Data Pump Masterclass

MakeIT, May 2025

Photo by Hasmik Ghazaryan Olson on Unsplash

Oracle

DBAs

run the world

DANIEL OVERBY HANSEN
Distinguished Product Manager
Database Upgrade, Migrations & Patching

dohdatabase

@dohdatabase.com

https://dohdatabase.com

Get the Slides

https://dohdatabase.com/slides

Find Slides and Much More on Our Blogs

MikeDietrichDE.com

Mike.Dietrich@oracle.com

dohdatabase.com

Daniel.Overby.Hansen@oracle.com

DBArj.com.br

Rodrigo.R.Jorge@oracle.com

AlexZaballa.com

Alex.Zaballa@oracle.com

Web Seminar	Slides
Episode 16 (replaces Episode 1 from Feb 2021) Oracle Database Release and Patching Strategy for 19c and 23c *115 minutes – May 10, 2023*	
Episode 17 From SR to Patch – Insights into the Oracle Database Development process *55 minutes – June 22, 2023*	
NEW Episode 18 Cross Platform Migration – Transportable Tablespaces to the Extreme *145 min – February 22, 2024*	
Episode 2 AutoUpgrade to Oracle Database 19c *115 minutes – Feb 20, 2021*	
Episode 3 Performance Stability, Tips and Tricks and Underscores *120 minutes – Mar 4, 2021*	
Episode 4 Migration to Oracle Multitenant	

Recorded Web Seminars

https://MikeDietrichDE.com/videos

More than 35 hours of technical content, on-demand, anytime, anywhere

AGENDA

14:30
Welcome
Presentation
Getting Started

15:30
Lab intro
Hands-on lab

17:15
Presentation
Top Tips

Getting Started

Data Pump Architecture

Dump File

Imported into database

Exported to dump file

Copied over the network

Network Link

Start import and
fetch data directly

Comparison

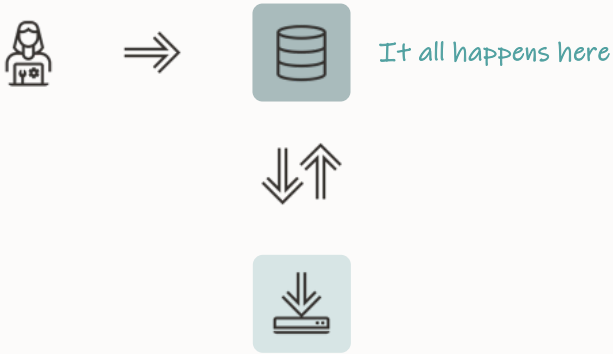
DUMP FILE	NETWORK
Requires access to file system	SQL*Net connectivity
Requires disk space for dump files	No extra disk space needed
Full functionality	Restricted functionality

Pro tip: Read more about how Data Pump moves data

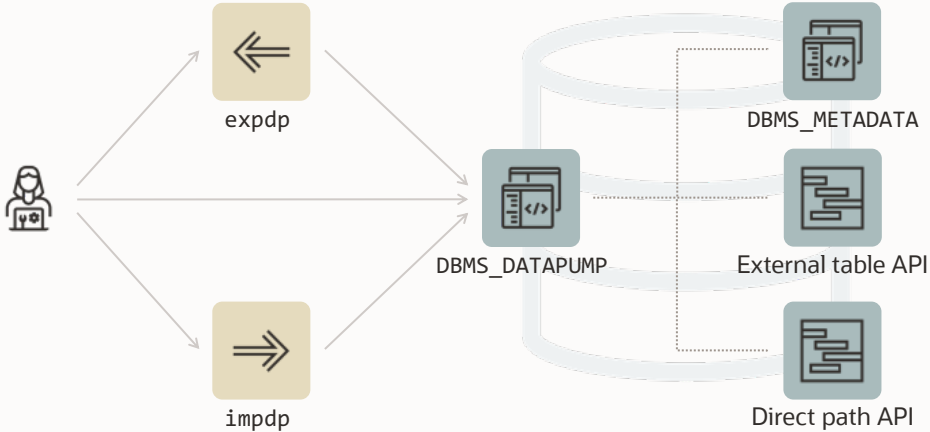
Architecture

Data Pump is server-based,

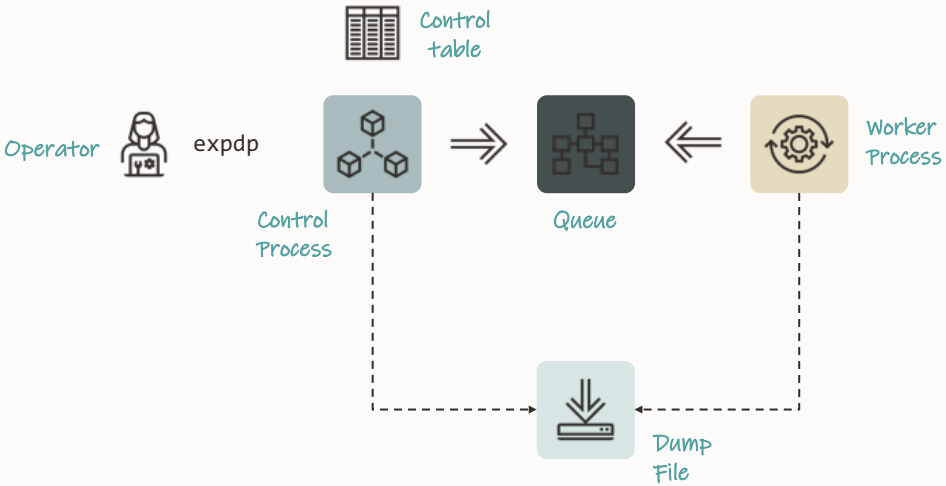
not client-based



Architecture



Architecture



Control Table

Control Table

A regular heap table containing:

- Job info and parameters
- Current status
- Object information
- Index into the dump files
- Enables restarts

Control Table

Data Pump uses control table to keep track

- Workers create/update records with COMPLETION_TIME
- Restart: Workers check for records with missing COMPLETION_TIME

OBJECT_TYPE	START_TIME	COMPLETION_TIME
TABLESPACE	12-SEP-2021:9:04.01	12-SEP-2021:9:05.23
USER	12-SEP-2021:9:05.27	

Control Table

Example of an import

- Workers track import status via STATE and STATUS

OBJECT	OBJECT_SCHEMA	OBJECT_NAME	PROCESSING_STATE	PROCESSING_STATUS
TABLE	SCOTT	EMP	W	C
TABLE	SCOTT	DEPT	U	C
INDEX	SCOTT	IDX1_EMP	R	C
INDEX	SCOTT	IDX1_DEPT	R	C

- R = objects were Retrieved (exported)
- C = objects are Current (successfully imported
- W = objects are Written (imported)
- U = objects are Unknown (import started but did not finish)

Control Table

Control Table

- Control table is dropped upon successful completion of a job
- Optionally, kept using `KEEP_MASTER=Y`
- Can be queried like any other table
- Last object exported to dump file & first imported
- Can be queried for troubleshooting

Unloading and Loading

Data Files

Used for transportable tablespace

Only metadata is unloaded into/loaded from dumpfile

Data remains in data files

Direct Path

External Tables

Insert as Select

Conventional Path

Pro tip: Cross-endian data migration requires data files are converted

Unloading and Loading

Data Files

Direct Path

External Tables

Insert as Select

Conventional Path

Unloads from / load into data files directly

Circumvents SQL layer

Fast

Not usable in all situations

Pro tip: Data Pump automatically selects the best unload/load method

Unloading and Loading

Data Files

Direct Path

External Tables

- Use SQL layer to unload to / load from external table
- Can use `APPEND` hint for faster load
- Very good parallel capabilities
- Dump file format similar to direct path

Insert as Select

Conventional Path

Pro tip: Data unloaded with Data Pump is not compatible with a regular external table (`CREATE TABLE ... ORGANIZATION EXTERNAL ...`)

