

Container only – New Features for Multitenant in Oracle 21c

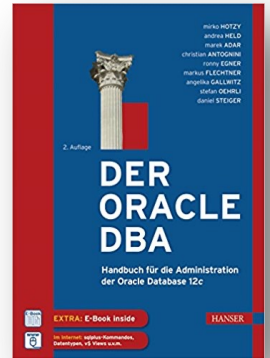
Markus Flechtner

HI!



MARKUS FLECHTNER PRINCIPAL CONSULTANT

- Trivadis Germany GmbH
- Studied Mathematics a long time ago
- Focus
 - Oracle High Availability
 - Database Upgrade + Migration
- Teacher: RAC, New Features, Multitenant, PostgreSQL
- Twitter @markusdba
- Blog: markusdba.net
- Co-author of the book The Oracle DBA (2016)

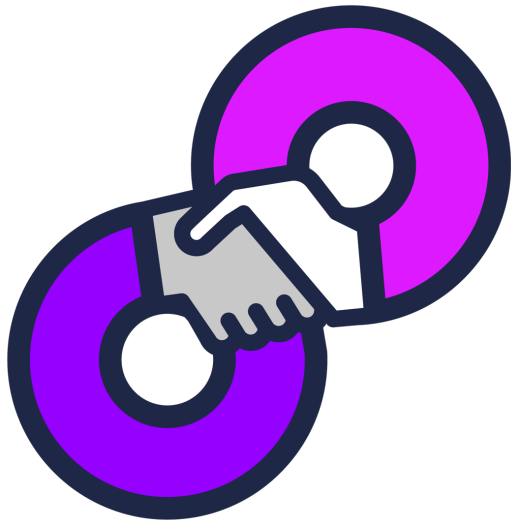


ORACLE®
ACE

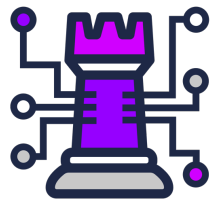
DOAG



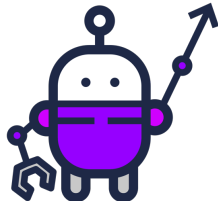
trivadis Part of Accenture



- Together we are 1500 specialists at 34 locations in Switzerland, Germany and Austria with a focus on **Data & Applied Intelligence**.
- Together we support you in the **intelligent end-to-end use of your data**.
- We cover the entire spectrum: from the **development and operation of data platforms and solutions**, to the refinement of data as well as **consulting and training**.
- We achieve this through the unique combination of Trivadis' **technological expertise** and Accenture's **strategic know-how** in the field of data.



**BUSINESS
ANALYTICS**



**BUSINESS
AUTOMATION**



**MODERN DATA-
PLATFORMS**



**MODERN CLOUD-
INFRASTRUCTURE**

trivadis
Part of **Accenture**

AGENDA

- Introduction
- Upgrade & Migration to Multitenant
- Resource Management & Security
- Multitenant & Oracle GridInfrastructure/RAC
- Multitenant & DataGuard
- Miscellaneous
- Conclusion

Introduction

REMEMBER ...

- The classical Non-CDB Architecture is desupported with Oracle Database 21c
- Multitenant Architecture („CDB Architecture“) is the only architecture

REMEMBER ...

- Oracle Database 21c is an "Innovation Release"
 - Available for Linux x86-64bit, Windows and HP-UX only
 - Use it for testing and evaluating new features, but not for production systems
- Support for Oracle Database 21c
 - Premier Support ends June 30, 2023
 - No extended support, error correction ends June 30,2023
- Expected next Long Term Support Release: Oracle Database 23c

Upgrade & Migration to Multitenant

10 AUTOUPGRADE

- AutoUpgrade is a well known tool for upgrading and migrating your database
 - Available since 2019
 - Download the latest release from MOS (Note 2485457.1)
 - Recommended method for database upgrades

- Automatic upgrade of Oracle Databases
 - PreUpgrade Checks & FixUps
 - Database Upgrade
 - PostUpgrade Tasks (e.g. TimeZone Upgrades)
 - Migration from Non-CDB to PDB

11 USING AUTOUPGRADE FOR A NON-CDB TO PDB MIGRATION & UPGRADE TO 21C

- Create the target CDB (21c) beforehand
- Autoupgrade will
 - Plug in the Non-CDB as a PDB
 - Upgrade the PDB to 21c

