How Autonomous is the Oracle Autonomous Data Warehouse?

Christian Antognini / Dani Schnider



@chrisantognini antognini.ch/blog
@dani_schnider danischnider.wordpress.com

Agenda

1. Introduction
2. Set Up
3. Connectivity
4. Loading Data
5. ETL Performance
6. Query Performance
7. Monitoring Performance
8. Miscellaneous
9. Conclusion

trivadis makes IT easier.

Introduction

trivadis makes IT easier.

Larry Ellison, ATP Announcement, 7th August 2018

“There is nothing to learn,
there is nothing to do”



trivadis
makes IT easier.

Set Up

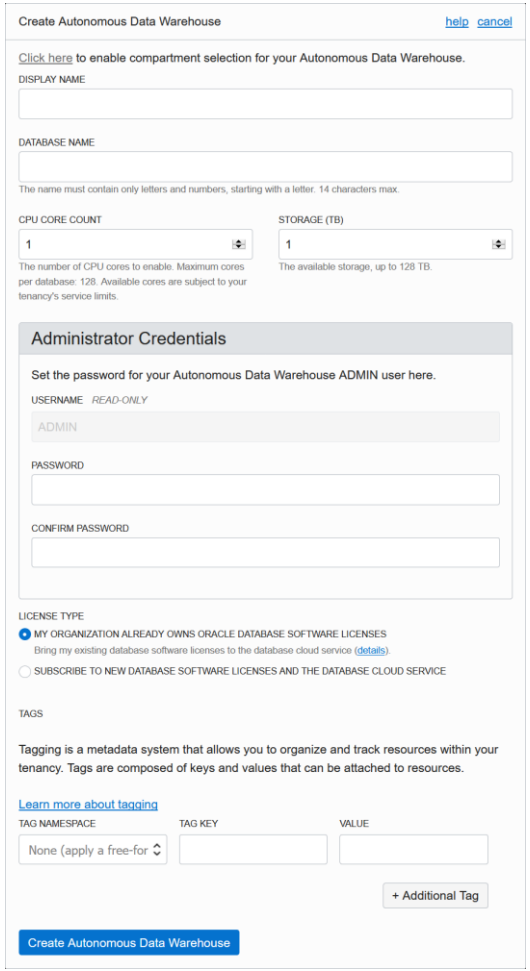
trivadis makes IT easier.

Creating

Number of CPU (1..128)

PDB admin password

Tags (optional)



Name

Storage (1..128 TB)

License type

trivadis makes IT easier.

Initialization Parameters That Can Be Modified

APPROX_FOR_AGGREGATION

APPROX_FOR_COUNT_DISTINCT

APPROX_FOR_PERCENTILE

AWR_PDB_AUTOFLUSH_ENABLED[1]

OPTIMIZER_CAPTURE_SQL_PLAN_BASELINES[2]

OPTIMIZER_IGNORE_HINTS

OPTIMIZER_IGNORE_PARALLEL_HINTS

PLSCOPE_SETTINGS

PLSQL_CCFLAGS

PLSQL_DEBUG

PLSQL_OPTIMIZE_LEVEL

PLSQL_WARNINGS

TIME_ZONE[2]

NLS_*

[1]System level only [2]Session level only

trivadis makes IT easier.

Resource Manager

A plan (DWCS_PLAN) with three consumer groups is pre-configured

Consumer Group	CPU	Session Pool	PX Server Limit	DOP Limit
HIGH	4	Unlimited	50	CPU_COUNT[1]
MEDIUM	2	Unlimited	84	4
LOW	1	2*CPU_COUNT[1]		1
OTHER_GROUPS	1	Unlimited		1

[1]When several instances are used, CPU_COUNT <> “CPU Core Count”

trivadis makes IT easier.

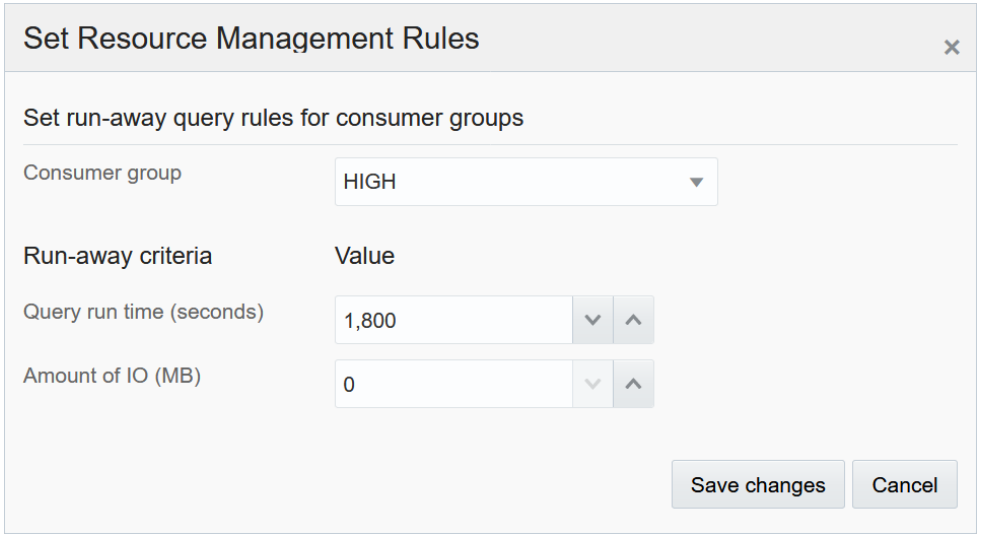
Resource Manager – Run Away Queries

Except for OTHER_GROUPS, two thresholds to configure when queries are cancelled can be set

- Run time in seconds
- Amount of disk I/O in MB

Remark:

- 0 = detection disabled



trivadis makes IT easier.

