



Oracle Data Analytics and Autonomous Data Warehouse service on the Cloud

CERN openlab Technical Workshop

Manuel Martin Marquez

23/01/2018

Oracle Autonomous Technologies



Self-Driving – Automates all database and infrastructure management, monitoring, tuning



Self-Securing – Protects from attacks



Self-Repairing – Minimizes or eliminates downtimes including planned maintenance



Reduce Costs
Risks
Focus on Innovation

Oracle Autonomous Technologies and ML



Infrastructure Management

Detection and recovery
offailed/sick server
storage orswitch/link



Operations

Hang Management
Anomaly Detection
Maintenance Slot Identification
Bug Identification and
Prioritization



Workload Optimizations

Query Optimizer
Real-time statistics
Automatic Indexing

Oracle Autonomous Technologies



Autonomous Data Warehouse (ADW)

Columnar
Data Summaries
Memory speeds Joins, Aggs, etc
Real-time Statistics



Autonomous Transaction Processing (ATP)

Row format
Indexes
Memory for Caching to avoid IO
Real-time Statistics

ADW – CERN Industrial IoT (SCADA)



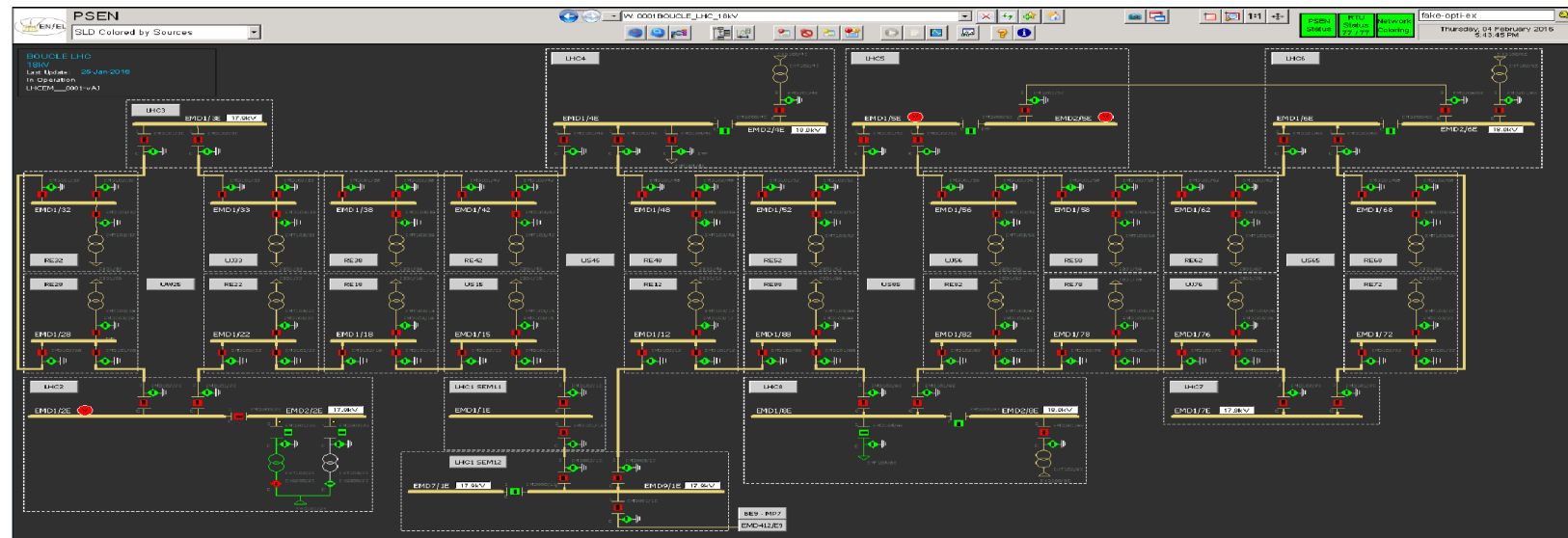
PSEN schema - the first step

- 750GB of SCADA data about Electrical Network
- Contains IOT partitioned tables
- One big IOT partitioned table of 620G