12 AUTOUPGRADE CONFIGURATION FILE

- Example configuration file:

```
global.autoupg_log_dir=/home/oracle/testconv

testconv1.dbname=NCDB
testconv1.start_time=NOW
testconv1.source_home=/u00/app/oracle/product/19c
testconv1.target_home=/u00/app/oracle/product/21c
testconv1.sid=NCDB
testconv1.log_dir=/home/oracle/testupgrade
testconv1.upgrade_node=localhost
testconv1.target_version=21.4
testconv1.target_cdb=TESTCDB
testconv1.target_pdb_copy_option=file_name_convert=
    ('/u01/oradata/NCDB','/u01/oradata/TESTCDB/TPDB')
testconv1.target_pdb_name=TPDB
[...]
```

13 „REPLAY UPGRADE“ ON PDB OPEN AND NON-CDB PLUG-IN

- When plugging an older PDB into a 21c-CDB, the PDB will automatically be upgraded to 21c
- A plugged-in Non-CDB will automatically be converted to a PDB, too
- Opening the PDB will take (much) longer and you will see the following messages in the alert.log file (beside others):

```
[...]  
TESTPDB(3):Starting Upgrade on PDB Open  
[...]  
TESTPDB(3):alter pluggable database application APP$CDB$CATALOG end upgrade  
TESTPDB(3):Completed: alter pluggable database application APP$CDB$CATALOG  
end upgrade  
2021-08-16T22:41:21.815101+02:00
```

14 „REPLAY UPGRADE“ (2)

- "Replay Upgrade" is controlled by two new database properties

```
alter database property set UPGRADE_PDB_ON_OPEN='true'  
Completed: alter database property set UPGRADE_PDB_ON_OPEN='true'  
alter database property set CONVERT_NONCDB_ON_OPEN='true'  
Completed: alter database property set CONVERT_NONCDB_ON_OPEN='true'
```

- After opening the PDB you have to compile all invalid objects in the PDB with the script utl_rp.sql to make all components in DBA_REGISTRY "valid".

Resource Management & Security

MANDATORY USER PROFILE FOR PASSWORD SECURITY

- A „mandatory profile“ can be created in CDB\$ROOT
- This profile will be used for all users
- The profile can contain
 - PASSWORD_VERIFY_FUNCTION
 - PASSWORD_GRACE_TIME
- Parameter MANDATORY_USER_PROFILE has to be set
- Additional profiles for the database users can exist

```
SQL> CREATE MANDATORY PROFILE
      2 C##ALL_USER_PROFILE
      3 LIMIT PASSWORD_VERIFY_FUNCTION
      4 ora12c_stig_verify_function
      5 CONTAINER=ALL;
```

Profile created.

```
SQL> ALTER SYSTEM SET
      2 mandatory_user_profile=
      3 'C##ALL_USER_PROFILE';
```

System altered.

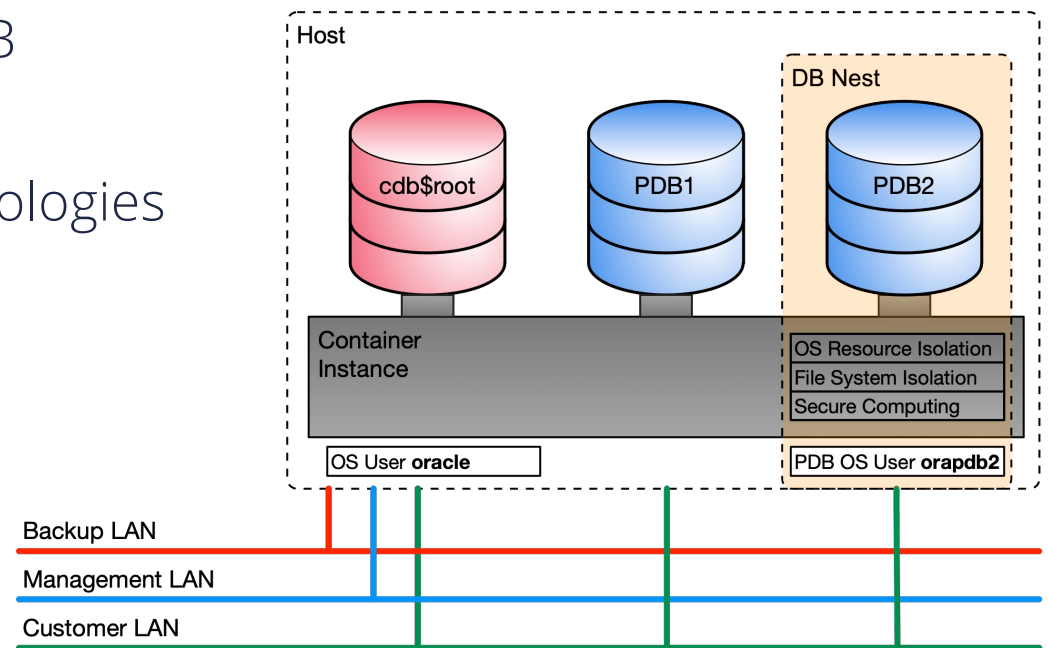
DATABASE RESIDENT CONNECTION POOLING (DRCP) PER PDB

- Using the new parameter `ENABLE_PER_PDB_DRCP` you can decide, whether ..
 - There is one connection pool for the entire CDB (default)
 - There are isolated connection pools for each PDB

```
SQL> alter system set ENABLE_PER_PDB_DRCP=true|false;
```