Unloading and Loading

Data Files

Direct Path

External Tables

Insert as Select

Used by network link imports only

Will disable use of direct path

Conventional Path

Not very common

Unloading and Loading

Data Files

Direct Path

External Tables

Insert as Select

Conventional Path

- Used as last resort
- Slower
- Import only

```
--Data Pump uses table statistics to determine the best access method.
--Inaccurate stats may cause Data Pump to make a bad decision

ACCESS_METHOD=[AUTOMATIC | DIRECT_PATH | EXTERNAL_TABLE |
               CONVENTIONAL_PATH | INSERT_AS_SELECT]
```


```
$ expdp dpuser/oracle schemas=app metrics=y

Export: Release 19.0.0.0.0 - Production on Thu Feb 10 16:51:42 2022
Version 19.14.0.0.0

Copyright (c) 1982, 2019, Oracle and/or its affiliates.  All rights reserved.

Connected to: Oracle Database 19c Enterprise Edition Release 19.0.0.0.0 - Production
Starting "DPUSER"."SYS_EXPORT_SCHEMA_02":  dpuser/******** schemas=app metrics=y
.
.
.
W-1 . exported "APP"."OBJECTS"              8.370 MB  67513 rows in 0 seconds using direct_path
W-1 . exported "APP"."SYS_EXPORT_SCHEMA_01" 329.5 KB   1615 rows in 0 seconds using direct_path
W-1      Completed 2 SCHEMA_EXPORT/TABLE/TABLE_DATA objects in 0 seconds
```

Metadata

A category of metadata is described by an object path

Examples:

`TABLE`
`TABLE/INDEX`
`TABLE/STATISTICS/TABLE_STATISTICS`
`TABLE/TRIGGER`

You can get a full list of object paths from these views:

`DATABASE_EXPORT_OBJECTS`
`SCHEMA_EXPORT_OBJECTS`
`TABLE_EXPORT_OBJECTS`

Metadata

Some metadata has dependencies.

Example: Excluding a table will also exclude

- Indexes
- Constraints
- Grants
- Triggers
- And the like upon that table

Example: Excluding an index will also exclude

- Statistics on that index

To migrate your data, you typically use Data Pump in schema or full mode

SCHEMA

Individual schemas and what they own

FULL

All schemas plus
more or less everything in the database

Full Export

Objects exported only in full export:

- Audit trail and policies
- Database Vault
- Directories
- Profiles and password verify function
- Public database links
- Public synonyms
- Roles
- SQL Management Objects (plan histories, SQL plan baselines, SQL profiles, etc.)
- Tablespaces
- Users (other than those specified in `SCHEMAS` parameter)
- Workspace manager (for schema export you need to use `DBMS_WM.Export_Schemas`)

...

Data Pump never exports grants on SYS objects

- Not even in a full export
- Add them manually following the import

Data Pump never exports AWR

- Not even in a full export
- Use `rdbms/admin/awrextr.sql`

Recommendations

Get the most out of Data Pump

Always use the Data Pump Bundle Patch

More than 200 functional and performance fixes

- Data Pump Recommended Proactive Patches For 19.10 and Above (Doc ID 2819284.1)

Importing a complete application with data drops from almost 2.5 hours to 48 minutes – by just applying the Data Pump bundle patch

A global provider of financial services

Bundle Patch

The patch is non-binary online installable

- Apply while the database instance is running
- Don't use Data Pump or `DBMS_METADATA`

The patch is not RAC rolling installable

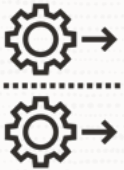
Ensure dictionary and fixed objects statistics are accurate

- Before export
- Before import
- Immediately after import

```
begin
   --dbms_stats.gather_dictionary_stats;
   dbms_stats.gather_schema_stats('SYS');
   dbms_stats.gather_schema_stats('SYSTEM');
   dbms_stats.gather_fixed_objects_stats;
end;
/
```


```
begin
   --dbms_stats.gather_dictionary_stats;
   dbms_stats.gather_schema_stats('SYS');
   dbms_stats.gather_schema_stats('SYSTEM');
   dbms_stats.gather_fixed_objects_stats;
end;
/
```

"After gathering dictionary stats, our Data Pump export went from 46 to 8 minutes"



Use parallel and multiple dump files

```
--Apply parallelism by simply specifying a degree
expdp ... parallel=8


--Use different parallel degree on import
impdp ... parallel=32
```


Oracle Cloud Infrastructure

Number of OCPUs

Number of ECPUs / 4

On-prem (x86-64)

2 x physical cores

On-prem (other)

Depends

Parallel Architecture

```
expdp ... parallel=4
```

Parallel Architecture

Control process

Worker processes

Parallel Architecture

```
expdp ... parallel=4
```

```
select /*+ parallel(2) */ * from t1
select * from t2
select * from t3
select * from t4
```

Worker 4 goes idle

Control process

Worker processes

```
--Use %L to allow multiple dump files
expdp ... parallel=8 dumpfile=exp%L.dmp
```


```
--Split dump files into minor files for easier transport
expdp ... parallel=8 dumpfile=exp%L.dmp filesize=10G
```


```
-- After export, store a checksum in the dump file.
-- Detects in-flight corruption or alteration.
-- Specify other algorithms using checksum_algorithm parameter.

expdp ... checksum=yes

impdp ... verify_checksum=yes
          verify_only=yes
```


For best protection against dump file tampering, use encrypted dump files

- Checksum is a weaker protection
- Requires Advanced Security Option

```
-- Protect your dump files from alteration by using encryption
-- Creating an encrypted dump file requires Advanced Security Option

expdp ... encryption=all encryption_algorithm=AES256
```


```
-- Protect your dump files from alteration by using encryption
-- Creating an encrypted dump file requires Advanced Security Option

expdp ... encryption=all encryption_algorithm=AES256
```

New default value

Transportable jobs can use parallel in Oracle Database 21c

```
-- Any transportable jobs can now run in parallel
-- Parallel unload/load of metadata provide a significant performance boost

expdp ... full=y transportable=always parallel=16
expdp ... tablespace=<list> parallel=16

impdp ... parallel=16
```


Parallel Transportable | Benchmark

Oracle E-Business Suite database

600.000+ objects

Export parallel 1	2h 2m
Import parallel 1	6h 44m
Total	**8h 46m**

Export parallel 16	1h 8m
Import parallel 16	1h 23m
Total	**2h 31m**

Speed up imports by using NOVALIDATE constraints

A Constraint Can Be

VALIDATED

All data in the table obeys the constraint.
The database guarantees that data is good.

NOT VALIDATED

All data in the table may obey the constraint.
The database does not know if data is good.

Most constraints are VALIDATED

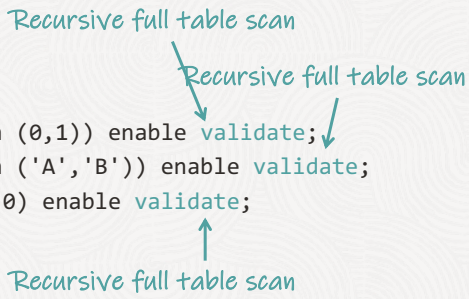
On import, Data Pump creates constraints in the same state as in the source

```
--Example of which commands Data Pump import might execute as part of an import

create table sales ( .... );

insert into sales as select ... ;

alter table sales add constraint c_sales_1 check (c1 in (0,1)) enable validate;
alter table sales add constraint c_sales_2 check (c2 in ('A','B')) enable validate;
alter table sales add constraint c_sales_3 check (c3 > 0) enable validate;
```



```
-- Add constraints with NOVALIDATE keyword regardless of state in source database
-- Significantly speeds up add constraints for larger tables

impdp ... transform=constraint_novalidate
```


```
--Transforming constraints to NOVALIDATE to speed up import

alter table sales add constraint c_sales_1 check (c1 in (0,1)) enable novalidate;
alter table sales add constraint c_sales_2 check (c2 in ('A','B')) enable novalidate;
alter table sales add constraint c_sales_3 check (c3 > 0) enable novalidate;
```

```
--Transforming constraints to NOVALIDATE to speed up import

alter table sales add constraint c_sales_1 check (c1 in (0,1)) enable novalidate;
alter table sales add constraint c_sales_2 check (c2 in ('A','B')) enable novalidate;
alter table sales add constraint c_sales_3 check (c3 > 0) enable novalidate;
```