Improve performance of most challenging data retrieval and analytics scenario

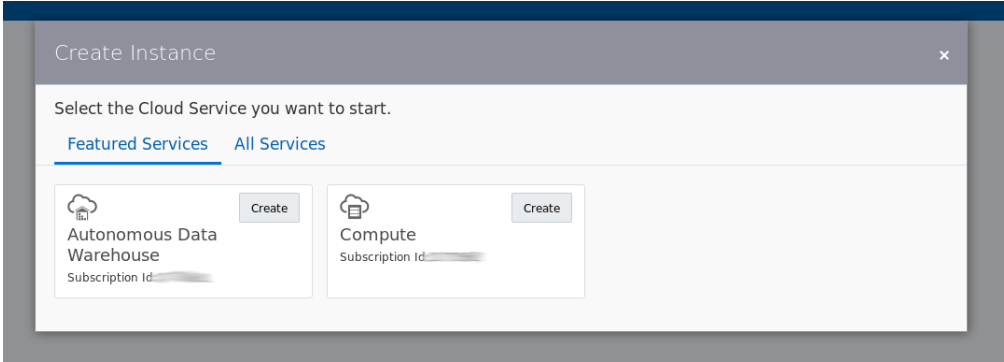
- Trends
- Events
- Alarms
- Historical event



ADW – CERN Industrial IoT (SCADA)



Create the instance



Create Autonomous Data Warehouse [help](#) [cancel](#)

COMPARTMENT
[Redacted]

Oracle recommends that you create this resource in a compartment other than the root. [Learn why.](#)

DISPLAY NAME
DB 201810161801

DATABASE NAME
DB201810161801
The name must contain only letters and numbers, starting with a letter. 14 characters max.

CPU CORE COUNT
1
The number of CPU cores to enable. Maximum cores per database: 128. Available cores are subject to your tenancy's service limits.

STORAGE (TB)
1
The available storage, up to 128 TB.

Administrator Credentials

Set the password for your Autonomous Data Warehouse ADMIN user here.

USERNAME READ-ONLY
ADMIN

PASSWORD

CONFIRM PASSWORD

LICENSE TYPE

MY ORGANIZATION ALREADY OWNS ORACLE DATABASE SOFTWARE LICENSES
Bring my existing database software licenses to the database cloud service [\(details\)](#).

SUBSCRIBE TO NEW DATABASE SOFTWARE LICENSES AND THE DATABASE CLOUD SERVICE

TAGS

Tagging is a metadata system that allows you to organize and track resources within your tenancy. Tags are composed of keys and values that can be attached to resources.

[Learn more about tagging](#)

| TAG NAMESPACE | TAG KEY | VALUE |
|------------------------------|---------|-------|
| None (apply a free-form tag) | | |

[Create Autonomous Data Warehouse](#) [+ Additional Tag](#)

[Create Autonomous Data Warehouse](#)

ADW – CERN Industrial IoT (SCADA)



Create the instance (CLI)

Oracle Cloud Infrastructure console screenshot showing the Autonomous Data Warehouse page. The page title is "Autonomous Data Warehouses in Compartment xxx". The "COMPARTMENT" dropdown is set to "Compartment xxx". The "STATE" dropdown is set to "Any state". A table lists two Autonomous Data Warehouses:

| Name | State | Database Name | CPU Core Count | Storage |
|------------------------|-----------------|---------------|----------------|---------|
| adwc1 | Provisioning... | adwc1 | 1 | 1 |
| PSENDB | Available | PSENDB | 6 | 10 |