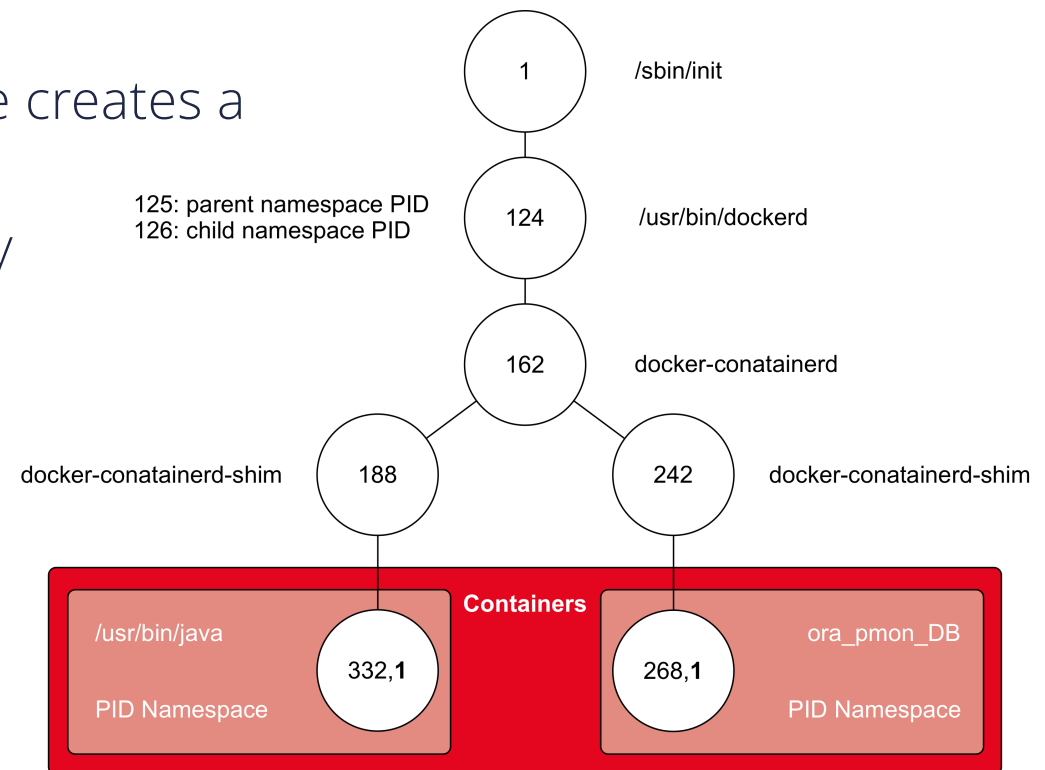
ORACLE DB NEST

- Control and isolation of...
 - ... OS resources used by a PDB
 - ... File system isolation per PDB
 - ... Secure computing
- Concept analog to container technologies like Docker
 - Use of Linux Namespaces
 - Use of CGROUPS
- Available on Linux x86-64bit only



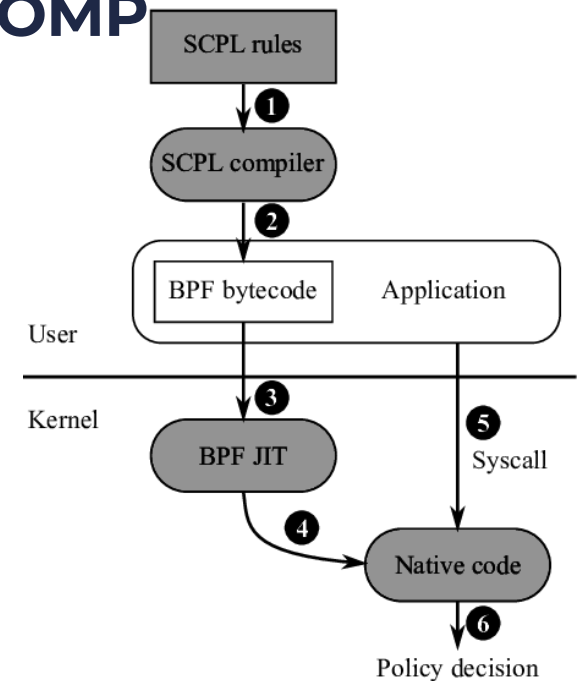
KERNEL NAMESPACES

- Linux kernel function for isolation and virtualization of system resources
- When a DB Nest is launched, Oracle creates a set of namespaces for that DB Nest
- Processes within a DB Nest see only its namespace
- Namespaces Types
 - Process namespace
 - User ID namespace
 - Mount namespace
 - ..