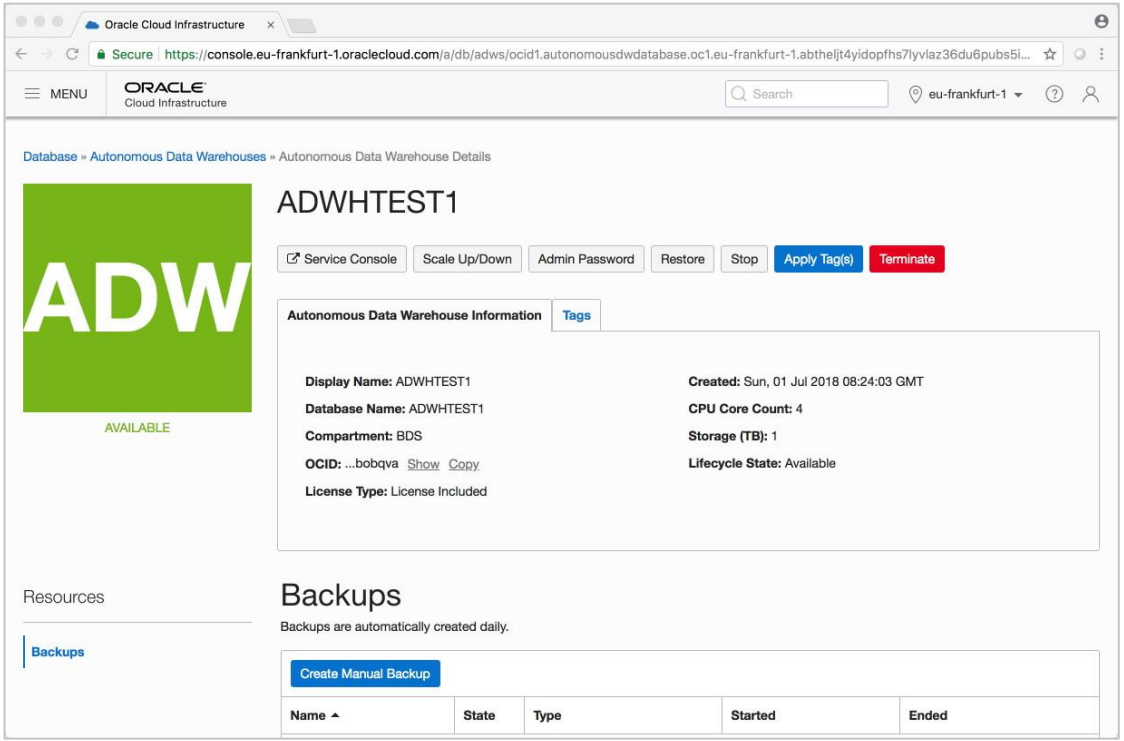
Connectivity

trivadis makes IT easier.

Service Management via Browser

Used for

- Service Console
- Start/Stop DB
- Scale up/down
- Restore
- Management of credential



trivadis makes IT easier.

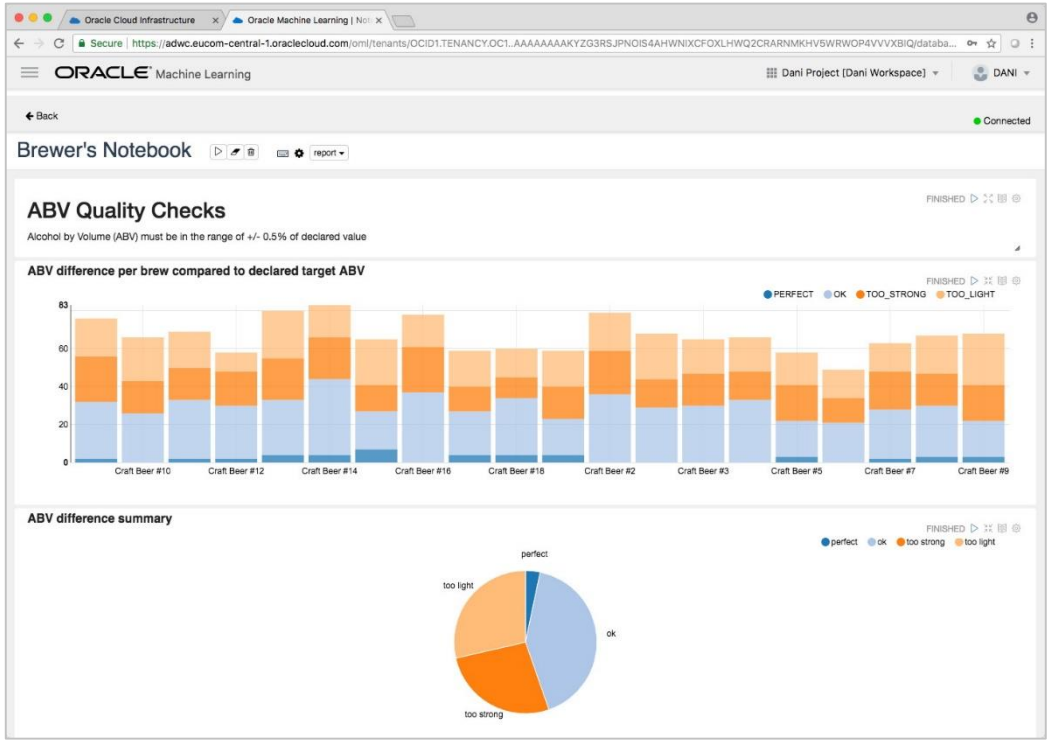
Oracle Machine Learning

Notebook style application for advanced SQL users

- Apache Zeppelin
- Interactive data analysis
- Graphical reports

Typical Users

- Data scientists
- Developers
- Business users



trivadis makes IT easier.

