU (%) | 78,0% | 8,2% |
| olvasás válaszidő max (ms) | 3,41 | 90-100 |

**AIX**

**47%**-ban az adatbázis szerver User I/O eseményekre vár; **CPU 8,2%**

A User I/O-nak, mint várakozási osztálynak az átlagos válaszideje ugyan alacsony, de ha csak a direct path read-et nézzük, amire az adatbázis idő 30%-a elmegy, ott látszik, hogy közel 3.7x a válaszidő az elfogadott, normál válaszidőnek, ami 10ms lenne.

Ez az az várakozási esemény, amitől a leginkább szenvedünk AIX-on, és ami fel tud menni átlag **90-100**ms környékére is, ahogy növekszik a terhelés.

Ez az I/O alrendszer lassulása okozhatja, amit adatbázis szerver oldali CPU növeléssel nem fogunk tudni megoldani.

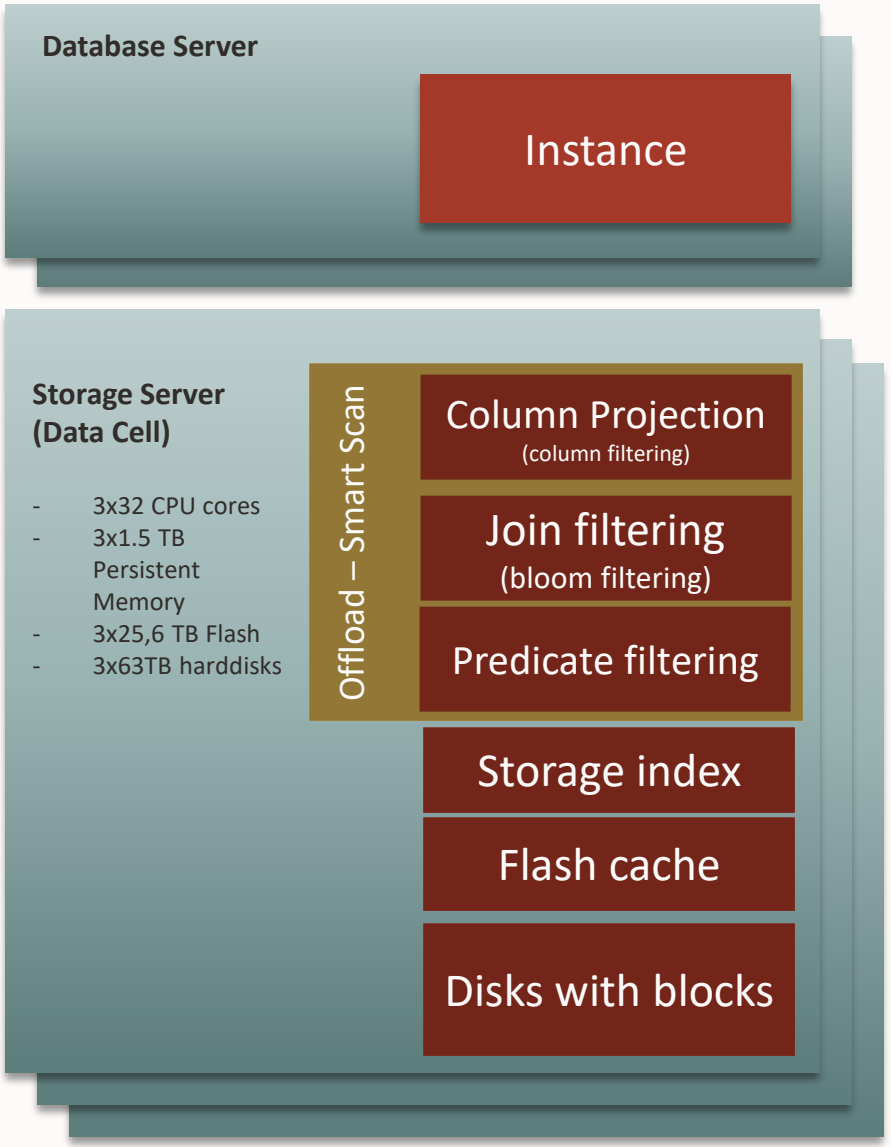
**EXA**

Az Exadatán ezzel szemben leginkább CPU-ra van elkönyvelve az idő **78%**-a, az adatbázis idő ezen kívül **8%**-ban I/O-ra, 4%-ban Concurrency, illetve 3%-ban klaszteres várakozásokra megy el.