20 SSCOMP – SECURE COMPUTING MODESCOMP SECURE COMPUTING MODE

- Linux kernel feature
- Filter out system calls that are...
 - ... unnecessary
 - ... malicious
- Restrict the actions available within the container
- seccomp uses Berkeley Packet Filters
- Well known / used in Container environments e.g. Docker



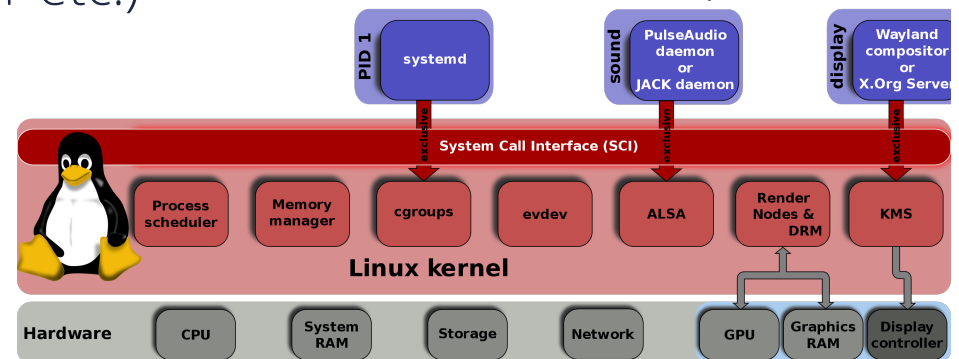
21 CONTROL GROUPS (CGROUPS)

- cgroups is a Linux kernel feature
- mainlined into the Linux kernel since 2007
- Allows to limit that limits, accounts for, and isolates the resource usage of a collection of processes
- Possibility of limiting and isolating the consumption of resources
- Heavily used in Container (runc, Docker etc.)
- CPU, memory, maximum number of PIDs, (network, disk I/O)

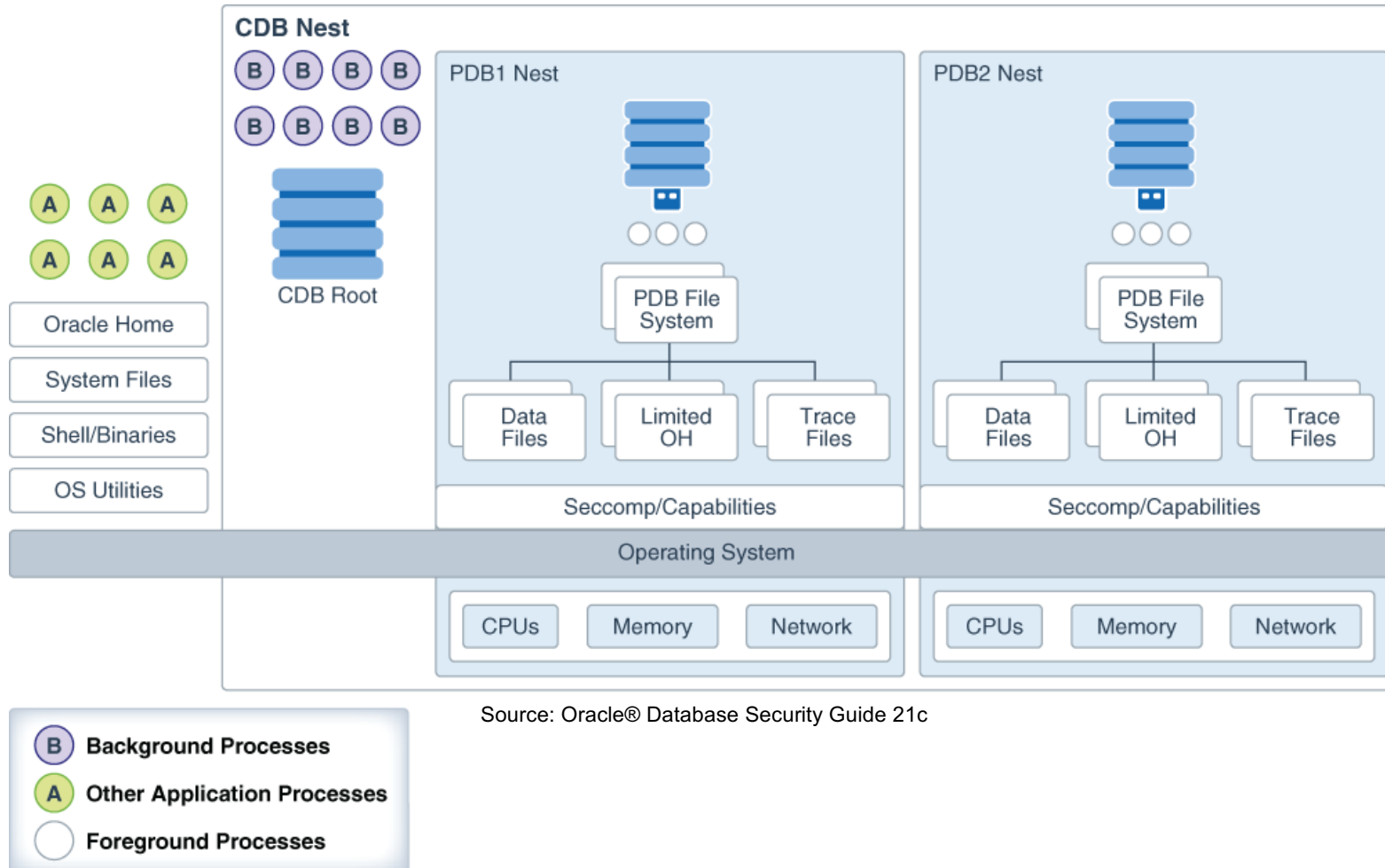
cgroups are responsible for resource management. It makes sense, to grant some daemon exclusive access to this functionality to avoid lots of problems.
systemd-nspawn