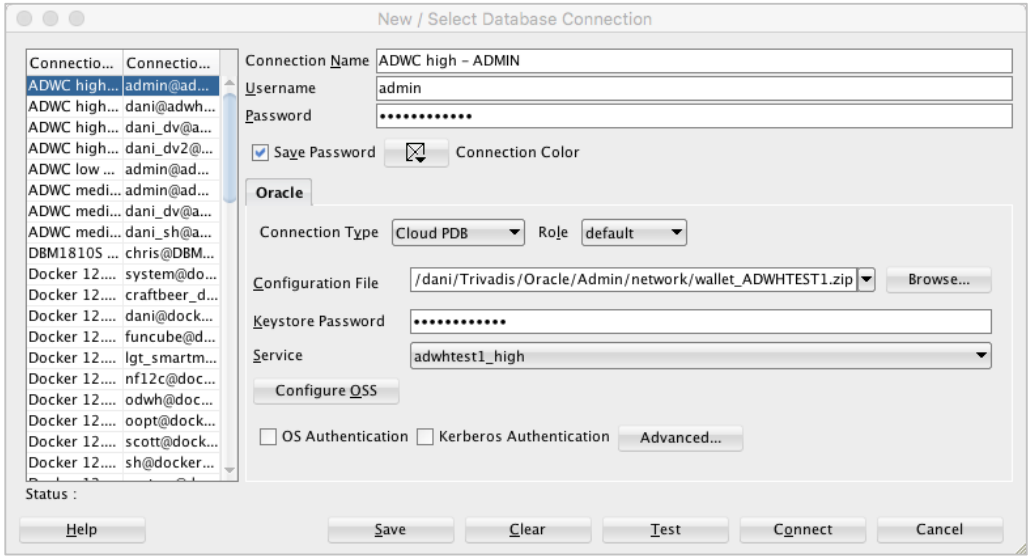
Connection with SQL Developer

Download client credential file

- Via Service Console

Create new connection

- Type “Cloud PDB”
- Import credential file
- Enter keystore password (only for versions < 18.2)
- Select service level (LOW, MEDIUM, HIGH)

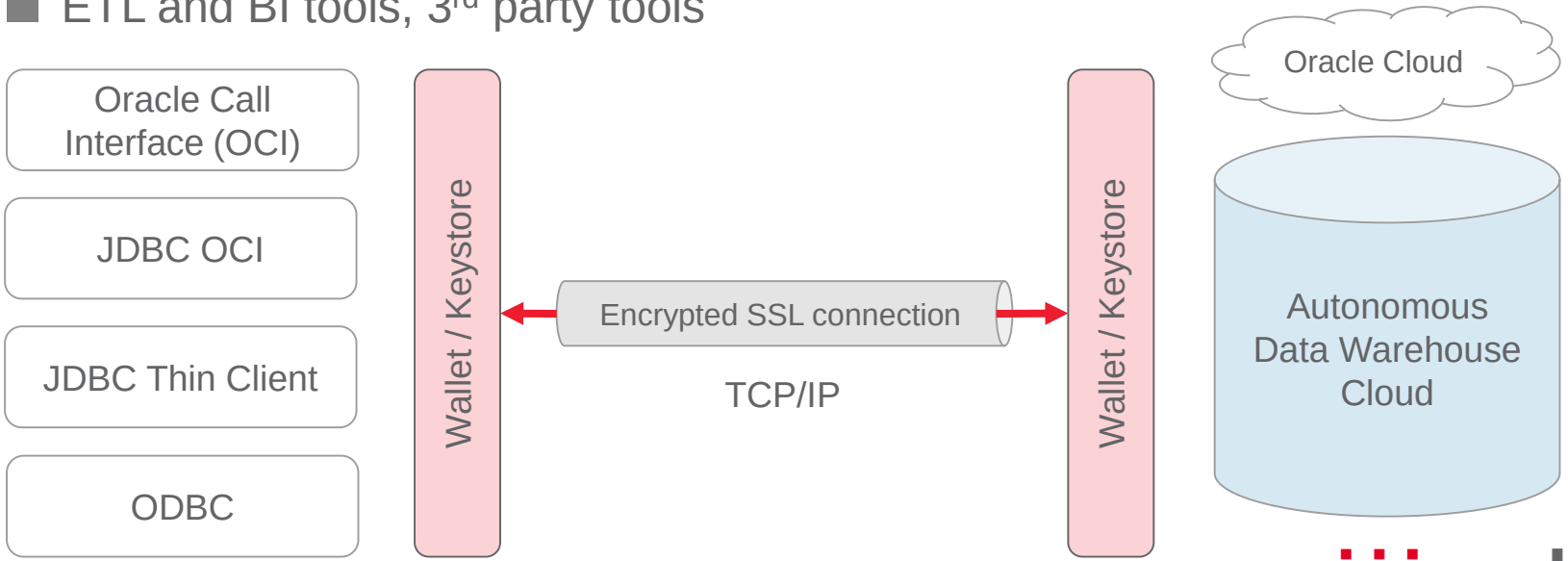


trivadis makes IT easier.

