

ORACLE

Oracle 数据库19c管理之 Data Pump

19^c

Ling Mei
SEHub Data Management China Team
2021-07

Oracle 免责声明

以下内容旨在概述 Oracle 产品的总体发展方向。该内容仅供参考，不可纳入任何合同。本文档不承诺提供任何材料、代码或功能，也不应将其作为购买决策的依据。

此处所述有关 Oracle 产品任何特性或功能的开发、发布、时间安排和定价可能会发生变更，且均由甲骨文公司自行决定。



议程

- Oracle Data Pump介绍
- DB 19c中Data Pump新特性
- Data Pump 最佳实践与性能优化
- Data Pump 用户案例
- Demo演示

19^c

Oracle Data Pump概述

- 基于服务器端的高速数据/元数据移动工具
- Oracle数据泵由三个不同的组件组成:
 - 命令行客户端、expdp和impdp
 - DBMS_DATAPUMP PL/SQL包(也称为数据泵API)
 - DBMS_METADATA PL/SQL包(也称为元数据API)
- 提供五种数据迁移方式
 - Data file Copying
 - Convetntional
 - Direct Path
 - External Tables
 - Network link support

Oracle 逻辑备份与恢复

Data Pump

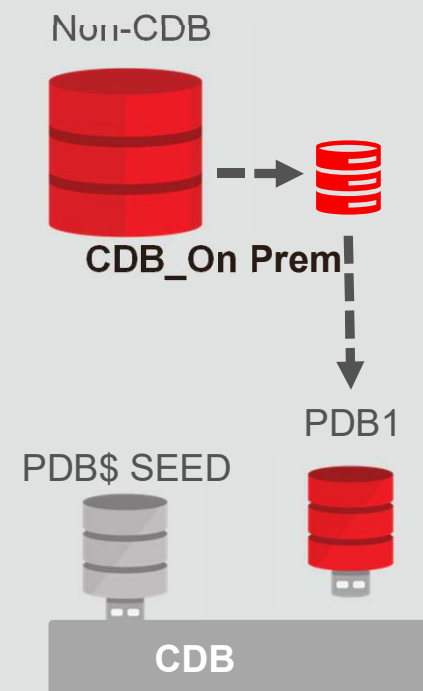
使用场景

- ✓ 不同字节序
- ✓ 源库版本为10g及其以上
- ✓ 不同数据库版本
- ✓ 从非CDB到PDB
- ✓ 可能更改数据库架构

注: 9i及其以前版本的源库要使用 exp/imp

迁移过程:

- 准备好新的目标PDB
- 使用expdp导出数据
- 复制 dumpfiles 到目标端
- 用impdp导入到目标库



使用Data Pump进行数据迁移场景1

通过Data Pump将数据迁移到Oracle DB 19c:

Step 1: 从源库导出数据;

Step 2: 安装DB 19c并创建目标数据库;