At the bottom of the table, it says "Displaying 2 Autonomous Data Warehouses < Page 1 >".

```
2. oracle@itdbsma:~/oci (ssh)
[oracle@itdbsma:~/oci]$ oci db autonomous-data-warehouse create --from-json-file:///home/oracle/.oci/request.json
{
  "data": {
    "compartment-id": "ocid1.tenancy.oc1.aaaaaaaanssic...",
    "connection-strings": null,
    "cpu-core-count": 1,
    "data-storage-size-in-tbs": 1,
    "db-name": "adwc1",
    "defined-tags": {},
    "display-name": "adwc1",
    "freeform-tags": {},
    "id": "ocid1.autonomousdwdatabase.oc1.eu-frankfurt-1.aaaaaaaanssic...",
    "license-model": "LICENSE_INCLUDED",
    "lifecycle-details": null,
    "lifecycle-state": "PROVISIONING",
    "service-console-url": null,
    "time-created": "2018-10-16T20:29:59.548000+00:00"
  },
  "etag": "b9eb2a89"
}
[oracle@itdbsma:~/oci]$ cat request.json
{
  "compartment-id": "ocid1.tenancy.oc1.aaaaaaaanssic...",
  "db-name": "adwc1",
  "display-name": "adwc1",
  "admin-password": "ad...",
  "cpu-core-count": 1,
  "data-storage-size-in-tbs": 1,
  "license-model": "LICENSE_INCLUDED"
}
[oracle@itdbsma:~/oci]$
```

ADW – CERN Industrial IoT (SCADA)



That is required to upload data pump export files

Oracle Cloud Infrastructure console screenshot showing the 'Create Bucket' dialog. The dialog includes fields for 'BUCKET NAME', 'STORAGE TIER' (Standard selected), and 'TAGS'. The background shows the Object Storage page with a list of buckets.

```
[oracle@itdbsma .oci]$ oci os bucket create --name test_bucket
{"data": {
  "compartment-id": "ocid1.tenancy.oc1..aa...",
  "created-by": "ocid1.user.oc1..aaaaaaa...",
  "defined-tags": {},
  "etag": "bf1e0fef-ffe7-4d5e-84d2-39e39e6f018b",
  "freeform-tags": {},
  "kms-key-id": null,
  "metadata": {},
  "name": "test_bucket",
  "namespace": "cernopenlab1",
  "object-lifecycle-policy-etag": null,
  "public-access-type": "NoPublicAccess",
  "storage-tier": "Standard",
  "time-created": "2018-10-17T09:29:49.567000+00:00"
},
"etag": "bf1e0fef-ffe7-4d5e-84d2-39e39e6f018b"
}
```


ADW – CERN Industrial IoT (SCADA)



Moving Data to ADWC

ADW principles: data is imported in its simplest form:

No indexes

No partitions

No IOTs

No materialized views

...

Data pump allows the needed transformations

ADW – CERN Industrial IoT (SCADA)



Moving Data to ADWC



Export Data

```
exclude=index, cluster, indextype, materialized_view, materialized_view_log, materialized_zonemap, db_link  
data_options=group_partition_table_data  
parallel=n  
schemas=schema_name  
dumpfile=export%u.dmp
```



Upload data to Object Storage

```
oci os object bulk-upload --bucket-name PSEN BUCKET 1 -- src-dir /mnt/oci/ - part-size 64 --parallel-upload-count 10
```



Import Data into ADW

- Parallel set to the number of CPUs you have
- Partitioned tables are converted to non-partitioned
- All segment attributes are ignored
- IOTs are converted to regular tables
- PK and unique indexes renamed to constraint name
- Same exclusion as during export

ADW – CERN Industrial IoT (SCADA)