Manual Connection Configuration

Connections using Oracle Net Services (SQL*Net)

- For SQL*Plus, SQLcl, Toad, ...
- ETL and BI tools, 3rd party tools

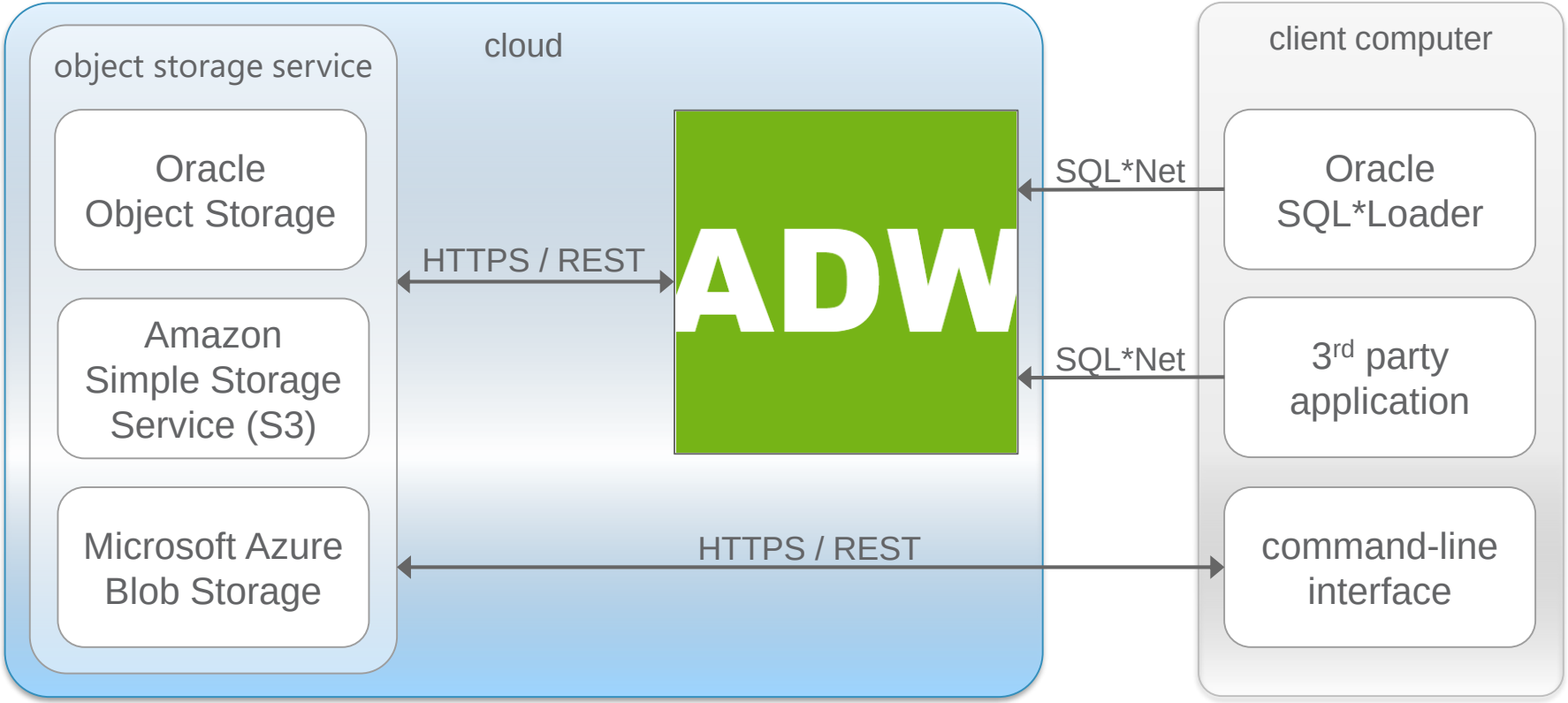


trivadis makes IT easier.

Loading Data

trivadis makes IT easier.

Loading Data



trivadis makes IT easier.

DBMS_CLOUD

It provides features to

- manage credentials to access an OSS
- manage external tables that can be used to query data stored in an OSS
- handle objects stored in an OSS
- handle files stored on ADW in DATA_PUMP_DIR

Data Pump 12.2 supports importing (but not exporting) a file stored in an OSS

```
ALTER DATABASE PROPERTY SET default_credential = 'ADMIN.CHRIS'
```

```
impdp dumpfile=default_credentials:https://…/test.dmp
```


trivadis makes IT easier.

Example – Loading Data into a 1TB TPC-DS Schema (1)

Step 1: Upload the text files from the file system of a server hosted in the cloud (same region) to the Oracle's object storage

```
oci os object bulk-upload --bucket-name tpcds
                          --src-dir /data/tpcds
                          --include "*.dat"
```

- By default, such a statement uses 10 parallel threads
- The load took 6 hours

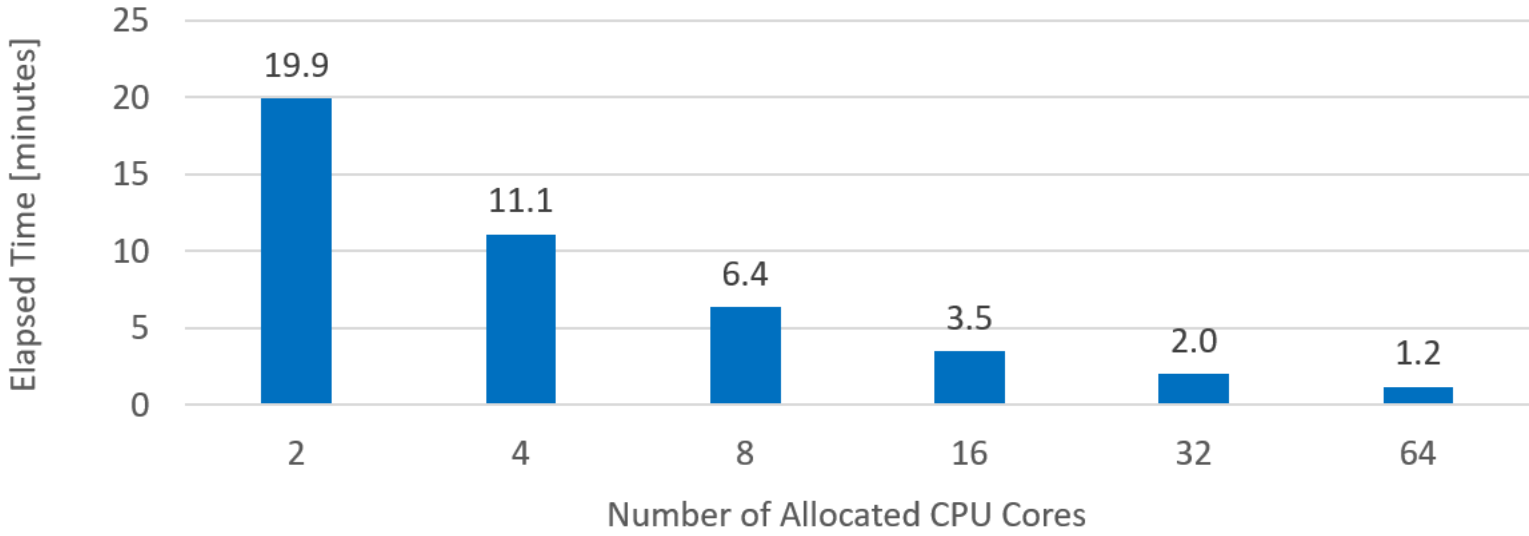
Step 2: Load the data into the database through DBMS_CLOUD