Step 3: 将数据导入到新的19c 数据库中。



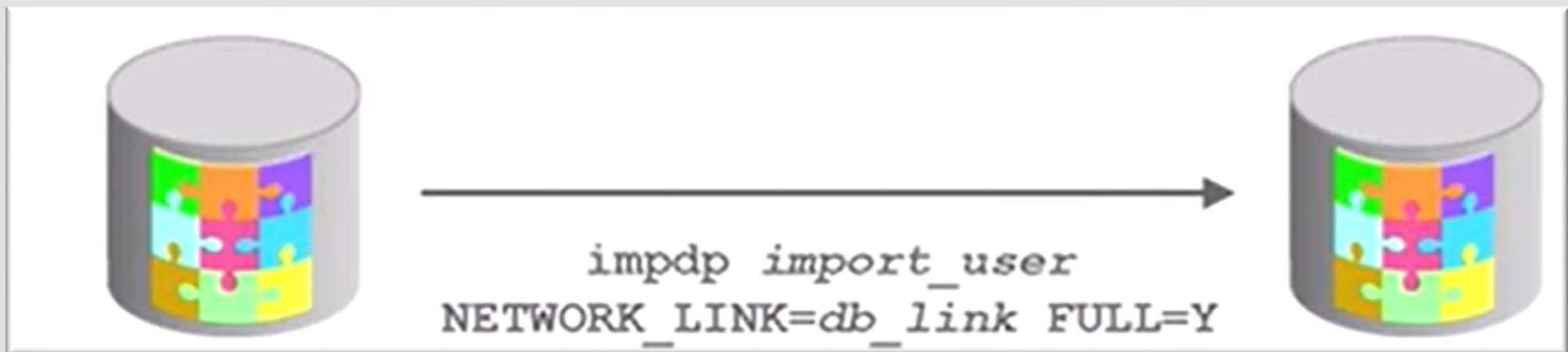
源数据库
11g/12c/18c/19c

目标数据库
19c

使用Data Pump进行数据迁移场景2

通过Network Link

通过Data Pump进行全库导入时，可以通过database link进行跨库导入。
用户在进行导出操作时需要有DATAPUMP_EXP_FULL_DATABASE role;
用户在进行导入操作时需要有DATAPUMP_IMP_FULL_DATABASE role。



源数据库
11g/12c/18c/19c

目标数据库
19c

数据泵与RMAN相关特性对比

特性	RMAN	Data Pump
脱机备份	支持。实例需要到Mount状态	不支持
联机备份	支持	支持
增量备份	支持	不支持
坏块检测	支持，识别坏块并记录在 V\$database_block_corruption	支持。 在导出日志中标识损坏的块
备份资料库	支持。	不支持
备份到介质管理器	支持	不支持
备份初始化参数文件	支持	不支持
备份密码文件以及网络相关文件	不支持	不支持

议程

- Oracle Data Pump概述
- DB 19c中Data Pump新特性
- Data Pump 最佳实践与性能优化
- Data Pump User Case
- Demo演示

19^c

DB 19c中Data Pump新特性

- ◆ 支持云对象存储作为数据加载的存储目标
- ◆ 可移动表空间的测试模式
- ◆ 允许表空间在TTS导入期间保持只读状态
- ◆ 支持资源使用限制
- ◆ 在导入操作时排除ENCRYPTION子句
- ◆ 支持对象存储文件的通配符

Data Pump | New in Oracle 19c



支持云对象存储作为数据加载的存储目标

- 该特性将REST客户端API支持添加到Oracle数据库，支持客户端API从对象存储区读取Datapump转储文件
- 使用expdp导出时直接将dump文件加载到对象存储上; 使用impdp导入时从对象存储直接将数据加载到数据库中。
- 引入该特性的好处
 - 在受管理的云服务中（如EECS和DWCS）用户不能访问操作系统，只能访问数据库服务。
 - 加载数据并将客户从本地迁移到云服务是一项艰巨的任务。如果能够直接访问对象存储中的文件，客户可以将他们的Data pump转储文件上传到对象存储中，方便使用新的对象存储REST API将它们加载到数据库中。

使用该特性

Impdp客户端CLI现在接受一个名为“credential”的新命令行参数。

该参数为访问对象存储桶所需的用户名和密码的凭据对象的名称。

当用户连接到数据库来运行impdp时，Data pump层将验证凭据是否存在，并且用户是否有权读取凭据。

如果有任何错误，则向impdp客户端返回一个错误。

凭据验证是通过查询SYS.USER_CREDENTIALS表。

示例:

```
impdp system/manager table_exists_action=replace credential = TEST  
dumpfile=https://usoracleinterxxxx.storage.oraclecloud.com/v1/odbstest-  
usoracleinterxxxx/dnfs/dump.dat trace=1800300 logfile=work_import.log
```



Data Pump | New in Oracle 19c

支持云对象存储作为数据加载的存储目标

- PDB引入了一个名为default_credential的新PDB属性，可以为指定的PDB设置此属性。
- 如果impdp 命令行没有指定身份认证凭证名，Data pump和对象存储模块可以使用default_credential对象检索用户名和密码，以访问对象存储桶。
- 当使用默认身份凭据运行impdp时，credential参数不需要指定。

Example:

```
SQL> alter database property set default_credential = 'SYSTEM.OPCTEST';
```

```
impdp system/manager table_exists_action=replace logfile=work_import4.log  
dumpfile=DEFAULT_CREDENTIAL:https://usoracleinterxxxx.storage.oraclecloud.com/v1/odbstest-  
usoracleinterxxxx/dnfs/testdump.dat
```

Data Pump | New in Oracle 19c

支持云对象存储作为数据加载的存储目标

相关注意点

- 新的impdp 参数: *credential*
- 识别新的凭据对象。云对象存储的凭据将使用凭据对象进行管理。可以使用DBMS_CREDENTIAL创建凭据对象。CREATE_CREDENTIAL过程。此过程接受用户名和密码。数据被加密保存在PDB字典中，拥有访问权限的用户和PDB的用户才能查询凭据。

