

# CERN openlab technical workshop Oracle Projects: Status and Plans

December 2016

Liana Lupsa and Antonio Nappi

# Topics



## Oracle Databases and monitoring

- The Oracle Cloud
- Oracle Database In-Memory
- Other openlab activities

## Data Analytics

- Data Visualization and Discovery
- Advanced Analytics
- Data Management & ETL

## Java and Virtualisation

- WebLogic
  - Current Status
  - Moving to containers
  - Future Architecture
- Oracle REST Data Service
- Oracle Traffic Director
- Oracle Java Cloud
- Oracle Backup Cloud

# The Oracle Cloud



## Why we want to test the Oracle Cloud?

- Assess the Cloud as possible disaster recovery solution
- Estimate effort needed for migration
- Feasibility of using the Cloud with the CERN environment
- Cost savings

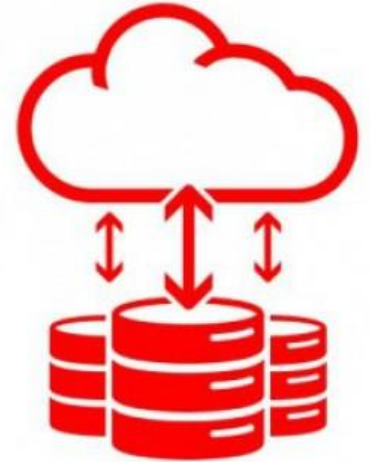


# The Oracle Cloud

## What do we want to test?

### › Database as a service

- Data Guard replication
  - Provisioning
  - Configuration management
  - Functional testing
  - Latency
  - Environment considerations
  - Needed space: size of selected applications + logs
- As a part of full stack of DB-based apps



# The Oracle Cloud

- › Database Schema Service
  - For the needs of Application testing
  - Any of the applications previously mentioned
  
- › Database backup
  - Repeat functional tests done with a data centre
  - Backup from ADG in cloud
  - Backup from DB in cloud
  - Backup size can be decreased if there is a need



# The Oracle Cloud

Initial kick-of meeting – 2 Dec



**CERNTest1**

Status: Stopped

Version: 11.2.0.4

Edition: Enterprise Edition

Submitted On: Dec 2, 2016 9:42:18 AM UTC

OCPUs: 1\*

Memory: 7.5 GB\*

Storage: 185 GB



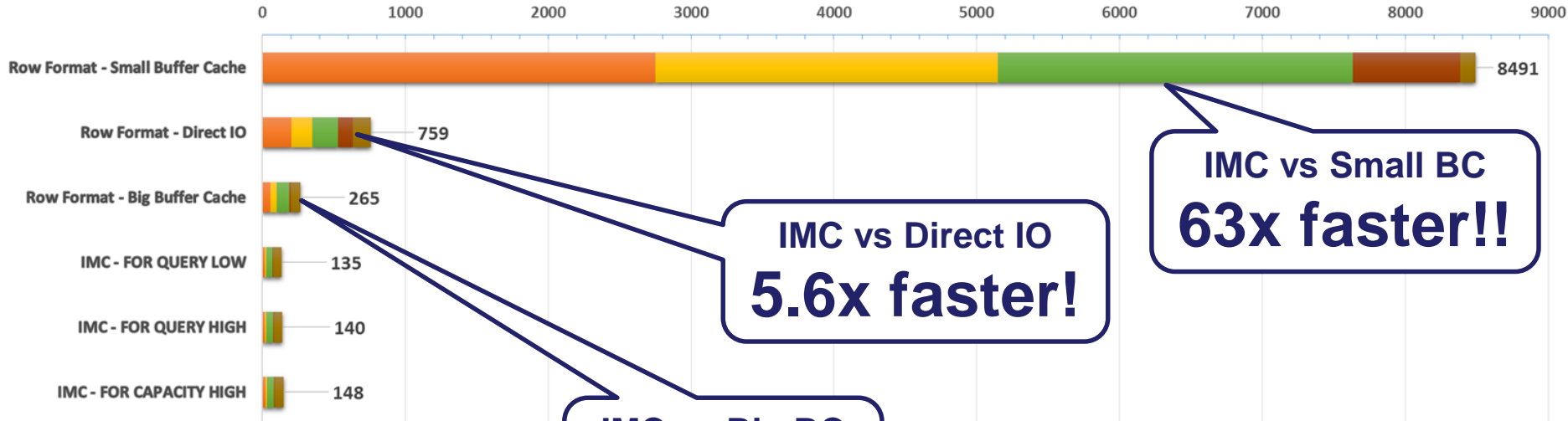
	Service Name	Service Type	Operation	Status	Start Time	End Time
▶	CERNTest1	Oracle Database Cloud Service	Stop Service	Succeeded	Dec 4, 2016 4:57:30 PM UTC	Dec 4, 2016 4:59:47 PM UTC