- The number of allocated CPU cores determines how fast data is loaded
- With 32 cores it took 67 minutes

trivadis makes IT easier.

Example – Loading Data into a 1TB TPC-DS Schema (2)

Example of scalability: load 22GB into the CATALOG_RETURNS table



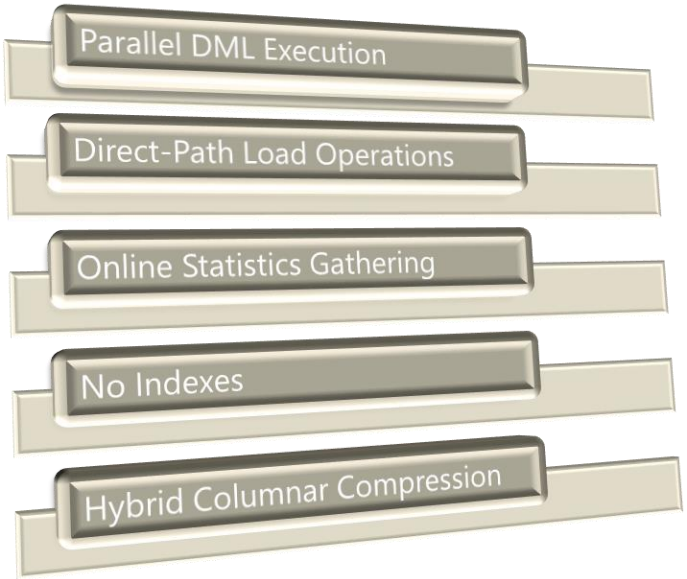
source: https://antognini.ch/2018/07/observations-about-the-scalability-of-data-loads-in-adwc/

trivadis makes IT easier.

ETL Performance

trivadis makes IT easier.

Prepared for High ETL Performance



Good ETL performance is possible

- Setup is optimized for DWH and ETL
- But details must be considered

trivadis makes IT easier.

Parallel DML Execution and Direct-Path Operations

Parallel DML (PDML) is enabled by default

- Only if CPU core count > 1 and consumer group is *MEDIUM* or *HIGH*
- Hint /*+ *parallel* */ can be added (usually not needed)
  - Set *optimizer_ignore_parallel_hints* = *FALSE* (default: *TRUE*)

Direct-Path INSERT is used

- For parallel DML
- If hint /*+ *append* */ is added

trivadis makes IT easier.

Parallel DML Execution and Consumer Groups

```
-------------------------------------------------------------------------------------------
| Id  | Operation                          | Name        |    TQ  |IN-OUT| PQ Distrib |
-------------------------------------------------------------------------------------------
|   0 | INSERT STATEMENT                   |             |        |      |            |
|   1 |  LOAD AS SELECT                    | CUSTOMERS_1 |        |      |            |
|   2 |   PX COORDINATOR                   |             |        |      |            |
|   3 |    PX SEND QC (RANDOM)             | :TQ10000    |  Q1,00 | P->S | QC (RAND)  |
|   4 |     OPTIMIZER STATISTICS GATHERING |             |  Q1,00 | PCWC |            |
|   5 |      PX BLOCK ITERATOR             |             |  Q1,00 | PCWC |            |
|   6 |       TABLE ACCESS STORAGE FULL    | CUSTOMERS   |  Q1,00 | PCWP |            |
-------------------------------------------------------------------------------------------

   - automatic DOP: Computed Degree of Parallelism is 8
```

HIGH

MEDIUM

```
-------------------------------------------------
| Id  | Operation                  | Name        |
-------------------------------------------------
|   0 | INSERT STATEMENT           |             |
|   1 |  LOAD TABLE CONVENTIONAL   | CUSTOMERS_1 |
|   2 |   TABLE ACCESS STORAGE FULL| CUSTOMERS   |
-------------------------------------------------

   - automatic DOP: Computed Degree of Parallelism is 1 because of no expensive parallel operation
   - PDML disabled because object is not decorated with parallel clause
   - Direct Load disabled because no append hint given and not executing in parallel
```

LOW

trivadis makes IT easier.

Online Statistics Gathering for Direct-Path Loads

Statistics are gathered automatically

Unlike 12c, this works also

- for non-empty tables
- for histograms

```
-----------------------------------------------------
| Id | Operation                          | Name     |
-----------------------------------------------------
|  0 | INSERT STATEMENT                   |          |
|  1 |  LOAD AS SELECT                    | TARGET   |
|  2 |   PX COORDINATOR                   |          |
|  3 |    PX SEND QC (RANDOM)             | :TQ10000 |
|  4 |     OPTIMIZER STATISTICS GATHERING |          |
|  5 |      PX BLOCK ITERATOR             |          |
|  6 |       TABLE ACCESS STORAGE FULL    | SOURCE   |
-----------------------------------------------------
```

Two new undocumented parameters

- *_optimizer_gather_stats_on_load_all* (default: TRUE)
- *_optimizer_gather_stats_on_load_hist* (default: TRUE)

trivadis makes IT easier.