[MOS Note 2457398.1](#)

示例:

```
BEGIN
  DBMS_CREDENTIAL.CREATE_CREDENTIAL(
    credential_name => 'TEST',
    username        => '*****',
    password        => '*****');
END;
```

Data Pump | New in Oracle 19c

可移动表空间的测试模式

- 可迁移表空间的测试模式支持下面两种场景下的只导出元数据的测试。
 - 可迁移表空间 (Transportable tablespace)
 - 完全可迁移(Full Transportable)
- 修复Bug 21471988-FORCE OPTION for DATA PUMP来支持TTS和全传输的测试。

DataPump通过这个特性为可传输操作提供了一个测试模式；该模式下不需要将所有表空间设置为只读，就可以执行元数据导出，以获得如下感知：

- a) 需要多长时间
- b) 是否存在关闭检查未发现的不可预见的问题；
- 解决的问题

消除了源数据库表空间为只读模式的限制；DataPump要求在导出可传输表空间操作之前将表空间设置为只读。将支持当前隐藏的仅用于API的DataPump参数TTS_CLOSURE_CHECK作为命令行参数。

将表空间设置为只读可确保数据处于一致状态；此外可传输表空间导出操作将对可传输表空间集进行闭包检查。

实际上有些客户的表空间在只读模式下的时间是有限的。

这些客户要求能够限制表空间处于只读模式的时间长度，并能够获得datapump可传输导出操作的时间估计。

Data Pump | New in Oracle 19c

可移动表空间的测试模式

如何使用该特性

可以在data pump expdp命令行通过设置以下参数值，来指定要作为可传输导出操作的一部分执行的闭包检查级别。

语法和描述

TTS_CLOSURE_CHECK = [ON | OFF | FULL | **TEST_MODE**]

ON – 指示执行闭包检查，以确保可传输表空间集不包含集合外的引用。

OFF – 指示执行无闭包检查。用户负责验证可传输表空间集的包容性。

FULL –指示执行完整的多向闭包检查 (仍然支持TTS_FULL_CHECK和TRANSPORT_FULL_CHECK参数)，
以确保可传输表空间集中没有远程引用

TEST_MODE –表示表空间不需要处于只读模式。

此选项仅用于测试目的;提供可传输表空间导出操作的计时要求。

数据泵导入无法使用生成的导出转储文件。

注意: 参数值ON, OFF, FULL 是互斥的。 TEST_MODE 仅仅是expdp的一个选项。

Data Pump | New in Oracle 19c

可移动表空间的测试模式

- 可以使用现有过程DBMS_DATAPUMP.SET_PARAMETER设置TTS_CLOSURE_CHECK参数值。
TTS_CLOSURE_CHECK选项使用数字位掩码值设置:

```
DBMS_DATAPUMP.KU$_TTS_CLOSURE_CHECK_ON is the value 1  
DBMS_DATAPUMP.KU$_TTS_CLOSURE_CHECK_OFF is the value 2  
DBMS_DATAPUMP.KU$_TTS_CLOSURE_CHECK_FULL is the value 4  
DBMS_DATAPUMP.KU$_TTS_CLOSURE_CHECK_TEST is the value 8
```

下面的命令禁用所有闭包检查并启用测试模式:

```
SYS.DBMS_DATAPUMP.SET_PARAMETER(jobhdl, 'TTS_CLOSURE_CHECK',  
DBMS_DATAPUMP.KU$_TTS_CLOSURE_CHECK_OFF+  
DBMS_DATAPUMP.KU$_TTS_CLOSURE_CHECK_TEST);
```

DataPump执行闭包检查的时间可能很长，有时是不必要的，特别是当客户知道包含可传输表空间集时。使用TTS_CLOSURE_CHECK OFF选项运行数据泵可传输操作将减少数据泵可传输操作完成所需的时间。

Data Pump | New in Oracle 19c

可移动表空间的测试模式

关于错误ORA-39050: 不兼容的参数异常

有两个现有的DataPump命令行参数用来控制多向闭包检查:

- `TTS_FULL_CHECK=[YES | NO]`
YES – indicates full multi-directional closure check
NO – indicates containment closure check
- `TRANSPORT_FULL_CHECK=[YES | NO]`
YES – indicates full multi-directional closure check
NO – indicates containment closure check