Moving Data to ADWC



Import Data

Parallelism

| SID | EVENT | MODULE | SQL_TEXT |
|------|---|------------------|--|
| 4301 | wait for unread message on broa... udi@dbnile-clie... | | BEGIN :1 := sys.kupc\$que_int.get_status(:2, :3); END; |
| 6986 | wait for unread message on broa... Data Pump Master | Data Pump Master | BEGIN :1 := sys.kupc\$que_int.receive(:2); END; |
| 7344 | enq: TM - contention | Data Pump Worker | INSERT /*+ APPEND ENABLE_PARALLEL_DML PARALLEL("EVENTHISTORY_00000008",1)+*/ |
| 7879 | enq: TM - contention | Data Pump Worker | INSERT /*+ APPEND ENABLE_PARALLEL_DML PARALLEL("EVENTHISTORY_00000008",1)+*/ |
| 8419 | Datapump dump file I/O | Data Pump Worker | INSERT /*+ APPEND ENABLE_PARALLEL_DML PARALLEL("EVENTHISTORY_00000008",1)+*/ |
| 8596 | enq: TM - contention | Data Pump Worker | INSERT /*+ APPEND ENABLE_PARALLEL_DML PARALLEL("EVENTHISTORY_00000008",1)+*/ |

Importing Issues

```
... imported "PSEN"."EVENTHISTORY_00000008": "EVH_00000008_2017042800" 603.1 MB 9586464 rows
.. imported "PSEN"."EVENTHISTORY_00000008": "EVH_00000008_2015082700" 573.0 MB 9115136 rows
.. imported "PSEN"."EVENTHISTORY_00000008": "EVH_00000008_2015071100" 574.4 MB 9115083 rows
.. imported "PSEN"."EVENTHISTORY_00000008": "EVH_00000008_2016102300" 593.9 MB 9428783 rows
ORA-39014: One or more workers have prematurely exited.
ORA-39029: worker 10 with process name "DW09" prematurely terminated
ORA-31671: Worker process DW09 had an unhandled exception.
ORA-00600: internal error code, arguments: [4832], [0xABB5E23C8], [], [], [], [], [], [], [], []
ORA-06512: at "SYS.DBMS_STATS", line 1726
ORA-06512: at "SYS.DBMS_STATS", line 14728
ORA-06512: at "SYS.DBMS_STATS", line 31265
ORA-06512: at line 1
ORA-06512: at "SYS.DBMS_SQL", line 1721
ORA-06512: at "SYS.KUPD$DATA", line 1148
ORA-06512: at "SYS.KUPD$DATA", line 1252
ORA-06512: at "SYS.KUPD$DATA", line 3424
ORA-06512: at "SYS.KUPD$DATA", line 4719
ORA-06512: at "SYS.KUPD$DATA", line 6417
ORA-06512: at "SYS.KUPW$WORKER", line 22345
ORA-06512: at "SYS.KUPW$WORKER", line 5628
ORA-06512: at "SYS.KUPW$WORKER", line 13365
ORA-06512: at "SYS.KUPW$WORKER", line 2397
ORA-06512: at line 2
```

ADW – CERN Industrial IoT (SCADA)



Moving Data to ADWC



Import Data

Where are my logs?

Alert logs

Trace files

Data pump logs

```
select PAYLOAD from V$DIAG_TRACE_FILE_CONTENTS where  
2 TRACE_FILENAME='ehs1pod8_ora_115533.trc' order by line_number fetch  
3 first 15 rows only  
4 /
```

PAYLOAD

```
SET SERVEROUTPUT ON SIZE 1000000  
DECLARE  
  l_file          UTL_FILE.file_type;  
  l_location      VARCHAR2(100) := 'DATA_PUMP_DIR';  
  l_filename      VARCHAR2(100) := 'import_PSEN_03082018.log';  
  l_text          VARCHAR2(32767);  
BEGIN  
  -- Open file.  
  l_file := UTL_FILE.fopen(l_location, l_filename, 'r', 32767);  
  
  BEGIN  
    LOOP  
      UTL_FILE.get_line(l_file, l_text, 32767);  
      DBMS_OUTPUT.put_line( l_text) ;  
    END LOOP;  
  EXCEPTION  
    WHEN NO_DATA_FOUND THEN  
      NULL;  
  END;  
  
  -- Close the file.  
  UTL_FILE.fclose(l_file);  
END;  
/
```

```
-----  
-----  
s/ehs1pod/ehs1pod8/trace/ehs1pod8_ora_115533.trc  
on Release 12.2.0.1.0 - 64bit Production  
X64_181006  
uct/12.2.0.1/dbhome_1  
  
x86_64  
:41:01 PST 2018  
  
)  
  
: 8
```

ADW Insights



Comparison ADW and on-premise:

- Different Schemas
 - IoT Tables – Compressed Tables + PK Index
 - Executions plans changed a lot
- Very specific use case
 - But that is real-life



Missing Statistics (after data pump import):

- We gather them again
- Due to no use recommended parameters



High compression:

- **Hybrid Columnar Compression (HCC)** reduces tables size by a factor of 10
- Full scans are smaller!