Database validates new rows

Benchmark, 1 billion rows

Importing VALIDATE constraints

```
10-AUG-24 00:32:28.716: W-1 Processing object type TABLE_EXPORT/TABLE/TABLE_DATA
10-AUG-24 00:36:42.762: W-1 . . imported "FUSION"."hwr_topic_tl" 151.2 GB 1044625000 rows in 254 seconds using external_table
10-AUG-24 00:45:41.226: W-1 Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/CONSTRAINT
10-AUG-24 00:55:35.787: W-1      Completed 7 CONSTRAINT objects in 594 seconds
```

Importing NOVALIDATE constraints

```
10-AUG-24 00:14:56.050: W-1 Processing object type TABLE_EXPORT/TABLE/TABLE_DATA
10-AUG-24 00:19:10.311: W-1 . . imported "FUSION"."hwr_topic_tl" 151.2 GB 1044625000 rows in 254 seconds using external_table
10-AUG-24 00:29:20.841: W-1 Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/CONSTRAINT
10-AUG-24 00:29:21.101: W-1      Completed 7 CONSTRAINT objects in 1 seconds
```

NOVALIDATE constraints prevent the optimizer from certain query rewrites

- Check `QUERY_REWRITE_INTEGRITY`

```
alter table sales add constraint c_sales_1 check (c1 in (0,1)) enable novalidate;

------------------------
------- GO LIVE -------
------------------------

#Validate constraints
#Optionally, use parallel query
alter session force parallel query;
alter table sales modify constraint c_sales_1 enable validate;
```


Exceptions

Data Pump always validates certain constraints:

1. On `DEFAULT ON NULL` columns
2. Used by a reference partitioned table
3. Used by a reference partitioned child table
4. Table with Primary key OID
5. Used as clustering key on a clustered table

Validate constraints after import, or even after go-live

- Still requires a full scan of the table
- But can use parallel query
- And no table lock!