If your sound card can do hardware mixing, and your Linux device driver supports this feature, then multiple programs can access your sound card at the same time and you hear them all simultaneously!
PulseAudio daemon does software mixing. Without hardware or software mixing, only one program can access the sound card; as a result, you cannot have Audacious AND VLC put out sound at the same time!
JACK daemon does the same but targets professional audio editors.

DRM manages the GPU
KMS manages the display controller (CRT) The display controller usually sits on the die of the GPU, and communicates with the monitor, e.g. changes the resolution or the refresh rate.
David Herman split DRM and KMS, then added "render nodes" to the DRM.
X.Org doesn't need to be root any longer, but it's still wise (technically necessary?) to grant it exclusive access to the KMS.



22 ARCHITECTURE OF A CDB NEST



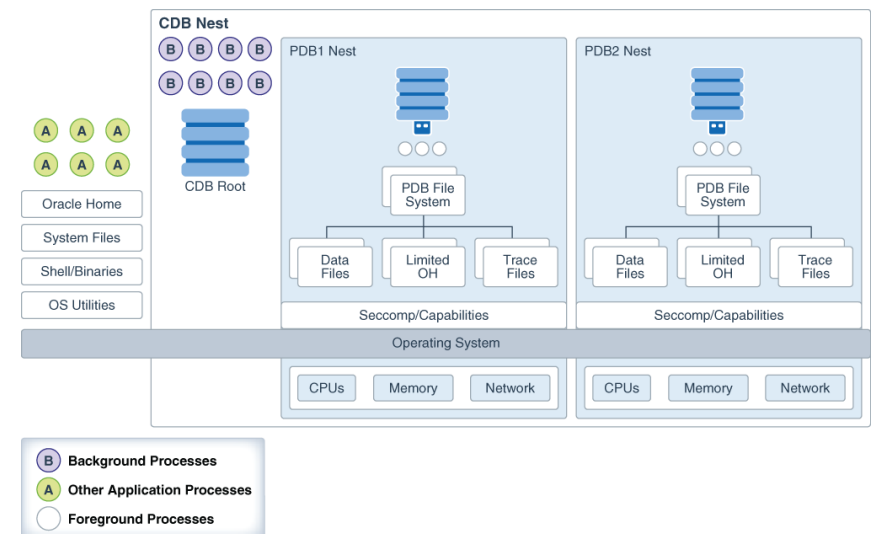
Source: Oracle® Database Security Guide 21c

23 GOAL OF DB NEST

- DB Nest is the Oracle solution for database instance and PDB protection
- Enables a database instance to run in a protected, virtualized environment.
- DB Nest isolate database instance from...
 - ... another database instance
 - ... other applications
 - ... as well as PDBs from each other and from the CDB

24 DB NEST PROPERTIES

- Operating system isolation
 - OS resources like process ID, user, and mount
- File system isolation
 - Visibility for file system entities
 - A **pivot root** in Linux namespaces is equivalent to chroot
 - A **bind mount** enables the contents of one directory to be accessible in a different directory
- Resource management
 - Control and monitor the resources of a nest
- Secure computing mode (seccomp)
 - seccomp to filter out system calls