# Oracle Database In-Memory



- › Physics Data Analysis
- › **Administrative Data Warehouse**
  - HR data and personal records, financial data, orders / purchases, resource usage planning, electronic recruitment and many others
- › LHCb DIRAC bookkeeping system
  
- › **In-Memory tests - Real life queries from BI applications**
  - Captured from DB SQL history, and provided by application users
  - **Performance** testing – query response time under different DB configurations
    - **Big** and **small Buffer Cache** – 180GB / 32GB
    - Forced **Direct IO** (direct path read)
    - **In-Memory columnar** format - With different compression levels

## Queries on large data sets >10GB - response time (s)



	IMC - FOR CAPACITY HIGH	IMC - FOR QUERY HIGH	Row Format - Big Buffer Cache	Row Format - Direct IO	Row Format - Small Buffer Cache
Query 1 (ERP)	21.3	19.5	58.2	203.9	2748.5
Query 2 (ERP)	10.8	9.6	44.1	147.3	2400.1
Query 3 (ERP)	49.3	48.1	83.5	180.5	2483.6
Query 4 (FP)	6.5	2.1	12.3	101.7	751.4
Query 5 (ERP)	59.8	61.1	66.5	125.7	107.6

**IMC vs Big BC  
2x faster**

**IMC vs Direct IO  
5.6x faster!**

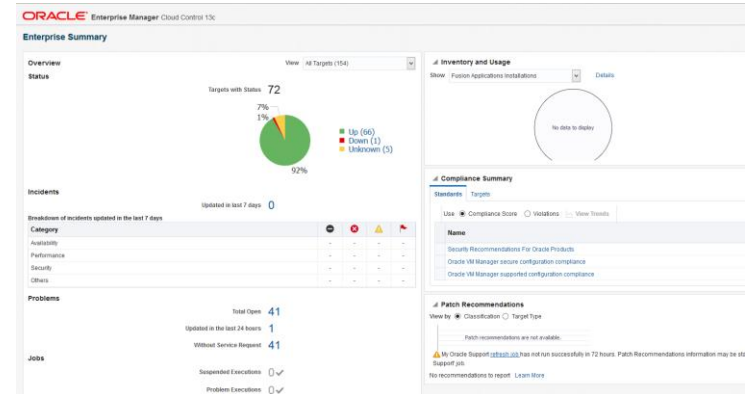
**IMC vs Small BC  
63x faster!!**



# openlab activities

- › Upgrade our monitoring system – **Enterprise Manager 13c**
  - Successfully managed to upgrade the test environment
  - Managed to configure Authentication with CERN Active Directory on the upgraded OEM
  - Configured Custom Certificates for OEM

- › Early access to Oracle Database **12.2**
  - Provided intensive feedback to Oracle



# Data Analytics Scenario

Control and Monitoring Systems



Intelligent, Predictive and Proactive Systems

Visualization &  
Discovery

Advanced Analytics

Data Management &  
ETL

# Visualization and Discovery

## Goals

Visualize any type of data  
Self-service data-lab (data analysed by domain experts)  
Collaborative environment