Statistics Gathering for Conventional Loads

Attention: statistics are not gathered automatically

Call DBMS_STATS with default values

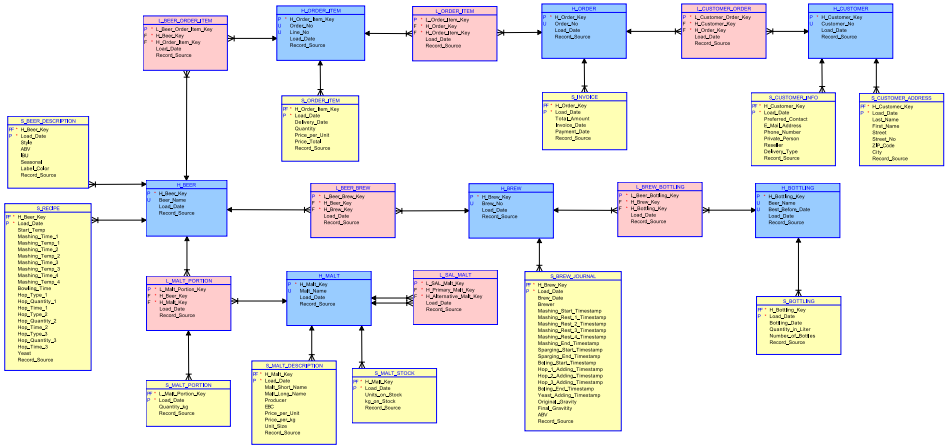
Automatic Statistics Gathering job is enabled, but maintenance windows are disabled

trivadis makes IT easier.

Example: Loading Data Vault Schema

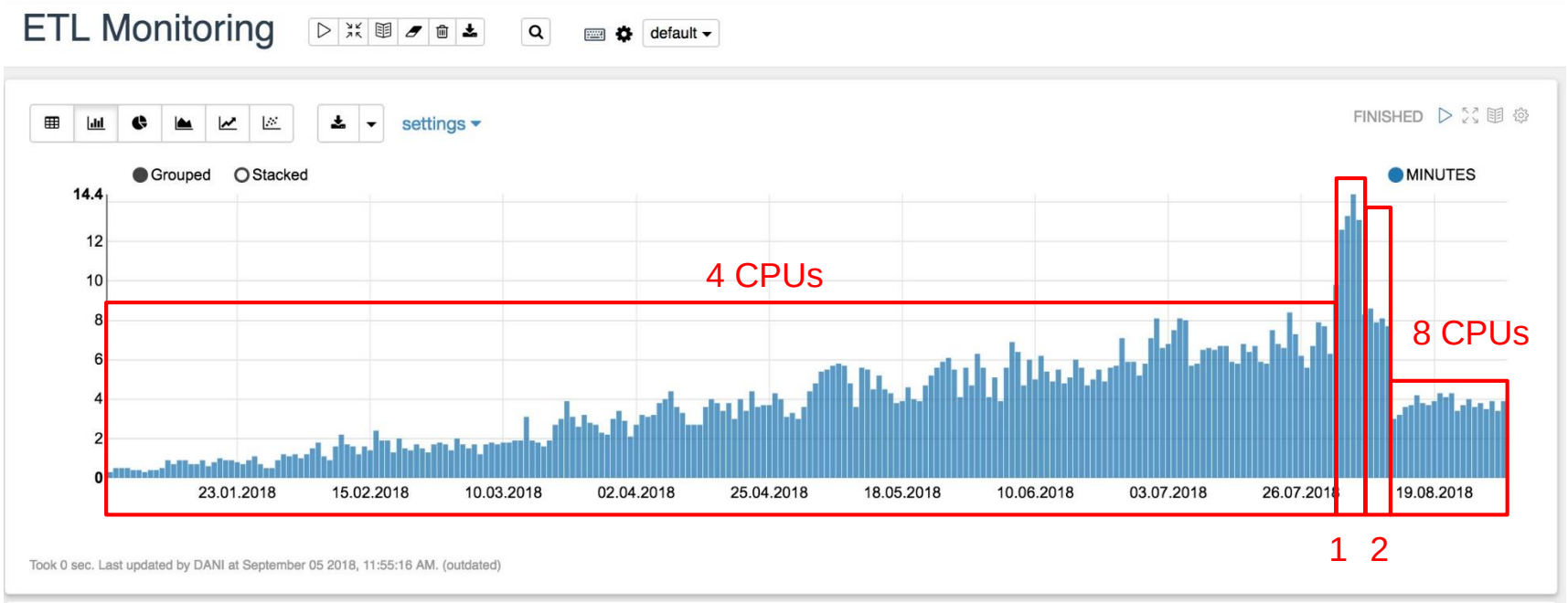
Data Vault Model of Craft Beer Brewery

- 25 target tables (Hubs, Links, Satellites)
- 20 beers, 272K customers
- 16M orders, 173M order items
- “Daily loads” for 8 months
  - Random data generator
  - Data Vault load patterns
  - Delta detection and versioning
  - 1.4M order items / day



trivadis makes IT easier.

Example: Loading Data Vault Schema



trivadis makes IT easier.

Query Performance

trivadis makes IT easier.

Result Cache

It is enabled by default on ADW

- RESULT_CACHE_MODE = FORCE

For queries being re-executed, it can lead to a tremendous performance improvement

To avoid caching, use the NO_RESULT_CACHE hint

- Even if OPTIMIZER_IGNORE_HINTS = TRUE (default on ADW)

trivadis makes IT easier.