这两个参数在19C中仍继续支持, 但和TTS_CLOSURE_CHECK对应值使用是冲突的(报告ORA-39050)。

`TTS_FULL_CHECK=YES` or `TRANSPORT_FULL_CHECK=YES` conflicts with `TTS_CLOSURE_CHECK=ON|OFF`

`TTS_FULL_CHECK=NO` or `TRANSPORT_FULL_CHECK=NO` conflicts with `TTS_CLOSURE_CHECK=FULL`

Data Pump | New in Oracle 19c

允许表空间在TTS导入期间保持只读状态



TRANSPORTABLE=NEVER | ALWAYS | **KEEP_READ_ONLY** | NO_BITMAP_REBUILD

重复执行TTS操作，而无需再次还原文件

在多个数据库之间共享表空间

示例:

```
impdp system DIRECTORY=dpump_dir DUMPFILE=dumpfile_name  
TRANSPORT_DATAFILES=datafile_name TRANSPORTABLE=KEEP_READ_ONLY  
表空间已安装并且可用于只读访问
```



Data Pump | New in Oracle 19c

允许表空间在TTS导入期间保持只读状态

该特性通过在DataPump导入命令行上的TRANSPORTABLE参数添加新选项实现。

impdp的TRANSPORTABLE参数有两个新选项:

TRANSPORTABLE = [NEVER | ALWAYS | **KEEP_READ_ONLY** | **NO_BITMAP_REBUILD**]

➤ **KEEP_READ_ONLY**

表空间在导入更新期间和导入完成后保持只读状态;

可传输数据文件将保持只读模式;

该模式下允许将数据文件插入到其他兼容给的数据库中进行只读访问。

➤ **NO_BITMAP_REBUILD**

允许表空间变为读/写状态;

不为表空间重新生成空闲空间位图;

不回收未使用的数据段, 减少导入操作时间;

可以使用dbms_space_admin.tablespace_rebuild_bitmaps重新构建位图映射。

Data Pump | New in Oracle 19c

允许表空间在TTS导入期间保持只读状态

适用场景:

- 部分用户希望在TTS导入期间保持表空间只读的能力，因为这允许他们在两个不同的数据库上挂载相同的表空间数据文件，需要这些文件保持只读状态。
- 某些大型表空间重新构建空闲空间所花费的时间可能较长，尤其是对服务级别有严格要求的客户。或者如果暂时不需要重新计算位图。添加不重新生成位图的选项将减少大规模迁移的停机时间。

注意事项:

- 当TRANSPORTABLE=KEEP_READ_ONLY指定时，由于数据文件处于只读模式，包含TSTZ列数据的表无法更新，将从导入中删除。此外，数据文件位图无法重建。
- 当TRANSPORTABLE=NO_BITMAP_REBUILD被指定时，如果有必要数据文件可能被置于读写模式，以便更新包含TSTZ列数据的表。

Data Pump | New in Oracle 19c

支持资源使用限制

这个特性引入了两个新的数据库参数：

- MAX_DATAPUMP_JOBS_PER_PDB
在容器多租户数据库环境中，PDB并行数据泵的最大数目
- MAX_DATAPUMP_PARALLEL_PER_JOB
每个Oracle数据泵作业允许的并行进程的最大数目

该特性也可以在非多租户环境中使用。

- 对在DataPump中实现资源使用策略的改进;
- 解决了在PDB中无法杀死DataPumpworker slave的问题(Reference Bug 27457666 - UNABLE TO KILL DATA PUMP WORKER SESSIONS).
- 当多个用户可能在给定的数据库环境中执行DataPump作业时，这些参数使DBA能够更好地控制资源的使用。

Data Pump | New in Oracle 19c

在导入操作时的ENCRYPTION子句



在导入操作时启用列加密

TRANSFORM=OMIT_ENCRYPTION_CLAUSE:[Y|N]

适用场景：引入此新的转换参数是为了方便从具有TDE的任何源数据库迁移到不具有TDE的目标数据库。

例如当源数据库有加密的列时，迁移到没有启用TDE的目标库

示例:

```
impdp hr DIRECTORY=dpump_dir1 DUMPFILE=hr.dmp SCHEMAS=hr  
TRANSFORM=OMIT_ENCRYPTION_CLAUSE:Y
```

描述:

该特性适用与表/表空间/物化视图对象类型;