## Recent Activities

Deployment of [Oracle Big Data Discovery 1.3.2](#) on Exalytics machine  
Work with the UX (user experience) Oracle team for Big Data Discovery  
Use cases:

- [FCC RAMS](#) project (Reliability, Availability, Maintainability and Safety studies for the Future Circular Collider)

## Next steps

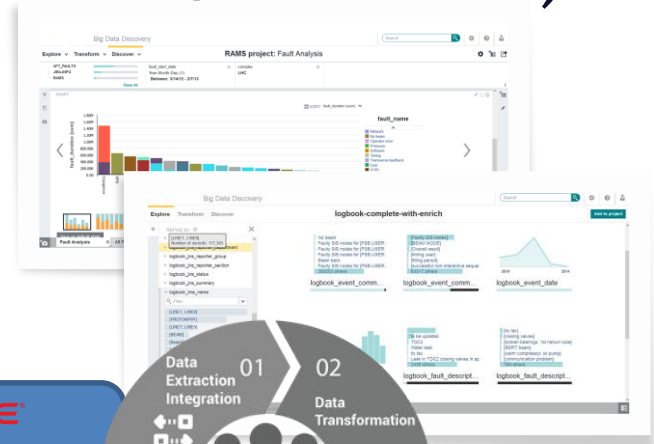
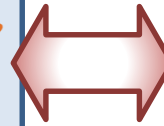
Integration of Kerberos with Oracle Big Data Discovery  
Notebooks for advanced analytics as part of the data-lab  
Early access to Oracle Big Data Discovery releases



# Discovery platform overview (Data-lab)

- Power Converters
- Cryogenics
- Machine Protection
- Accelerator Major Events
- Accelerator Fault Tracking
- Accelerator Logging
- Operations logbook

NoSQL  
XML  
JSON  
Text  
RDBMS



# Advanced Analytics

## Goals

- Predictive maintenance and system optimization (Batch)
- Real-time analytics

## Recent Activities

- Data analytics using Oracle R Advanced Analytics for Hadoop  
Implementation of **error analysis** for FCC reliability studies  
[http://openlab.cern/publications/technical\\_documents/oracle-r-technologies-data-analytics-and-machine-learning-hybrid](http://openlab.cern/publications/technical_documents/oracle-r-technologies-data-analytics-and-machine-learning-hybrid)
- Evaluation of analytic frameworks (H2O, PredictionIO, Oracle Stream Explorer)  
Real-time Postmortem analysis using Oracle Stream Explorer and Kafka  
[http://openlab.cern/publications/technical\\_documents/evaluation-intuitive-platforms-data-analytics](http://openlab.cern/publications/technical_documents/evaluation-intuitive-platforms-data-analytics)



# Advanced Analytics

## Ongoing work

- Oracle Machine Learning (OML) Notebook Beta program
- Data Integration - Kafka pilot service
  - Goal: Central IT Kafka service at CERN
  - Integration with streaming analytic frameworks  
Oracle Complex Event Processing and Stream Explorer
  - Use cases  
Streaming analytics for PostMortem analysis



# Data Management & ETL

## Goals

Integrate heterogeneous data repositories  
Facilitate ETL process for IT and CERN users

## Recent Activities

Developed framework to ingest data in Hadoop  
Self-service ETL with Oracle Big Data Discovery  
Developed application for Big Data Ingestion in HUE (Hadoop User Experience)  
[http://openlab.cern/publications/technical\\_documents/hue-application-big-data-ingestion](http://openlab.cern/publications/technical_documents/hue-application-big-data-ingestion)