Use with care if
you are transforming data on import

Also available in Oracle Database 19c via 19.23.0 Data Pump Bundle Patch

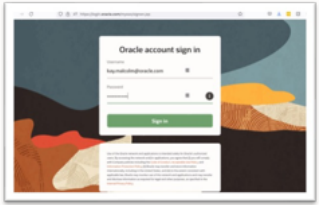
- Plus patch 37280692 - or be on 19.27

Lab Environment

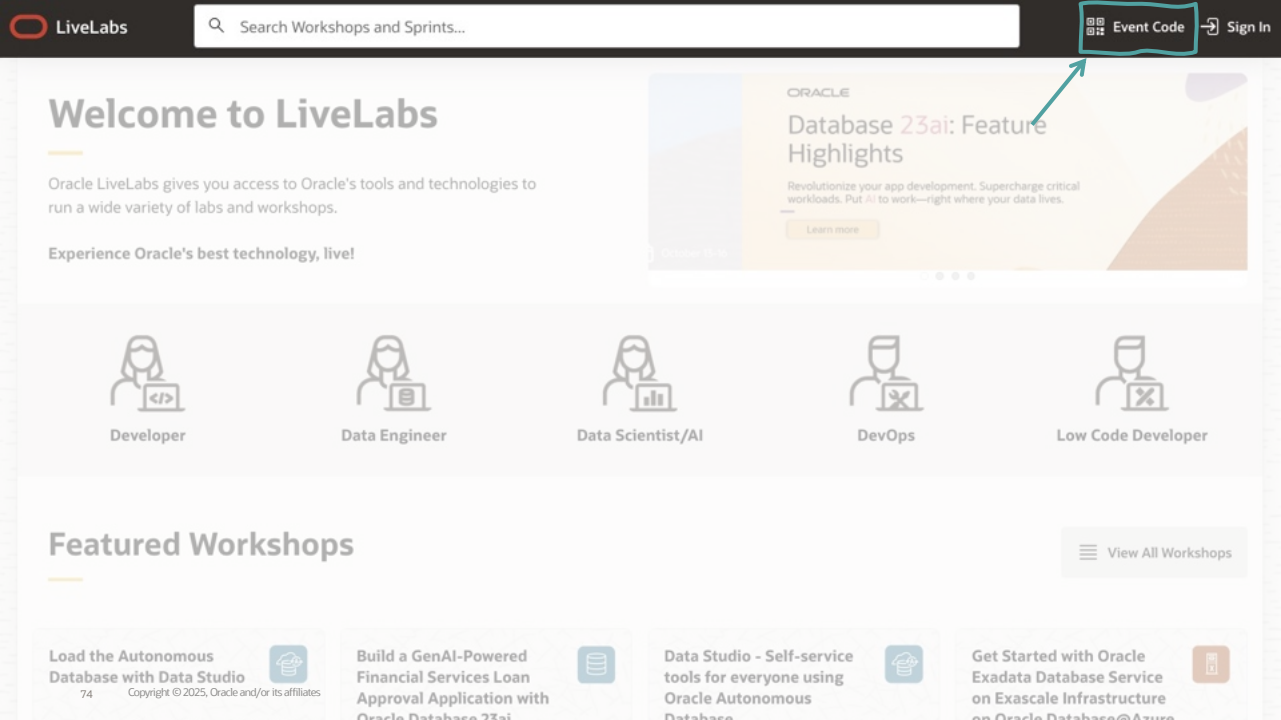
Oracle LiveLabs

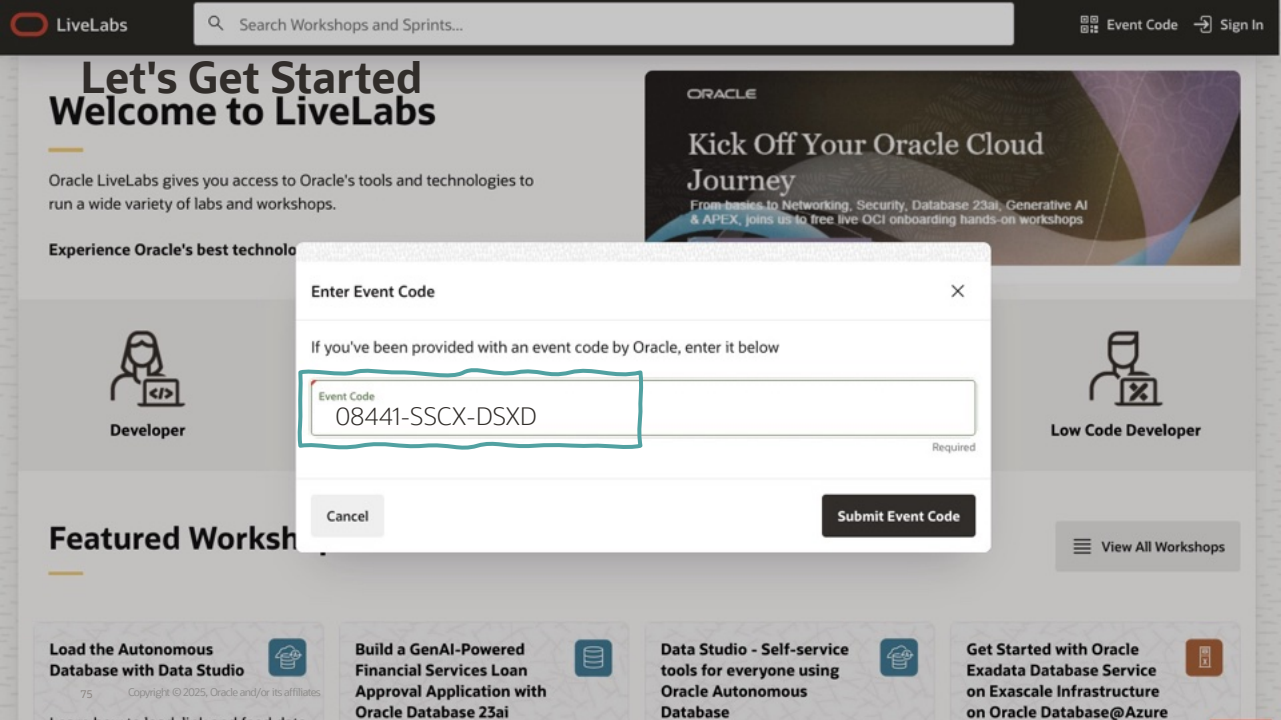
Requirements

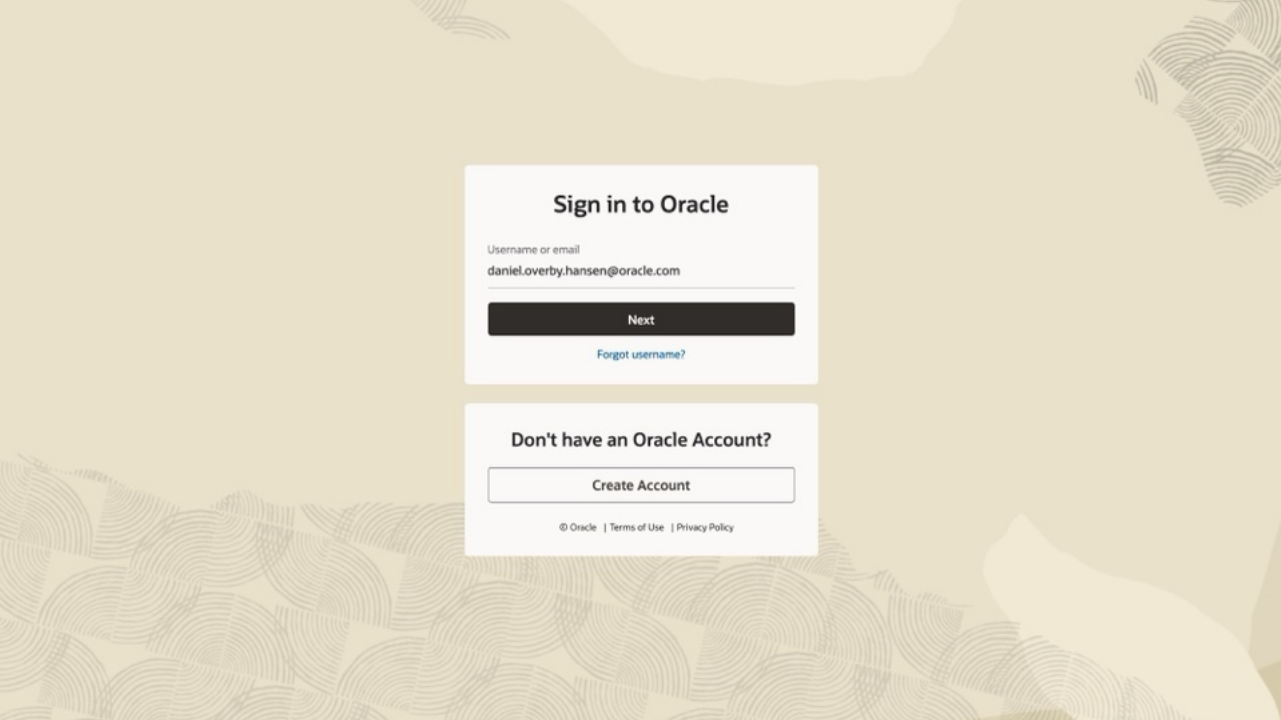
Laptop



Oracle Account

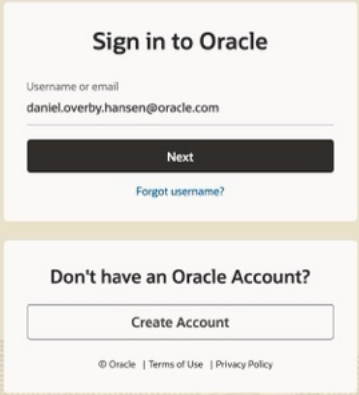



LiveLabs
Search Workshops and Sprints...
Event Code
Sign In
Let's Get Started
Welcome to LiveLabs
Oracle LiveLabs gives you access to Oracle's tools and technologies to run a wide variety of labs and workshops.
Experience Oracle's best technolo
ORACLE
Kick Off Your Oracle Cloud Journey
From basics to Networking, Security, Database 23ai, Generative AI & APEX, joins us to free live OCI onboarding hands-on workshops
Enter Event Code
If you've been provided with an event code by Oracle, enter it below
Event Code
08441-SSCX-DSXD
Required
Cancel
Submit Event Code
Developer
Low Code Developer
Featured Worksh
View All Workshops
Load the Autonomous Database with Data Studio
Build a GenAI-Powered Financial Services Loan Approval Application with Oracle Database 23ai
Data Studio - Self-service tools for everyone using Oracle Autonomous Database
Get Started with Oracle Exadata Database Service on Exascale Infrastructure on Oracle Database@Azure
75
Copyright © 2025, Oracle and/or its affiliates


Sign in to Oracle
Username or email
daniel.overby.hansen@oracle.com
Next
Forgot username?
Don't have an Oracle Account?
Create Account
© Oracle | Terms of Use | Privacy Policy

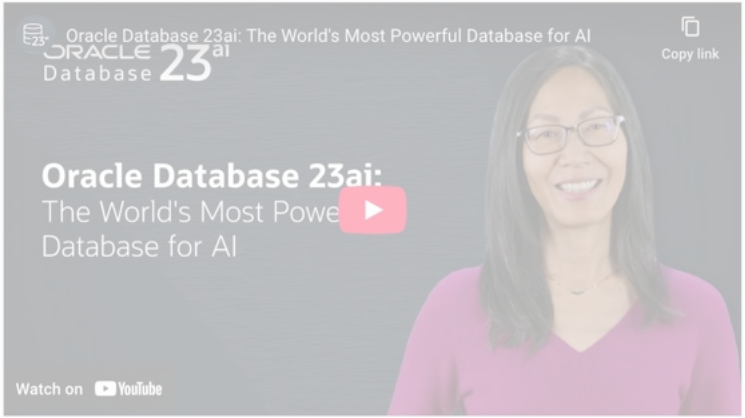
Use your Oracle Account to log in

- This is not your Oracle Cloud Tenancy account
- It is your Oracle SSO account



Supercharge data movement with Data Pump

Share
Start



4 hours

Organizer:
Oracle

Event Date:
Wednesday 21 May

Outline

- Configure Data Pump
- Apply best practices and speed up imports
- Customize and secure you exports and imports
- Monitor, troubleshooting and trace
- Upgrade, convert and downgrade

Prerequisites

- Familiarity with Oracle Database is required
- Data Pump experience is **not** a requirement
- Some understanding of database terms is helpful

Supercharge data movement with Data Pump

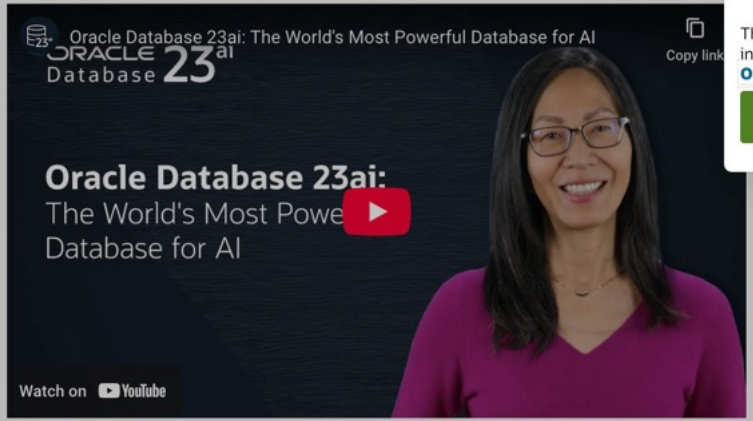
Gain practical experience with Oracle Data Pump; the powerful utility for fast, customizable, and

Supercharge data movement with Data Pump

Share
Start

The **Run on LiveLabs** button will dynamically create resources in an Oracle-owned tenancy for you to use for free!
Oracle account help | ***Oracle account signup***

Run on LiveLabs Sandbox



- Configure Data Pump
- Apply best practices and speed up imports
- Customize and secure you exports and imports
- Monitor, troubleshooting and trace
- Upgrade, convert and downgrade

Prerequisites

- Familiarity with Oracle Database is required
- Data Pump experience is **not** a requirement
- Some understanding of database terms is helpful

Supercharge data movement with Data Pump

Gain practical experience with Oracle Data Pump; the powerful utility for fast, customizable, and

LiveLabs

Search Workshops and Sprints...