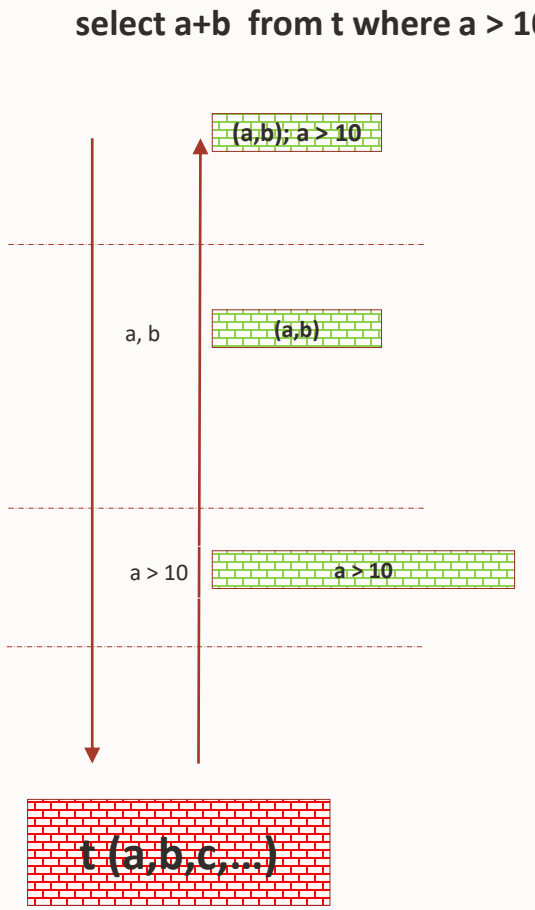
A Cluster-es és Concurrency várakozások a RAC-ra történő finomhangolással még tovább optimalizálhatóak lehetnek.

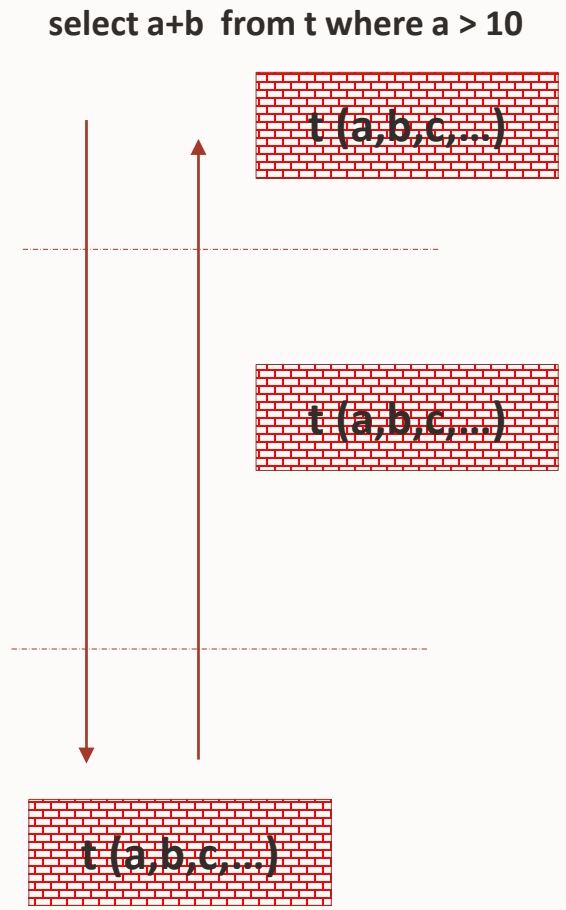
Fontos kiemelni, hogy a legnagyobb szeletet kiszakító olvasási műveletek válaszideje **2-4 ms** körül van.