可用于任何不可传输 (non-transportable import) 的导入作业类型

对应值含义如下

Y: 源库中的加密列不会在导入的表中加密

N: 列加密子句和源库中的创建代码一致



Data Pump | New in Oracle 19c

在导入操作时的ENCRYPTION子句

在导入时使用排除加密子句。

该特性仅允许用户在DataPump生成create object ddl时抑制加密ddl，没有影响或问题。

当要迁移的对象涉及加密的列时，它使从TDE环境迁移到非TDE环境变得更容易。

在现有的命令行参数TRANSFORM指定该值:

TRANSFORM=OMIT_ENCRYPTION_CLAUSE: [Y | N]

N -生成的create object ddl将包括相关的加密属性子句(默认)。

Y -生成的create object ddl将不包括任何加密属性子句。

这个转换参数的值也可以通过DBMS包的对应接口进行设置: DBMS_METADATA.SET_TRANSFORM_PARAM.

Example:

```
DBMS_METADATA.SET_TRANSFORM_PARAM(job_handle, 'OMIT_ENCRYPTION_CLAUSE', true);
```

Data Pump | New in Oracle 19c

支持对象存储中的dumpfile通配符

适用场景:当需要从对象存储服务中导入多个转储文件时, 可以使用基于url的转储文件名通配符简化Oracle自治数据库的导入命令, 简化导入多个转储文件。

- 可以在URL的文件名中指定通配符
- 不支持bucket-name组件中的通配符。

```
> impdp admin/password@ATPC1_high  
    directory=data_pump_dir credential=my_cred_name  
    dumpfile= https://objectstorage.example.com/v1/atpc/atpc_user/exp%u.dmp"
```

Data Pump | New in Oracle 21c

- 通过Checksum 校验 Dumpfile
expdp: CHECKSUM, CHECKSUM_ALGORITHM
impdp: VERIFY_CHECKSUM, VERIFY_ONLY
- 在同一个导出或导入操作中包含和排除对象
expdp include=table exclude=statistics
- 在导入时支持可选的索引压缩, 包括Oracle自治数据库
TRANSFORM=INDEX_COMPRESSION_CLAUSE:COMPRESS ADVANCED LOW
- 支持从Oracle自治数据库导出dump 文件到云对象存储
- 支持TTS (可传输表空间) 元数据并行导出
- 可恢复已停止的TTS表空间的导出和导入作业
- 导出和导入操作支持原生JSON数据类型

议程

- Oracle Data Pump概述
- DB 19c中Data Pump新特性
- **Data Pump 最佳实践与性能优化**
- Data Pump User Case
- Demo演示

19^c

Data Pump | 最佳实践 #1

始终使用par 参数文件

听起来奇怪，但是许多问题都是因为人们在命令行输入很长的Data Pump命令出错。

例:

```
DIRECTORY=my_data_pump_dir
DUMPFILE=dumpfile.dmp
LOGFILE=logfile.log
SCHEMAS=HR
EXCLUDE=STATISTICS
LOGTIME=ALL
METRIC=YES
FLASHBACK_TIME=SYSTIMESTAMP
expdp parfile=my_data_pump_parfile.par
```

Data Pump | 最佳实践 #2

一致的数据泵导出（数据一致性）：

- `FLASHBACK_SCN=<scn>`
- `FLASHBACK_TIME=SYSTIMESTAMP`

这两个参数不能同时使用；

如果使用了`Network_link`,`SCN`参数值为源库的对应`SCN`值；

`Flashback_scn`仅适用于数据库的闪回查询，不适用于闪回数据库，闪回Drop以及闪回数据归档功能；

对于不支持大`SCN`的版本，不能为网络导出或网络导入指定大`SCN`。

或者：

- `CONSISTENT=Y` [since Oracle 11.2 – Legacy Interface]

实际导出时取决于当前时间并使用`FLASHBACK_TIME`；

这会增加导出数据期间的`UNDO`量；

Data Pump | 最佳实践 #3

➤ 在导出过程中始终使用:

EXCLUDE=STATISTICS

重新创建新的统计信息或使用DBMS_STATS staging table进行统计信息迁移

LOGTIME=ALL

从 Oracle Database 12.1开始

METRICS=YES

对象数和时间记录在数据泵日志文件中

➤ 在导入过程中始终使用:

LOGTIME=ALL

从 Oracle Database 12.1开始

METRICS=YES

Data Pump | 最佳实践 #4

并行

➤ 使用 PARALLEL=n

11g之后的企业版export/import均可设置;