Supercharge data movement with Data P

Supercharge data movement with Data Pump

Gain practical experience with Oracle Data Pump; the powerful utility for fast,

Reserve Workshop

Attendee Email Address
daniel.overby.hansen@oracle.com

Attendee Timezone
EET (+03:00)

Required

* Start Workshop Now?

I consent to receive emails from LiveLabs for my reservation and I agree that I will not upload sensitive personal or company information to Oracle Cloud Infrastructure *

Preview Sandbox Instructions

Submit Reservation

My Reservations

Reservation submitted

All your current workshop reservations are shown below. You can edit active or pending reservations, view workshop details, attend an available workshop, or delete a reservation.

To access this page again click the user dropdown in the top right corner and select **My Reservations**

Note: The status of your reservations will be emailed to you. Check your mail for any status updates.

Supercharge data movement with Data Pump
Wednesday May 21st, 12:01pm (12:01) EET

Pending creation

Refresh the page a few times
until a lab environment is assigned

My Reservations

All your current workshop reservations are shown below. You can edit active or pending reservations, view workshop details, attend an available workshop, or delete a reservation.

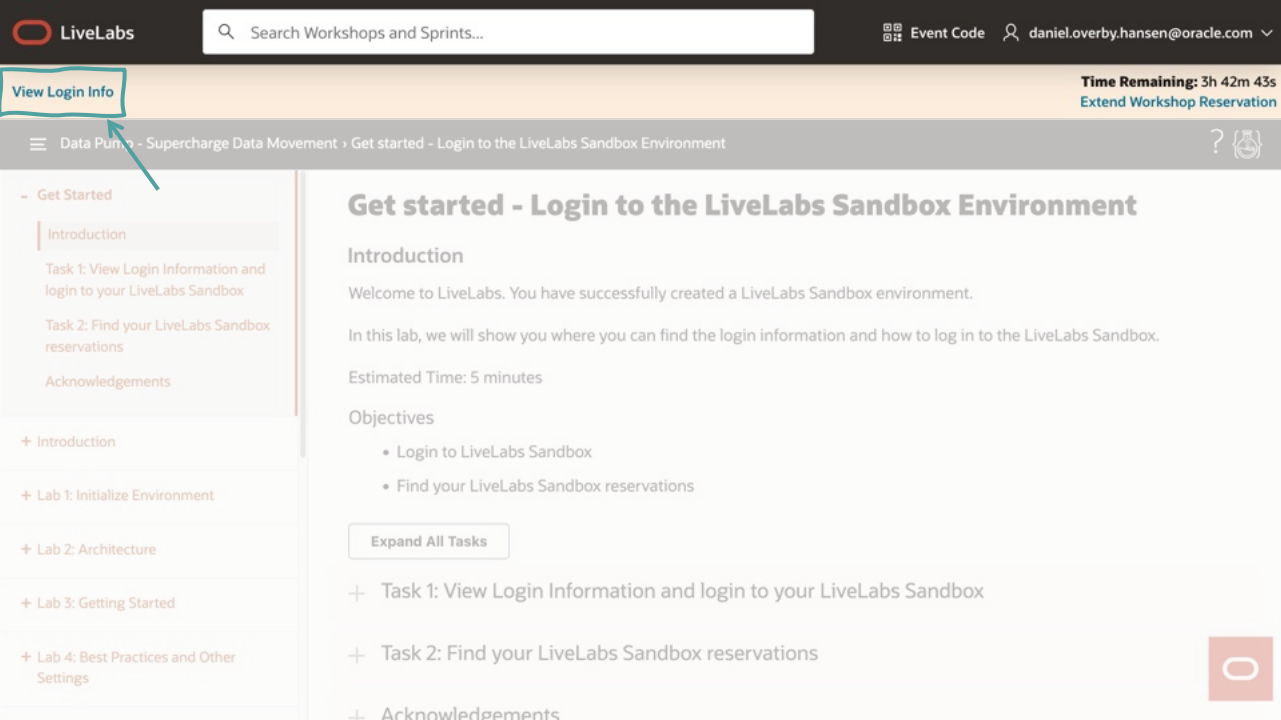
To access this page again click the user dropdown in the top right corner and select **My Reservations**

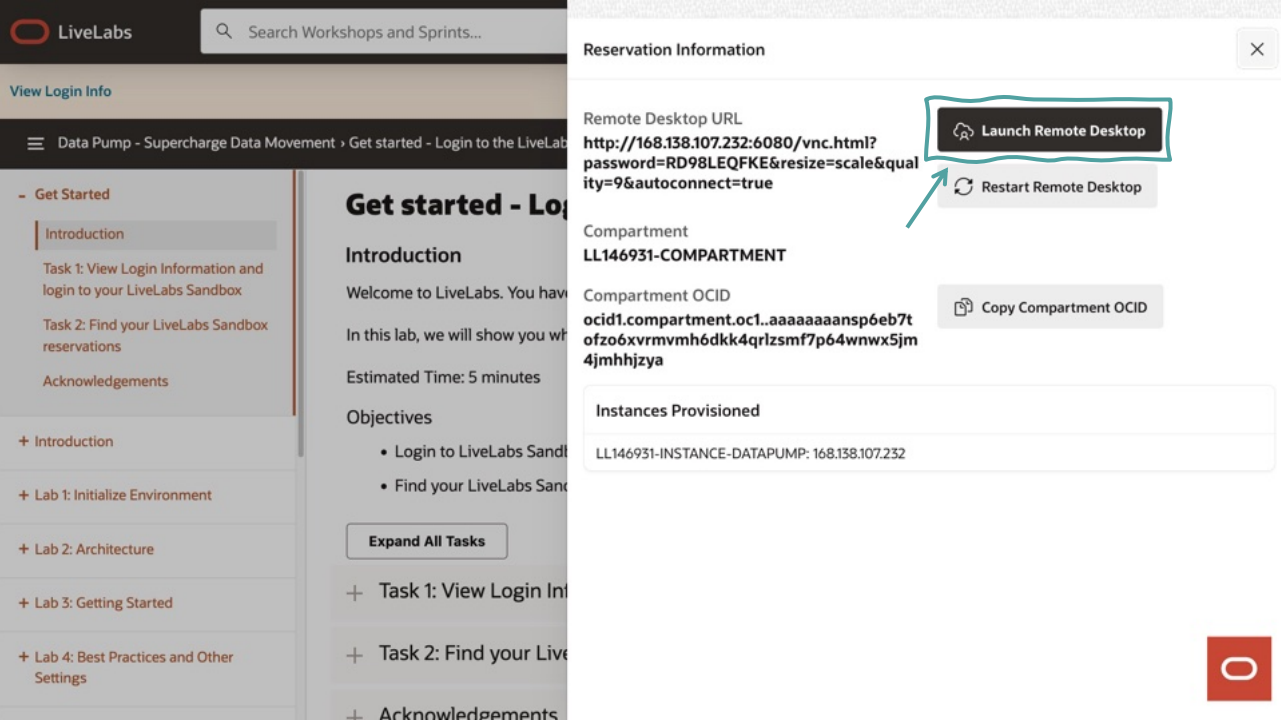
Note: The status of your reservations will be emailed to you. Check your mail for any status updates.