ADW Insights



Indexes:

- Default scenario is not use indexes on ADW
- HCC reduces full scans workloads
- Exadata Smart Scan -
- Storage Indexes



Seamless provisioning and fully elastic

- CPU and storage can be adjusted online, at anytime in few seconds!
- So you can start with few resources and grow only if needed



Optimization:

- Automatic and transparent access to Oracle optimization features

Autonomous Next Steps



Deeper analysis on performance and scale data volumes

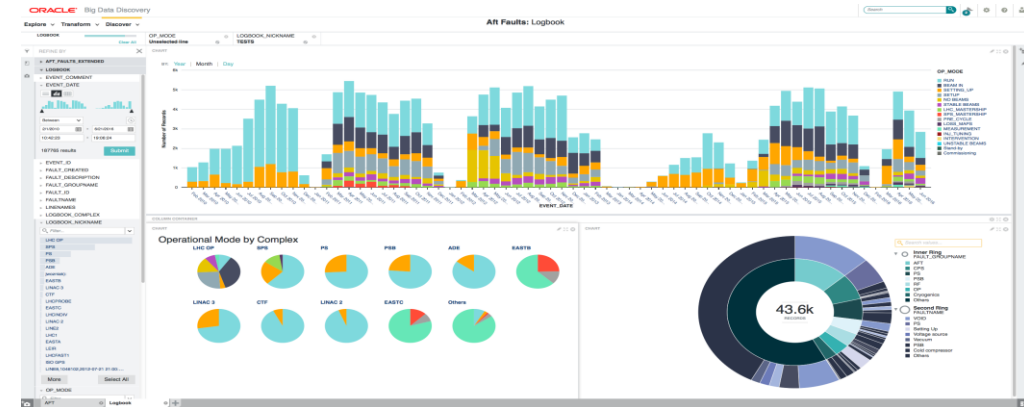
- Indexes
- HCC
- Exadata Smart Scan -
- Storage Indexes



Oracle Autonomous Transaction Processing



Oracle Autonomous Analytics Cloud Services:



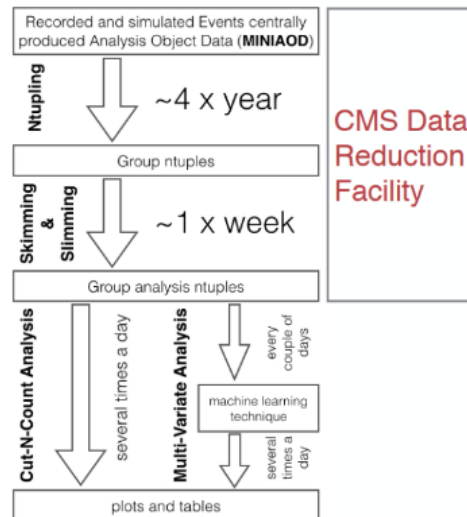
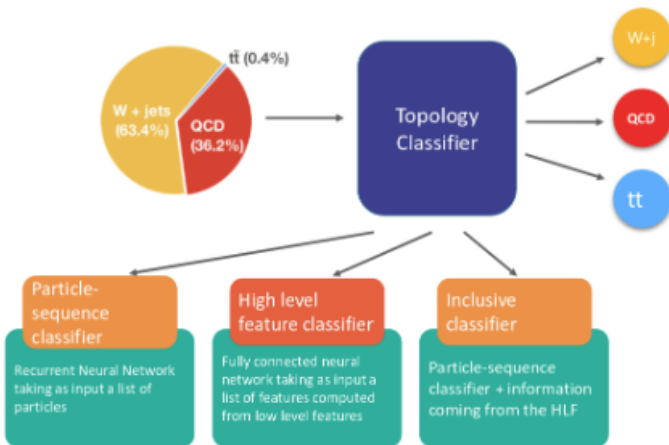
Machine Learning and Oracle Cloud Infrastructure (OCI)