25 CONFIGURATION OVERVIEW

- Introduction of new init.ora parameter
 - DBNEST_ENABLE – Enables or disables DB Nest
 - DBNEST_PDB_FS_CONF – Specifies the location of an optional file system configuration file. Set this parameter in the CDB root.
- Use of a dedicated broker configured in *listener.ora* by DEDICATED_THROUGH_BROKER_LISTENER
- Introduction of new commandline tools **dbnest** and **dbnestinit**
 - Allows to create, initialize and test DB Nests
- Requires additional OS package
 - nscd – A Name Service Caching Daemon (nscd)
 - sssd – System Security Services Daemon

26 CONFIGURATION

- Configure a dedicated broker in listener.ora

```
DEDICATED_THROUGH_BROKER_LISTENER=ON
```

- Enable the broker

```
ALTER SYSTEM SET use_dedicated_broker=TRUE;
```

- Enable DB Nest and restart the database

```
ALTER SYSTEM SET dbnest_enable=cdb_resource_pdb_all SCOPE=SPFILE;
```

- Check the alert.log for DB Nest

```
Instance running inside DB Nest (TDB200C_TDB200C)
```

```
...
```

```
PDBHR(3):DB Nest (PDB00003, 2968463207) open successful
```

27 CHECK DB NEST (1)

```
oracle@training21c:~/ [TVDCDB1] dbnest list
```

```
-----  
Id : Nest                : Parent                : : Tag                :      State  
-----  
1 : ORA_TVDCDB1_TVDCDB1 :                      : ORA_TVDCDB1_TVDCDB1 :      OPEN  
  Net State              :  
  Namespace State       : (pid=0,cnid=4026531836,pnid=4026531836,no namespace,type=0x0)  
  Resources              : (cpu=0)  
  Property enabled      : resources  
  Seccomp status        : (level=none)  
  FS Isolation          : (disabled)  
-----  
[...]  
-----  
3 : PDB00003            : ORA_TVDCDB1_TVDCDB1 : PDB01 (uid=2042280984) :      OPEN  
  Net State              :  
  Namespace State       : (pid=5791,cnid=4026532249,pnid=4026531836,type=0x7)  
  Resources              : (cpu=0)  
  Property enabled      : namespaces,resources  
  Seccomp status        : (level=strict1)  
  FS Isolation          : (default-config)  
-----
```

28 CHECK DB NEST (2)

- Existing Namespaces with dbnest

```
oracle@training21c:~/ [TVDCDB1] lsns
      NS TYPE      NPROCS      PID USER      COMMAND
4026531835 cgroup      82   1972 oracle /u00/app/oracle/product/21c/bin/tnslsnr LISTENER
4026531836 pid        80   1972 oracle /u00/app/oracle/product/21c/bin/tnslsnr LISTENER
4026531837 user        80   1972 oracle /u00/app/oracle/product/21c/bin/tnslsnr LISTENER
4026531838 uts         82   1972 oracle /u00/app/oracle/product/21c/bin/tnslsnr LISTENER
4026531839 ipc         82   1972 oracle /u00/app/oracle/product/21c/bin/tnslsnr LISTENER
4026531840 mnt          80   1972 oracle /u00/app/oracle/product/21c/bin/tnslsnr LISTENER
4026531992 net          82   1972 oracle /u00/app/oracle/product/21c/bin/tnslsnr LISTENER
4026532247 user         1   5791 oracle dbnestinit PDB00003
4026532248 mnt          1   5791 oracle dbnestinit PDB00003
4026532249 pid          1   5791 oracle dbnestinit PDB00003
4026532250 user         1   6195 oracle dbnestinit PDB00004
4026532251 mnt          1   6195 oracle dbnestinit PDB00004
4026532252 pid          1   6195 oracle dbnestinit PDB00004
```

29 ENTERING DB NESTS

- Use dbnest to enter the namespace of a nest e.g. opening a shell in this namespace

```
oracle@ol7db21:~/ [TDB210C] dbnest enter PDB00001
Entering nest namespace : PDB00001

oracle@ol7db20:~/ [TDB210C] exit
exit
Exiting nest namespace : PDB00001
```

- Sqlplus "/ as sysdba" does not work with Dbnest (use passwordfile)

30 FILE SYSTEM ISOLATION

- Directories can be excluded from a DB nest
- Create a blacklist file

```
vi /u00/app/oracle/nest/nest_blacklist.txt  
  
DBNEST_NO_FS_ROOT_MODE  
/bin  
/usr/bin
```

- Configure the init parameter **DBNEST_PDB_FS_CONF** to point to the blacklist file

```
ALTER SYSTEM SET DBNEST_PDB_FS_CONF=  
'/u00/app/oracle/nest/nest_blacklist.txt' SCOPE=spfile;
```

31 TOOL DBNEST (1)

- New commandline tools to configure / administer Oracle DB Nests
- Currently not yet documented
- Highly try and error to use it

```
oracle@db21:/u00/app/oracle/nest/ [TCPU21C] dbnest -h
Usage : dbnest <command> [options]
```

List of options and commands.

init [options]

 --stage <staging area path>

 --cgroup <base cgroups>

Initialize nest environment

Nest staging area path, Used for storing nest conf files, skeleton directories for nests etc.