# EXA DATA IO




# Conventional IO




**Smart Scan (Offload):** Data-intensive processing* runs in massively parallel Exadata Storage, bypassing network bottlenecks and freeing up DB CPUs

Quick Select: Custom | Time Range: Feb 22, 2023 4:00:00 PM - 8:02:00 PM | Time Zone: Browser (UTC+01:00) | Hide Activity Summary | Refresh

## Activity Summary (Average Active Sessions)


Performance | Health | ADDM | Top Consumers | Cells | Disks | Smart I/O | I/O Reasons | Configuration

## Smart I/O

| | |
|---|---|
| Bytes Requested: | 70 TB |
| Storage Index: | 39.6 TB |
| Flash Cache: | 4.6 TB |
| Offload: | 67.4 TB |
| Passthru: | 1.2 TB |
| Reverse Offload: | 95.9 MB |

fra304015exdcl06: 22.4 TB
22.4 TB processed by the cells and not returned to the database
96.18 % of bytes requested

## Flash Log

| | |
|---|---|
| Redo Bytes: | 1.6 MB |
| | 13M |
| Errors: | 0 |

## Flash Cache

| | |
|---|---|
| Configuration: | Mode - WriteBack |
| Space Usage: | 69.7 MB |
| I/O Requests: | 4K IOPS, 9K IOPS |
| I/O Throughput: | 1.7 KBPS, 1.9 KBPS |
| Read Misses: | 17.15 /s |