Supercharge data movement with Data Pump

Wednesday May 21st, 12:01pm (12:01) EET

Details Delete


LiveLabs
Search Workshops and Sprints...
Event Code
daniel.overby.hansen@oracle.com
View Login Info
Time Remaining: 3h 42m 43s
Extend Workshop Reservation
Data Pump - Supercharge Data Movement › Get started - Login to the LiveLabs Sandbox Environment
Get Started
Introduction
Task 1: View Login Information and login to your LiveLabs Sandbox
Task 2: Find your LiveLabs Sandbox reservations
Acknowledgements
Introduction
Lab 1: Initialize Environment
Lab 2: Architecture
Lab 3: Getting Started
Lab 4: Best Practices and Other Settings
Get started - Login to the LiveLabs Sandbox Environment
Introduction
Welcome to LiveLabs. You have successfully created a LiveLabs Sandbox environment.
In this lab, we will show you where you can find the login information and how to log in to the LiveLabs Sandbox.
Estimated Time: 5 minutes
Objectives
Login to LiveLabs Sandbox
Find your LiveLabs Sandbox reservations
Expand All Tasks
Task 1: View Login Information and login to your LiveLabs Sandbox
Task 2: Find your LiveLabs Sandbox reservations
Acknowledgements


LiveLabs
Search Workshops and Sprints...
View Login Info
Data Pump - Supercharge Data Movement › Get started - Login to the LiveLab
Get Started
Introduction
Task 1: View Login Information and login to your LiveLabs Sandbox
Task 2: Find your LiveLabs Sandbox reservations
Acknowledgements
Introduction
Lab 1: Initialize Environment
Lab 2: Architecture
Lab 3: Getting Started
Lab 4: Best Practices and Other Settings
Get started - Log
Introduction
Welcome to LiveLabs. You have
In this lab, we will show you wh
Estimated Time: 5 minutes
Objectives
Login to LiveLabs Sandb
Find your LiveLabs Sand
Expand All Tasks
Task 1: View Login Inf
Task 2: Find your Live
Acknowledgements
Reservation Information
Remote Desktop URL
http://168.138.107.232:6080/vnc.html?password=RD98LEQFKE&resize=scale&quality=9&autoconnect=true
Launch Remote Desktop
Restart Remote Desktop
Compartment
LL146931-COMPARTMENT
Compartment OCID
ocid1.compartment.oc1..aaaaaaaansp6eb7tofzo6xvrmvmh6dkk4qrlzsmf7p64wnwx5jm4jmhhjzya
Copy Compartment OCID
Instances Provisioned
LL146931-INSTANCE-DATAPUMP: 168.138.107.232

HTTPS-Only Mode Alert

Secure Site Not Available

You've enabled HTTPS-Only Mode for enhanced security, and a HTTPS version of **168.138.107.232:6080** is not available.

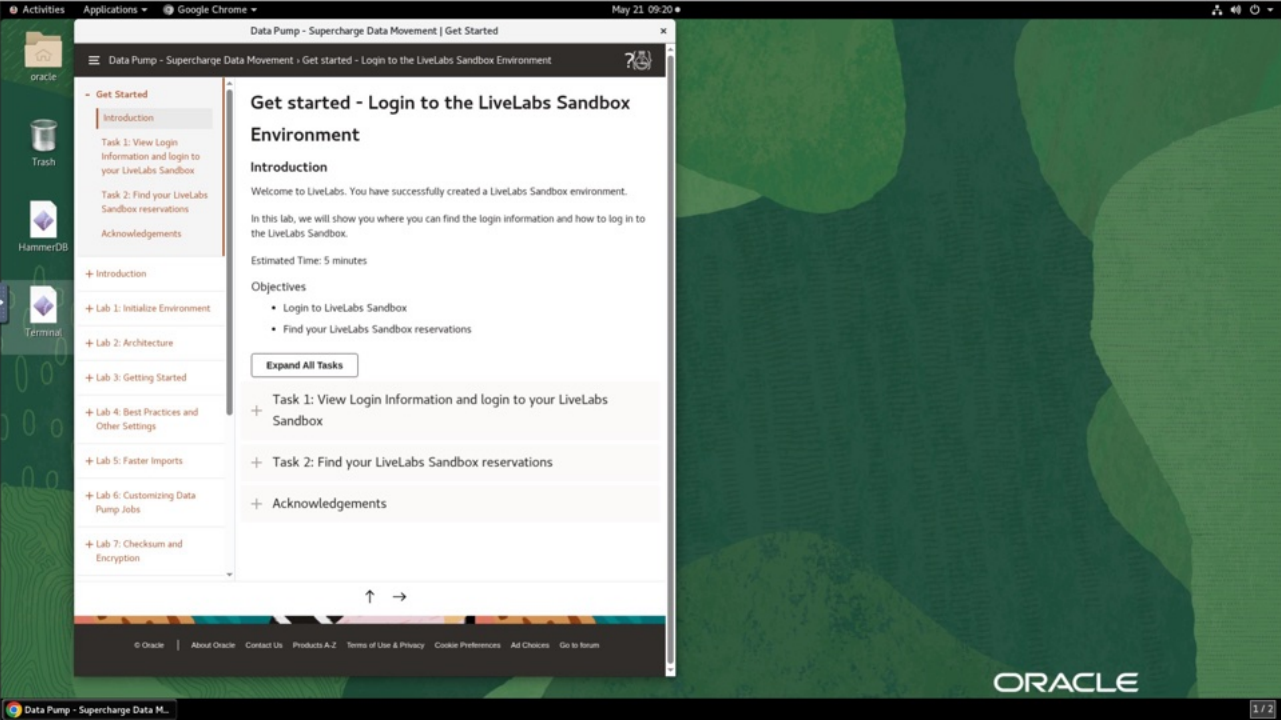
Learn More...

What could be causing this?

- Most likely, the website simply does not support HTTPS.
- It's also possible that an attacker is involved. If you decide to visit the website, you should not enter any sensitive information like passwords, emails, or credit card details.

If you continue, HTTPS-Only Mode will be turned off temporarily for this site.

Continue to HTTP Site | Go Back


Activities
Applications
Google Chrome
May 21 09:20
Data Pump - Supercharge Data Movement | Get Started
Data Pump - Supercharge Data Movement › Get started - Login to the LiveLabs Sandbox Environment
Get started - Login to the LiveLabs Sandbox Environment
Introduction
Welcome to LiveLabs. You have successfully created a LiveLabs Sandbox environment.
In this lab, we will show you where you can find the login information and how to log in to the LiveLabs Sandbox.
Estimated Time: 5 minutes
Objectives
Login to LiveLabs Sandbox
Find your LiveLabs Sandbox reservations
Expand All Tasks
Task 1: View Login Information and login to your LiveLabs Sandbox
Task 2: Find your LiveLabs Sandbox reservations
Acknowledgements
Get Started
Introduction
Task 1: View Login Information and login to your LiveLabs Sandbox
Task 2: Find your LiveLabs Sandbox reservations
Acknowledgements
Introduction
Lab 1: Initialize Environment
Lab 2: Architecture
Lab 3: Getting Started
Lab 4: Best Practices and Other Settings
Lab 5: Faster Imports
Lab 6: Customizing Data Pump Jobs
Lab 7: Checksum and Encryption
© Oracle
About Oracle
Contact Us
Products A-Z
Terms of Use & Privacy
Cookie Preferences
Ad Choices
Go to forum
oracle
Trash
HammerDB
Terminal
Data Pump - Supercharge Data M...
1 / 2

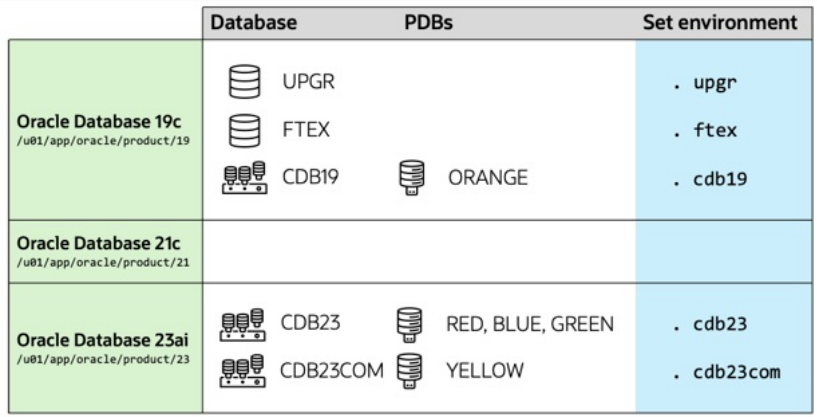
You can copy/paste from the instructions

- SHIFT + CTRL + C
- SHIFT + CTRL + V

Introduction

Data Pump - Hands-On Lab

Overview



Labs

1. Initialize
2. Architecture
3. Getting Started
4. Best Practices and Other Settings

Labs

5. Faster imports
6. Customizing Data Pump Jobs
7. Checksum and Encryption
8. Determining Import Success

Labs

9. Monitoring, Troubleshooting and Tracing
10. Upgrading, Downgrading and Converting
11. Using DBMS_DATAPUMP

Complete the lab in the listed order

Exercises

- Take your time and understand the tasks
- Finish the exercises later today
- The lab is available, for free, via Oracle LiveLabs

Let's Get Started

Data Pump Hands-On Lab

Let's Get Started

https://livelabs.oracle.com

08441-SSCX-DSXD

Top Tips

Supercharge Data Pump Performance

You get the fastest LOB operations with SecureFile LOBs

2007

Oracle Database 11*g* Release 1

```
-- Do you still have any old BasicFile LOBs in your database?

select * from dba_lobs where securefile='NO';
```


If exporting SecureFile LOBs is slow, apply 19.23.0 Data Pump Bundle Patch

- Alternatively, trick Data Pump with fake stats

By applying the Data Pump Bundle Patch our 4.3 TB export with huge LOBs went from over 21 hours to 3 hours 22 minutes

A European government agency

… Plus, by increasing parallel from 4 to 12 the export dropped to 1 hour 51 minutes

A European government agency

… Finally, we moved to faster ASM based storage bringing it to 1 hour 7 minutes

A European government agency

Do you still have BasicFile LOBs?