Valid base cgroup path if required to override the default path available on the system.

32 NEW TOOLS FOR DBNEST

```
oracle@training21c:~/ [rdbms21] dbnest -h
```

```
[...]
```

```
create <nest> [options]
```

```
--parent <parent_nest>
```

```
--nstype <combination of nstypes>
```

```
--cpu <cpu count>
```

```
--cpuids <list>
```

```
--cpu_excl <1|0>
```

```
--cpu_from_end <1|0>
```

```
--cpushares <cpushares>
```

```
--max_mem <maximum memory>
```

```
--max_swap <maximum memory>
```

```
[...]
```

Create a nest

Parent nest name

Combination of required namespace types. Defaults to user, pid and mount.

Number of cpus required for nest

Comma delimited cpuid range/list

e.g: 0,1-2 1-3,5 2-3 1-3,5-7 2,3

Alloc the CPUs exclusively or not

1 : alloc exclusive

0 : alloc shared (default)

Start CPU allocation from the end

1 : alloc from end

0 : alloc from cpu0 (default)

CPU shares for this nest

Max memory (in MB) for the nest

Max swap (in MB) for the nest

trivadis

Part of **Accenture**

Multitenant & Oracle GridInfrastructure/RAC

PDB AS A CLUSTER RESOURCE (1)

- PDBs are now cluster resources
- Example (Oracle Restart):

```
grid@fenrir:~/ [grinf21] crsctl stat res -t -w "TYPE = ora.pdb.type"
```

Name	Target	State	Server	State details
------	--------	-------	--------	---------------

Cluster Resources

ora.remcdb.pdb1.pdb

1	ONLINE	ONLINE	fenrir	STABLE
---	--------	--------	--------	--------

ora.remcdb.pdb2a.pdb

1	ONLINE	ONLINE	fenrir	STABLE
---	--------	--------	--------	--------

PDB AS A CLUSTER RESOURCE (2)

- There are new commands within the well-known tool `srvctl`
 - `srvctl add pdb`
 - `srvctl modify pdb`
 - `srvctl config pdb`
 - `srvctl modify pdb`
 - `srvctl start pdb`
 - `srvctl status pdb`
 - `srvctl stop pdb`
 - `srvctl enable pdb`
 - `srvctl disable pdb`
- New PDBs are automatically added as cluster resources

PDB AS A CLUSTER RESOURCE (3) - EXAMPLE

```
SQL> show pdbs
```

CON_ID	CON_NAME	OPEN MODE	RESTRICTED
2	PDB\$SEED	READ ONLY	NO
3	PDB1	READ WRITE	NO
4	PDB2A	READ WRITE	NO

```
SQL> !srvctl stop pdb -db REMCDB -pdb PDB1
```

```
SQL> show pdbs
```

CON_ID	CON_NAME	OPEN MODE	RESTRICTED
2	PDB\$SEED	READ ONLY	NO
3	PDB1	MOUNTED	
4	PDB2A	READ WRITE	NO

Multitenant & DataGuard

PDB SIDE RECOVERY (1)

- Until 19c
 - MRP got stuck when there was a hot-cloned, flashbacked or recovered (PITR) PDB
 - Reason: an recovery process was running but there was already MRP running in the background
 - Result: **ORA-1153 an incompatible media recovery is active**

PDB SIDE RECOVERY (2)

- In 21c (Active DataGuard):
 1. PDB on standby is automatically marked as "disabled"
 2. PDB is recovered in a separate session ("PDB side recovery", "PDB recovery isolation")
 3. PDB is re-enabled again
- In 21c (without ADG):
 - PDB files are copied
 - PDB ist marked "DISABLED AUTOMATIC RECOVERY" on Standby
 - MRP does not stop
 - DBA-Action:
 - Recover PDB from service (Primary)
 - Enable Recovery for PDB

PDB SIDE RECOVERY (3) - EXAMPLE

- On Primary

```
SQL> create pluggable database TESTPDB from DGPDB
      2 file_name_convert=('DGPDB','TESTPDB');
Pluggable database created.
```