Improve Query Performance

What cannot be done

- Use In-Memory Column Store

What should not be done

- Partition tables
- Create materialized views
- Create indexes

What can be done

- Scale up the number of CPU cores
- Use different service
- Use constraints to enable query transformations (e.g. join elimination)

trivadis makes IT easier.

How Are Queries on a Star Schema Optimized?

The star transformation is disabled (star_transformation_enabled=FALSE)

Instead, the query optimizer can use the vector transformation

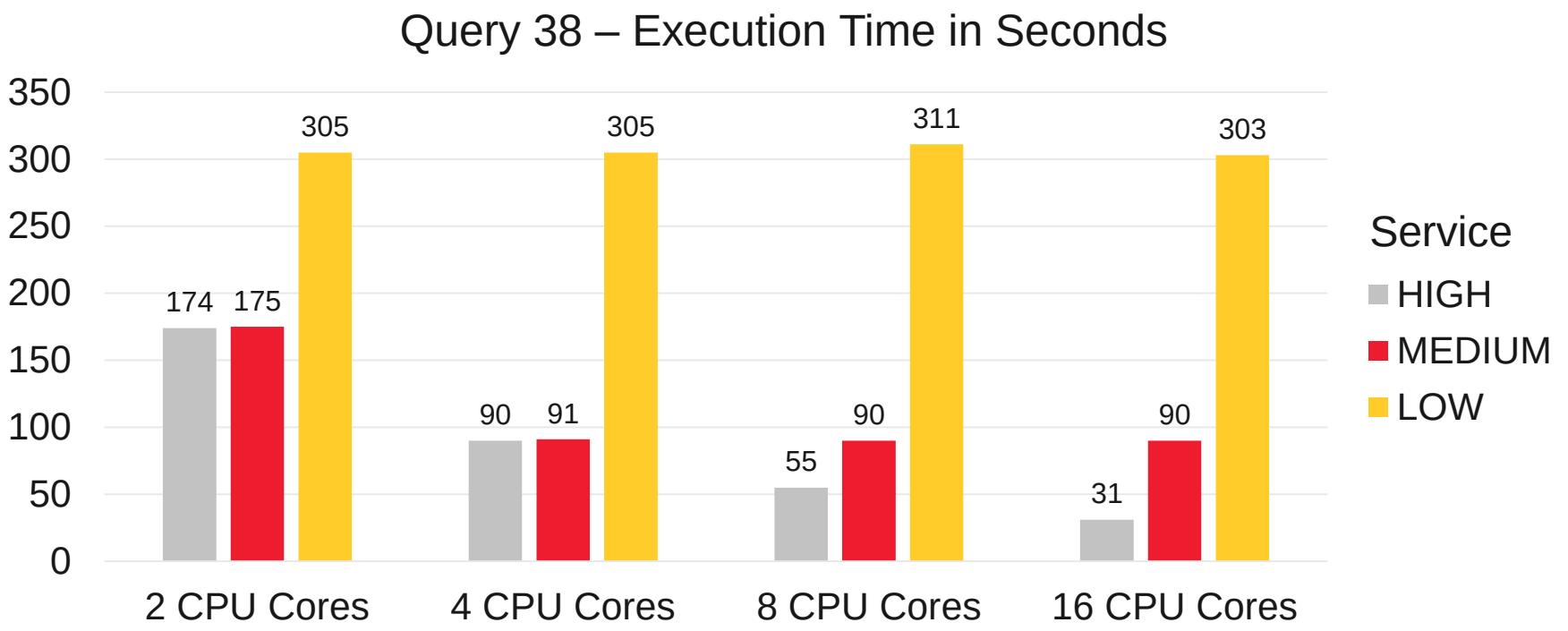
- Introduced in 12.1.0.2 for *In-Memory Aggregation*
- In many situations, faster than star transformation

```
Note
-----
   - vector transformation used for this statement
```

- Works even if tables are not populated in IMCS
- In-Memory is enabled (but not used) in ADW
  - INMEMORY_SIZE = 1073741824

trivadis makes IT easier.

Example – TPC-DS Queries

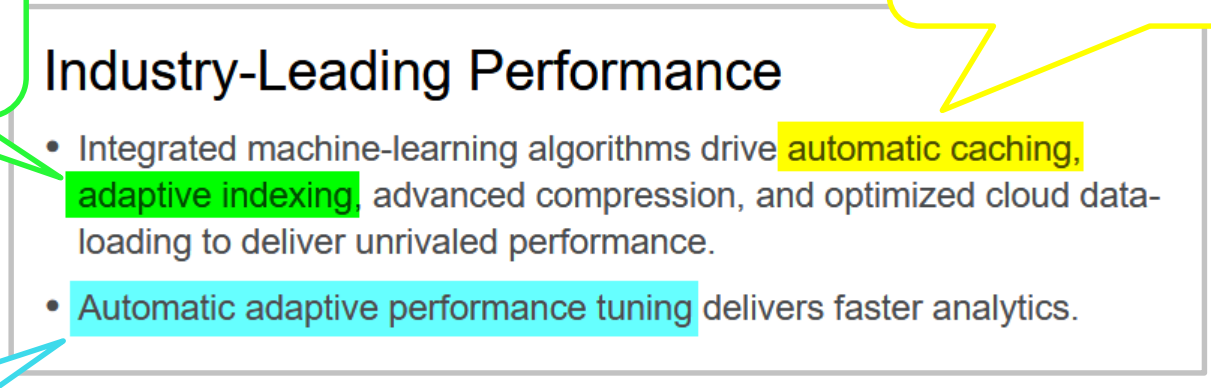


trivadis makes IT easier.

Automated Tuning?

Exadata storage indexes?

Result cache?
(in addition to disk I/O caching)



source: https://www.oracle.com/database/data-warehouse.html

Not observed ☹

trivadis makes IT easier.