The goal of the project is to deploy Physics Data processing and Machine learning on cloud resources, notably using CERN cloud and Oracle Cloud Infrastructure (OCI)



First Use cases



Data reduction from CMS Big Data project performing event selection and dimuon invariant mass calculations on 1 PB of data read from EOS



An event classifier ML pipeline “Topology classification with deep learning to improve real-time event selection at the LHC” (<https://arxiv.org/abs/1807.00083>)

Machine Learning and Oracle Cloud Infrastructure (OCI)



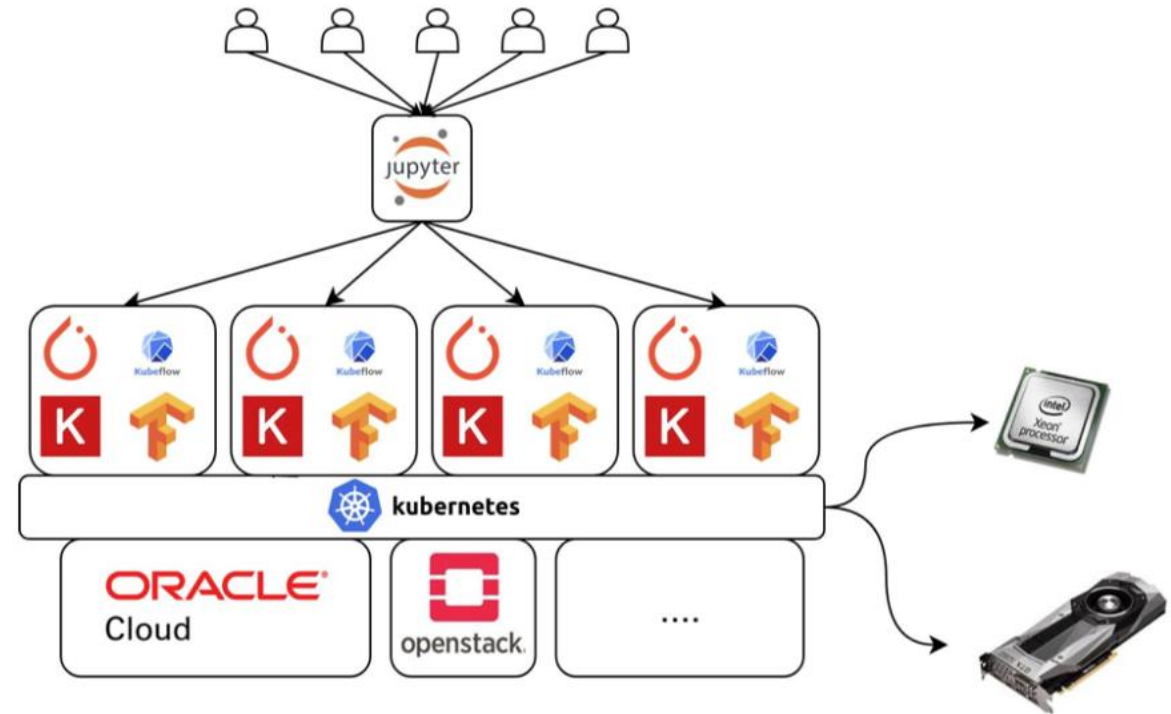
Scale rapidly from prototype



Share a cluster regardless of where it is deployed and of the chosen software stack



Fully featured environment, allows users to focus on analysis rather than data engineering and work collaboratively





QUESTIONS?

Manuel.Martin.Marquez@cern.ch