- Use DIY parallelism during export

```
--Converting a BasicFile LOB to SecureFile during import,
--is faster than not converting it.
--Overview of Oracle LOBs (Doc ID: 1490228.1)

impdp ... transform=lob_storage:securefile
```


Importing as BasicFile LOBs

```
... imported "SCHEMA"."TABLE"   31.83 GB  681025 rows in 804 seconds using direct_path
```

Importing as SecureFile LOBs

```
... imported "SCHEMA"."TABLE"   31.83 GB  681025 rows in 261 seconds using external_table
```


Do you still have **LONG** and **LONG RAW**?

- Deprecated since Oracle8i

```
-- Convert LONG to CLOB, and LONG RAW to BLOB on import
-- Be sure to change your application as well,
-- PL/SQL interface for accessing LOBs and LONGs are not the same

impdp ... transform=long_to_lob:y
```


Even faster index imports

Use index size to determine parallel degree on index creation

- Requires 23.8 and Data Pump Bundle Patch

Index Creation

```
impdp ... parallel=16
```

Really good for few big indexes

Before 12.1

Worker 1 `CREATE INDEX .... PARALLEL 16`

Index Creation

```
impdp ... parallel=16
```

From 12.1

Worker 1	`CREATE INDEX .... PARALLEL 1`
Worker 2	`CREATE INDEX .... PARALLEL 1`
...	`CREATE INDEX .... PARALLEL 1`
Worker 16	`CREATE INDEX .... PARALLEL 1`

Index Creation

```
impdp ... parallel=16
```

From 23

The best of both worlds

Worker 1	`CREATE INDEX .... PARALLEL 1`
Worker 2	`CREATE INDEX .... PARALLEL 8`
Worker 3	`CREATE INDEX .... PARALLEL 4`
Worker 4	`CREATE INDEX .... PARALLEL 3`

How Data Pump Create Indexes

1. Calculate the optimal parallel degree
2. Create indexes

How Data Pump Create Indexes

1 Calculate the optimal parallel degree

- Always parallel 1 when a table is less than 150 MB
- Customizable via `INDEX_THRESHOLD`
- Get optimal parallel degree using `EXPLAIN PLAN`

```
SQL> explain plan for create index i1 on t1(c1) parallel;

Explained.
```


```
SQL> explain plan for create index i1 on t1(c1) parallel;

Explained.

SQL> select * from table(dbms_xplan.display(format => 'ALL'));

...

Note
-----
   - automatic DOP: Computed Degree of Parallelism is 4 because of degree limit
   - estimated index size: 655K bytes
```

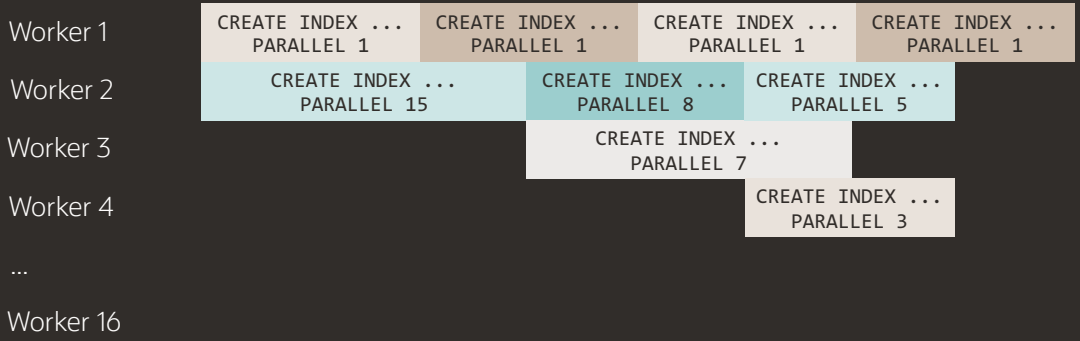

How Data Pump Creates Indexes

2 Create indexes

- One worker creates small indexes (parallel 1) in large batches
- The next worker starts with the biggest index (measured by optimal parallel degree)

How Data Pump Creates Indexes

```
impdp ... parallel=16
```



Benchmark

Importing with former index method

```
10-MAY-25 16:18:55.130: W-12 Processing object type SCHEMA_EXPORT/TABLE/INDEX/INDEX
10-MAY-25 16:36:46.902: W-30      Completed 480 INDEX objects in 1071 seconds
```

Importing with new index method

```
10-MAY-25 15:47:50.267: W-4 Processing object type SCHEMA_EXPORT/TABLE/INDEX/INDEX
10-MAY-25 15:59:17.006: W-3      Completed 480 INDEX objects in 686 seconds
```


Also available in Oracle Database 19c
via 19.26.0 Data Pump Bundle Patch

Dealing with optimizer statistics

Generally, we recommend excluding statistics from Data Pump export

- Use `exclude=statistics`

1 Include statistics in Data Pump

2 Exclude statistics in Data Pump
Regather statistics after import

3 Exclude statistics in Data Pump
Import statistics using `DBMS_STATS`

1 Include statistics in Data Pump

2 Exclude statistics in Data Pump
Regather statistics after import

3 Exclude statistics in Data Pump
Import statistics using `DBMS_STATS`

We have adopted this method for stats.
We migrated 60 TB database from AIX to Exadata using cross-platform transportable tablespace

Taoqir Hassan, comment on YouTube channel

Gathering stats from scratch took more than 36 hours.
We transported the statistics in less than 2 hours.

Taoqir Hassan, comment on YouTube channel

`EXCLUDE=STATISTICS`

Table statistics

Index statistics

Statistics preferences

Column usage information

```
EXCLUDE=STATISTICS
```

Table statistics

Index statistics

Statistics preferences

Column usage information

```
BEGIN
   DBMS_STATS.SET_TABLE_PREFS (
      OWNNAME => '...',
      TABNAME => '...',
      PNAME   => 'TABLE_CACHED_BLOCKS',
      PVALUE  => '42'
   );
END;
```


Table 171-131 SET_TABLE_PREFS Procedure Parameters

Parameter	Description
ownname	Owner name
tabname	Table name
pname	Preference name. You can set the default value for following preferences: • APPROXIMATE_NDV_ALGORITHM • AUTO_STAT_EXTENSIONS • CASCADE • DEGREE • ESTIMATE_PERCENT • GRANULARITY • INCREMENTAL • INCREMENTAL_LEVEL • INCREMENTAL_STALENESS • METHOD_OPT • NO_INVALIDATE • OPTIONS • PREFERENCE_OVERRIDES_PARAMETER • PUBLISH • STALE_PERCENT • TABLE_CACHED_BLOCKS
pvalue	Preference value. If NULL is specified, it will set the Oracle default value.