通常, $n = 2x < \text{CPU核心数} >$, 但可能需要调整

注意在RAC环境使用并行时, Directory需要指向所有实例都能访问的共享存储

Trigger/View/Sequence/Constraint/Object_grant/REF_constraint导出时不能并行

PARALLEL还定义了并行创建多少索引

11.2.0.4 and 12.1.0.2: apply patch for bug [22273229](#) to enable parallel import of constraints/indexes

➤ 12.2新特性: 并行Export/Import 元数据

但不适用于Transportable Tablespaces, Full Transportable Export/Import 能及通过 NETWORK_LINK

Data Pump | 最佳实践 #5 和 #6

- 将STREAMS_POOL_SIZE设置为合理的值(64MB ←… → 256MB)
- 当服务器位于不同的操作系统和存储/ SAN上时, 请使用NETWORK_LINK
 - 在目标数据库上通过database link启动impdp
 - 不会生成dump file
 - 从Oracle 12.2开始: 支持通过dblink进行直接路径加载 (包括 LONG 数据)
 - 使用: ACCESS_METHOD=DIRECT_PATH
 - 限制:
 - 不支持并行导元数据
 - 在Oracle 12.2之前不支持LONG和LAW格式
 - 数据并行性仅限于多个分区或表; 在一个大的非分区表上没有PQ 并行

Data Pump | 最佳实践 #7

- BasicFile (old) LOBs 总是很慢
- SecureFile LOBs 快一些
 - 特别是与分区结合使用
 - 使用DBMS_REDEFINITION或ONLINE TABLE MOVE (18c) 将Basic转换为SecureFile LOB
 - 在导入操作时, 使用DBMS_DATAPUMP或者DBMS_METADATA的LOB_STORAGE参数, 强制将旧的LOB转换为SecureFiles

Data Pump | 最佳实践 #8

➤ 关于数据库字符集

建议源库和目标库字符集一致

隐式字符集转换以目标库字符集为准，可能会导致报错

➤ 关于Version，尤其是在数据库升级降级时

注意参考数据库的Compatible参数设置

设置version参数时要慎重

Data Pump性能优化

- 不建议使用传统的IMP/EXP
- 和Data Pump性能相关的参数设置
 - DISK_ASYNC_IO=TRUE
 - DB_BLOCK_CHECKING=FALSE
 - DB_BLOCK_CHECKSUM=FALSE
- 相关需要调整的初始化参数
 - Processes
 - Sessions
 - Parallel_max_servers
 - Shared_pool_size
 - Undo_tablespace
- 多个用户Data Pump作业的资源管理调度
 - Max_datapump_jobs_per_pdb
 - Value:0-250,default 100
 - Auto: 50% of the sessions value
 - Max_datapump_parallel_per_job
 - Value: 0-1024,default 1024
 - Auto: 25% of the sessions value

议程

- Oracle Data Pump概述
- DB 19c中Data Pump新特性
- Data Pump 最佳实践与性能优化
- **Data Pump User Case**
- Demo演示

19^c

Data Pump 用户场景介绍

案例1: ADW Instance Migration by Data Pump

背景: 客户需要将当前用户报表系统所使用ADW实例迁移至另外一个租户的准生产环境。数据迁移验证后, 准生产切换为新的生产系统。

环境: ADW on Oracle OCI (DB 19c)

Storage: Object Storage on OCI

数据迁移方式: Data pump

数据量: 5T

迁移时间窗口: 一周

关键词: Parallel

object storage

通配符

```
Expdp username/password@adwsource filesize 10000MB  
dumpfile default_credential:https://objectstorage.regionnam  
e.oraclecloud.com/xxxxxxx/dmpfile/o/exp2019%L.dmp  
parallel xx encryption_pwd_prompt yes logfile exp2019.log  
directoty datapumpdir ....
```

```
Impdp username/password@adwtarget  
directory=datapumpdir  
dumpfile=default_credential:https://objectstorage.regionnam  
e.oraclecloud.com/xxxxxxx/dmpfile/o/exp2019%L.dmp  
parallel=xx  
Exclude=xxxxxx
```

.....

Demo演示

1 可移动表空间的测试模式 (Test Mode for Transportable Tablespaces)

TTS_CLOSURE_CHECK=test_mode

2 允许表空间在TTS导入期间保持只读状态 (Allow tablespaces to stay Read-Only during TTS import)

TRANSPORTABLE=KEEP_READ_ONLY

TRANSPORTABLE=NO_BITMAP_REBUILD

Q & A
Thank you