- On Standby (ADG)

```
SQL> show pdbs
      CON_ID CON_NAME                                OPEN MODE  RESTRICTED
-----
          2 PDB$SEED                                READ ONLY  NO
          3 DGPDB                                    MOUNTED
          4 TESTPDB                                    MOUNTED

SQL> select process,status,sequence# from v$managed_standby where
      2 process like 'MRP%';
PROCESS      STATUS SEQUENCE#
-----
MRP0         APPLYING_LOG                20
```


PDB SIDE RECOVERY (4) – ALERT.LOG - STANDBY

```
TESTPDB(4):Datafile #11 has been copied to the standby.  
2021-11-10T21:17:31.873127+01:00  
TESTPDB(4):PDB Side Media Recovery started for pdbid(4)  
TESTPDB(4):..... (PID:7306): Managed Recovery starting Real Time Apply [krsm.c:15901]  
TESTPDB(4):max_pdb is 4  
TESTPDB(4):..... (PID:7306): Media Recovery Waiting for T-1.S-20 (in transit)  
[krsm.c:6191]  
2021-11-10T21:17:31.949889+01:00  
  
[...]  
  
TESTPDB(4):..... (PID:7306): Side Recovery Complete [krds.c:1584]  
2021-11-10T21:17:53.329342+01:00  
all data files of pdbid 4 are brought online.  
Started logmerger process  
2021-11-10T21:17:53.361688+01:00
```

PDB SIDE RECOVERY (5) - EXAMPLE

- Without ADG (after creating a PDB as r/w-clone):

```
SQL> select name,RECOVERY_STATUS from v$pdb;
```

NAME	RECOVERY_STATUS
PDB\$SEED	ENABLED
DGPDB	ENABLED
TESTPDB	ENABLED
TESTPDB2	DISABLED AUTOMATIC RECOVER

```
SQL> select process,status,sequence# from v$managed_standby where  
process like 'MRP%';
```

PROCESS	STATUS	SEQUENCE#
MRP0	APPLYING_LOG	23

Miscellaneous

PDB POINT-IN-TIME RECOVERY

- In 21c a PDB can be recovered "to any time in the recent past"
- Flashback & Redolog-data must be available
- No Resetlogs on CDB-level
- Examples:
 - "Flashback" to a point in time before a PDB resetlogs operation
 - Point-In-Time-Recovery to a point in time before a PDB resetlogs operation

TIMEZONE SUPPORT IN DBCA

- A PDB can run in a different timezone than CDB\$ROOT (since Oracle Database 12c)

```
SQL> alter database set timezone='Europe/Berlin';
```

- 21c: you can configure the timezone of a PDB with dbca

```
oracle@tvd21c:~/ dbca -createPluggableDatabase -help
  -createPluggableDatabase - Command to Create a pluggable database.
    -pdbName <Pluggable database name>
[...]
```

**[-pdbTimezone <Specify PDB specific timezone offset from UTC or
timezone region. +HH:MM | -HH:MM | Region >]**

```
[...]
```

- Similar with "-configurePluggableDatabase"

Conclusion

47 CONCLUSION

- "Replay Upgrade" makes upgrade and migration to Multitenant "more smooth" 👍
- Oracle DB Nest is ...
 - a meaningful approach
 - definitely still a little shaky
 - highly depend on the OS
 - poor documentation
- PDB as cluster resource 👍
- PDB side recovery 👍
- and there's more to come ("PDB level DataGuard")

48 REFERENCES & MORE INFORMATION

- Oracle Database 21c Documentation
 - "Learning Database 21c New Features"
 - Security Guide – Ch. 15: Securing and Isolating Resources Using DbNest
- MOS-Note "PDB Side Recovery (Doc ID 2649208.1)"
- DB Nest
 - <https://mahmoudhatem.wordpress.com/2020/10/13/oracle-20c-dbnest-linux-namespaces-seccomp-capabilites-cgroups/>
 - <https://mahmoudhatem.wordpress.com/2020/10/20/a-first-hands-on-oracle-20c-dbnest-preview/>
 - .. And many more blog posts by Hatem Mahmoud
- Information on Linux Namespaces:
<http://ifeanyi.co/posts/linux-namespaces-part-1/>

QUESTIONS & ANSWERS

Kudos to Stefan Oehrli (@oradbach) and Christian Gohmann (@cgohmannde) for their support when preparing this presentation



MARKUS FLECHTNER

- Markus.flechtner@trivadis.com
- Twitter @markusdba
- Blog: markusdba.net



trivadis Part of Accenture

**TOGETHER WE ARE
#1 PARTNER FOR BUSINESSES TO
HARNESS THE POWER OF DATA
FOR A SMARTER LIFE**

trivadis
Part of Accenture