PL/SQL Packages and Types Reference, Oracle Database 19c

Data Pump exports table-level statistics preferences together with table statistics

- In full, schema and table mode
- In transportable, it is controlled by `USER_PREF_STATISTICS`

Data Pump never exports global statistics preferences

- Not even in a full export
- Define manually using `DBMS_STATS.SET_GLOBAL_PREFS`

`DBMS_STATS` has dedicated procedures for transporting table-level statistics prefs

You often use statistics preferences to solve a particular problem

- Evaluate whether that problem exists in the target environment

```
EXCLUDE=STATISTICS
```

Table statistics

Index statistics

Statistics preferences

Column usage information

Column Usage Information

- Information on how you join tables
- Used by the optimizer to determine when to create histograms
  `METHOD_OPT => ... SIZE AUTO`
- When missing, statistics gathering creates no or few histograms
- Stored internally in `SYS.COL_USAGE$`

When Data Pump transfers statistics,
it also transfers column usage information

EXCLUDE

`EXCLUDE=STATISTICS`

`COL_USAGE$` empty

REGATHER

First time only

```
METHOD_OPT =>
SIZE SKEWONLY
```

GO LIVE

Column usage information is updated

REGATHER

Use default

```
METHOD_OPT =>
SIZE AUTO
```

Importing statistics might be a bad idea

When source and target database do not match

When Importing Stats Is Bad

Fragmented table

Blocks	12000
Leaf blocks	11000
B-level	4
Clustering factor	10000

Compacted table

Blocks	12000
Leaf blocks	11000
B-level	4
Clustering factor	10000

```
DBMS_STATS.GATHER_TABLE_STATS(...
```

Blocks	5000
Leaf blocks	4000
B-level	2
Clustering factor	20000

When Importing Stats Is Bad

- Potentially a problem
  - Fragmented tables
  - Changing block size
  - Changing character set
  - Compress or decompress

  …

- Only a problem for table and index base statistics, column statistics remain accurate

Accurate statistics is the starting point for good performance

Bits and pieces

Time Zone File Version Check

Source
Version 43

Target
Version 42

```
create table t1 (
   ...
   c1 timestamp with timezone
   ...
)
```

```
Import: Release 19.0.0.0.0 - Production on Sun Sep 1 06:17:06 2024
Version 19.21.0.0.0

Copyright (c) 1982, 2019, Oracle and/or its affiliates.  All rights reserved.

Connected to: Oracle Database 19c Enterprise Edition Release 19.0.0.0.0 - Production
ORA-39002: invalid operation
ORA-39405: Oracle Data Pump does not support importing from a source database with
TSTZ version 43 into a target database with TSTZ version 42.
```


Works in 19.27 plus Data Pump Bundle Patch by converting the data

```
-- How do you deal with large Data Pump import log files?
-- In this example, the Data Pump import log file has almost 200.000 lines

$ du -h import.log
 29M   import.log

$ wc -l import.log
  189931 import.log
```

```
$ python3 dpla.py import.log

========================
 Data Pump Log Analyzer
========================

...

Operation Details
~~~~~~~~~~~~~~~~~
 Operation:                  Import
 Data Pump Version:          19.22.0.0.0
 DB Info:                    Oracle Database 19c EE Extreme Perf Release 19.0.0.0.0
 Job Name:                   DPJOB1
 Status:                     COMPLETED
   Processing:               -
 Errors:                     1267
   ORA- Messages:            1267
 Start Time:                 2024-04-11 09:30:55
 End Time:                   2024-04-12 10:33:01
 Runtime:                    25:03:06

Data Processing
~~~~~~~~~~~~~~~
 Parallel Workers:           128
 Schemas:                    27
 Objects:                    224755
 Data Objects:               188084
 Overall Size:               13.16 TB
```

```
$ python3 dpla.py import.log -e

========================
 Data Pump Log Analyzer
========================

...

ORA- MESSAGES DETAILS
~~~~~~~~~~~~~~~~~~~~~
(sorted by count):

 Message                                                                                                  Count
 -------------------------------------------------------------------------------------------------------- ---------
 ORA-39346: data loss in character set conversion for object COMMENT                                        919
 ORA-39082: Object type PACKAGE BODY created with compilation warnings                                      136
 ORA-39346: data loss in character set conversion for object PACKAGE_BODY                                    54
 ORA-39082: Object type TRIGGER created with compilation warnings                                            36
 ORA-39082: Object type PROCEDURE created with compilation warnings                                          29
 ORA-31684: Object type USER already exists                                                                  27
 ORA-39111: Dependent object type PASSWORD_HISTORY skipped, base object type USER already exists             27
 ORA-39346: data loss in character set conversion for object PACKAGE                                         18
 ORA-39082: Object type PACKAGE created with compilation warnings                                            10
 ORA-39082: Object type VIEW created with compilation warnings                                                7
 ORA-39346: data loss in character set conversion for object PROCEDURE                                        2
 ORA-39082: Object type FUNCTION created with compilation warnings                                            2
 -------------------------------------------------------------------------------------------------------- ---------
 Total                                                                                                     1267
 -------------------------------------------------------------------------------------------------------- ---------
```

```
$ python3 dpla.py import.log -o

========================
 Data Pump Log Analyzer
========================

...

 Object                              Count     Seconds     Workers     Duration
 ---------------------------------- ---------- ----------- ----------- ------------
 SCHEMA_EXPORT/TABLE/TABLE_DATA         188296     6759219         128      6759219
 CONSTRAINT                                767       37253           1        37253
 TABLE                                    2112        3225          51          156
 COMMENT                                 26442         639         128           18
 PACKAGE_BODY                              197         125         128            5
 OBJECT_GRANT                             5279          25           1           25
 TYPE                                      270           6           1            6
 ALTER_PROCEDURE                           149           5           2            3
 ALTER_PACKAGE_SPEC                        208           4           3            2
 PACKAGE                                   208           3           3            1
 PROCEDURE                                 149           2           2            1

 ...

 ---------------------------------- ---------- ----------- ----------- ------------
 Total                                  224755     6800515         128      6796697
 ---------------------------------- ---------- ----------- ----------- ------------
```

How about NOVALIDATE constraints?

≡ Data Pump Log Analyzer

Table Details

Search for Table...

Table	Rows	Size	Seconds	Part	Subpart
SALES.ORDERS	118914251151	1.73 TB	878854	278	4448
SALES.INVOICES	115668171592	4.33 TB	805901	588	9408
SALES.TRANSACTIONS	115720037994	3.61 TB	611891	451	7216
FINANCE.EXPENSES	35091517646	258.14 GB	112962	367	0
MARKETING.CAMPAIGNS	11621627768	458.93 GB	82801	16	0
HR.EMPLOYEES	19433932893	296.19 GB	66156	2254	0
SALES.DOCUMENTS	4743542596	345.97 GB	48117	589	9424
SALES.REPORTS	4744610748	263.63 GB	42904	440	7040
INVENTORY.EQUIPMENT	9824954344	51.01 GB	33290	130	0
HR.PARTNERS	3983265247	83.62 GB	16388	3046	0

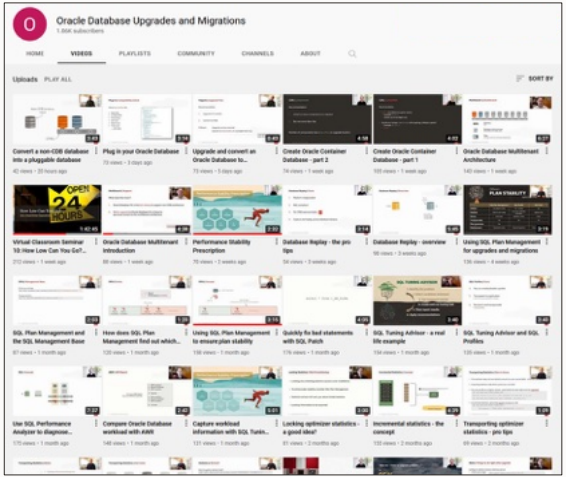
Data Pump Log Analyzer

- Free to use
- Download from GitHub
- Not an official Oracle tool
- Created by Marcus Doeringer
  Our migration superstar

Key Learnings

1. Use Data Pump Bundle Patch
2. Transform to NOVALIDATE constraints
3. Faster index creation

YouTube | Oracle Database Upgrades and Migrations



https://www.youtube.com/@upgradenow

- 300+ videos
- New videos every week
- No marketing
- No buzzword
- All tech

SCAN ME

Thank You