Monitoring Performance

trivadis makes IT easier.

Monitoring Performance and SQL Statements

Service Console with monitoring capabilities:

- Activity and utilization
  - Storage and CPU usage
  - Real-time or time period
- Running SQL statements
  - Time & wait statistics, I/O statistics
  - Runtime execution plan
  - Parallel processes
  - Downloadable real-time SQL Monitoring report

Default 8 days, can be changed with AWR settings

trivadis makes IT easier.

Autonomous Database | Activi

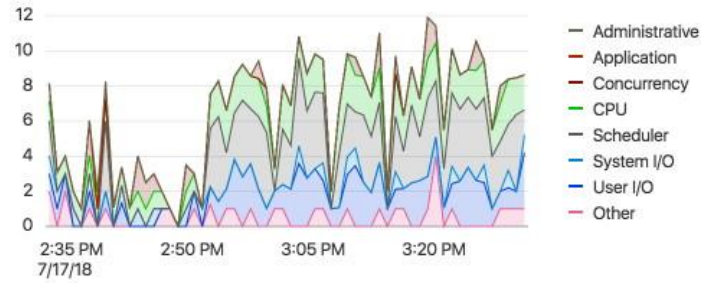
Secure https://adwc.eucom-central-1.oraclecloud.com/console/?nav=activity

ORACLE
Autonomous Data Warehouse Cloud

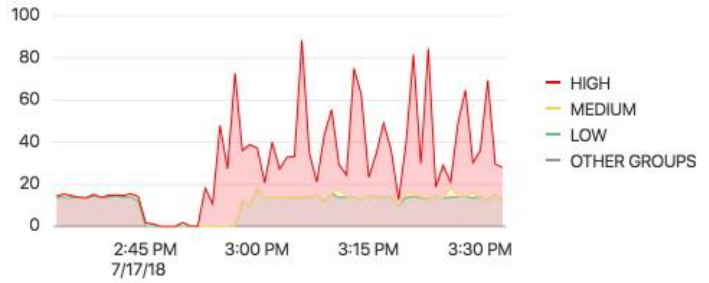
ADWHTEST1

- Overview
- Activity
- Administration

Database Activity



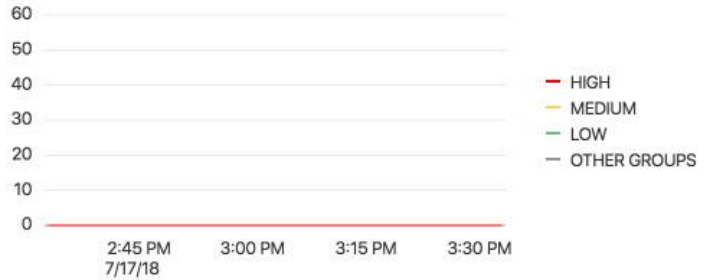
CPU Utilization (%)



Running Statements



Queued Statements



COPYRIGHT © 2018 ORACLE AND/OR ITS AFFILIATES. ALL RIGHTS RESERVED.

Miscellaneous

trivadis makes IT easier.

Backup & Recovery

Automatic (incremental) backups take place daily

- The start time cannot be set
- The retention period for automatic backups is 60 days

Manual backups can be initiated through the console

- They are stored in the object store
- The DEFAULT_CREDENTIAL and DEFAULT_BUCKETS database properties must be set

Recovery at any point-in-time can be initiated through the console

trivadis makes IT easier.

Patching

Oracle is patching the service on a regular basis

- No announcement
- No downtime

There is no way to “schedule” when the installation of the patches takes place

trivadis makes IT easier.

Oracle Support

As soon as a problem cannot be solve because of missing privileges, an SR has to be opened

Support has (almost) no visibility

- E.g. no access to the alert.log of the PDB
- E.g. no OS access

Support relies on Operations to fix things

The issues we experienced were fixed in a time frame going from one days to one week

trivadis makes IT easier.

Conclusion

trivadis makes IT easier.

Conclusion

- Appropriate setup for DWH
- Better than many manual configured DWHs
- Easy (limited) administration, ready to use
- No In-Memory (yet)
- Support in case of a service request can take too long
- Sometimes “shaky” (no control about patches / changes)
- Not as simple as it looks at first sight
- Knowhow about physical DB design still important







trivadis
makes IT easier.

Further Information in Blog Posts

- DBMS_CLOUD Package – A Reference Guide
  https://antognini.ch/2018/07/dbms_cloud-package-a-reference-guide/
- Which Privileges Are Required to Use the ADWC Service Console?
  https://antognini.ch/2018/07/which-privileges-are-required-to-use-the-adwc-service-console/
- Observations About the Scalability of Data Loads in ADWC
  https://antognini.ch/2018/07/observations-about-the-scalability-of-data-loads-in-adwc/
- External Tables in Autonomous Data Warehouse Cloud
  https://danischnider.wordpress.com/2018/07/04/external-tables-in-autonomous-data-warehouse-cloud/
- Gathering Statistics in the Autonomous Data Warehouse Cloud
  https://danischnider.wordpress.com/2018/07/11/gathering-statistics-in-the-autonomous-data-warehouse-cloud/
- 10 Tips to Improve ETL Performance – Revised for ADWC
  https://danischnider.wordpress.com/2018/07/20/10-tips-to-improve-etl-performance-revised-for-adwc/
- Star Schema Optimization in Autonomous Data Warehouse Cloud
  https://danischnider.wordpress.com/2018/09/13/star-schema-optimization-in-autonomous-data-warehouse-cloud/

trivadis makes IT easier.

Questions and Answers…

Christian Antognini
Dani Schnider



@chrisantognini antognini.ch/blog
@dani_schnider danischnider.wordpress.com

trivadis
makes IT easier.