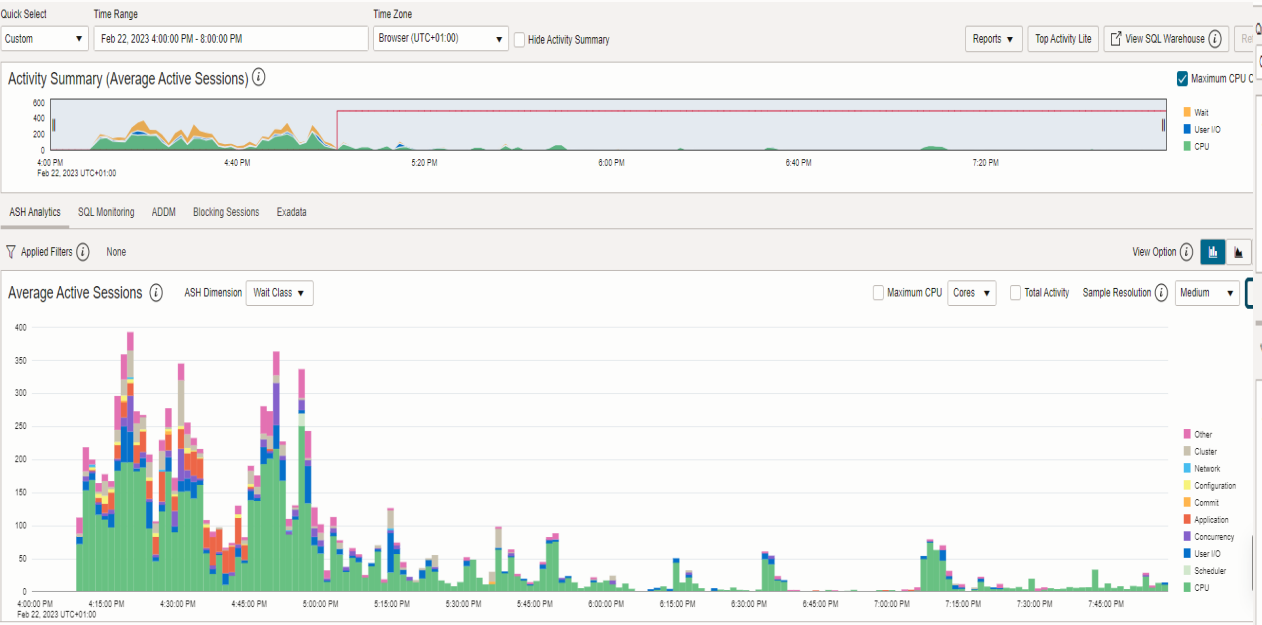
## Cell Details

Smart I/O | Flash Log | Flash Cache

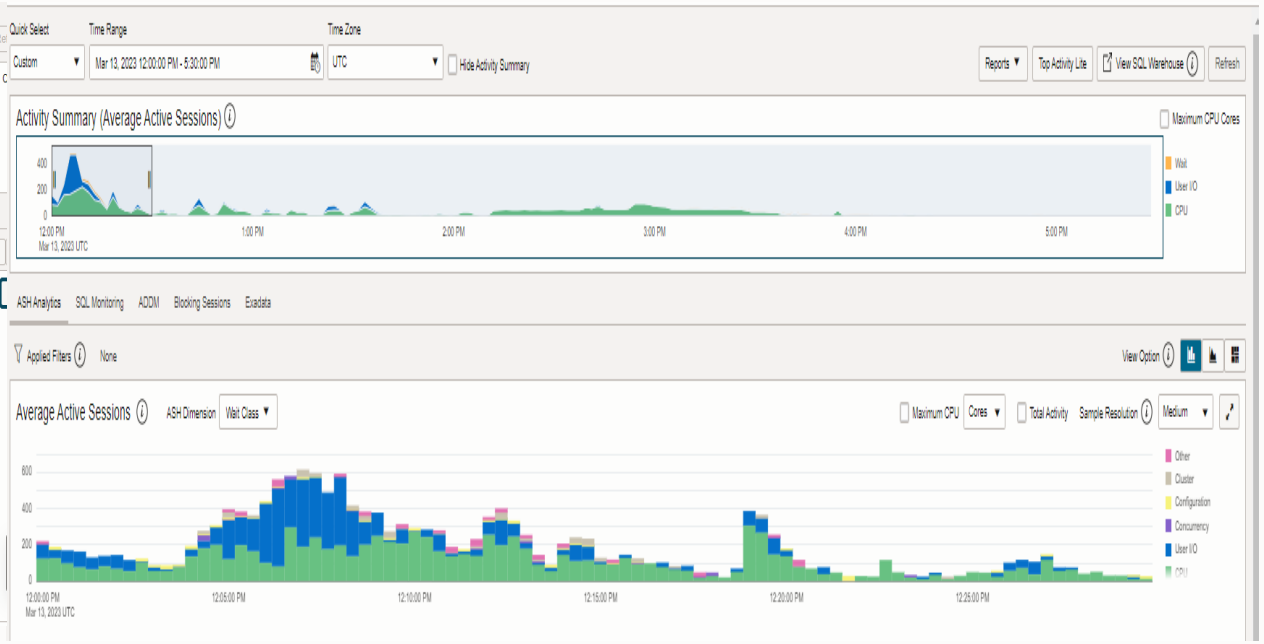
| Cell Name | Bytes Requested | Storage Index | Flash Cache | Offload | Passthru | Reverse Offload |
|---|---|---|---|---|---|---|
| All | 70 TB | 39.6 TB | 4.6 TB | 67.4 TB | 1.2 TB | 95.9 MB |
| fra304015exdcl07 | 23.4 TB | 13.2 TB | 1.6 TB | 22.5 TB | 407.7 GB | 29.2 MB |
| fra304015exdcl08 | 23.3 TB | 13.2 TB | 1.5 TB | 22.4 TB | 408.9 GB | 29.4 MB |
| fra304015exdcl06 | 23.3 TB | 13.2 TB | 1.5 TB | 22.4 TB | 405 GB | 37.3 MB |

# IO

## w/ Offload




## w/o Offload



# Skálázási lehetőség

| Módszer | |
|---|---|
| Manuális skálázás | - |
| Dynamic scaling tool<br>• Automatikus a VM node-ok load-ja alapján | + |
| DWH ütemező kiegészítése<br>• Automatikus a DWH töltések alapján<br>• Ez kiegészíthető a Dynamic Scaling tool alkalmazásával a DWH töltésen kívüli felhasználáshoz | ++ |