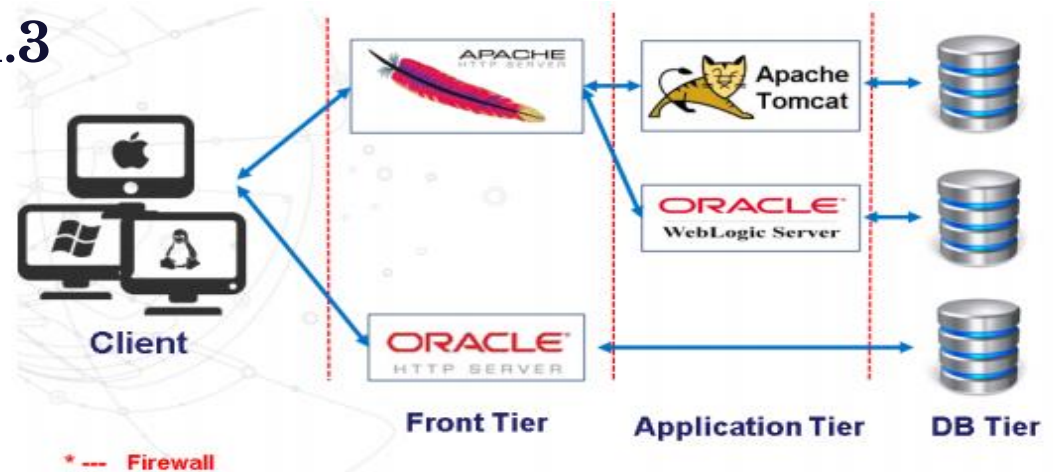


## Next steps

Evaluate hybrid RDBMS & Hadoop systems with Oracle Big Data SQL

# WebLogic : Numbers

- Oracle WebLogic 12.1.3
- ~258 Clusters
- ~530 JVMs
- ~105 VMs
- Users :
  - Engineering
  - Administration
  - IT
- Apache web server



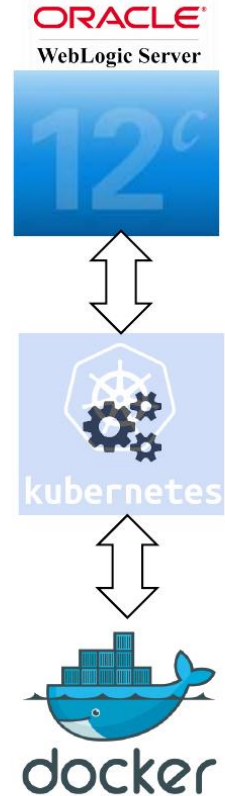
Credit: Sanadhi Sutandi @CERN summer student





# Why move to containers

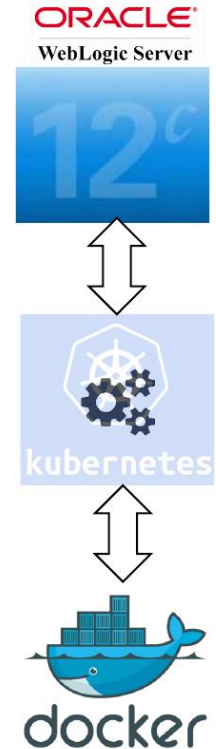
- **Increase developer productivity**
  - Simplify versioning control
  - Simple to maintain
  - Improve the testing since it is easy to reproduce.
- **Take advantages of WLS deployment via Docker containers :**
  - Easy to scale
  - Lightweight
  - Fast to deploy
  - Portability across machines





# Why move to containers

- **Migrate to WebLogic 12.2.1**
  - Take advantage of new features.
    - Partitions
    - Full REST API.
  - Enable Session Replication without affect the running infrastructure.
- **Make easier the deployment of WebLogic**
  - Run a complex application with least effort





# What we did

- **Set up Kubernetes cluster**
  - Started with physical machines, we moved to VMs.
- **Service IPs and Flannel for allow communication among pods.**
- **Custom images with custom scripts.**
  - One general image to pack
  - One custom for Admin and Managed Servers (Unpack and run script to configure)
- **Tried different architectures to allow the access to containers from external world.**



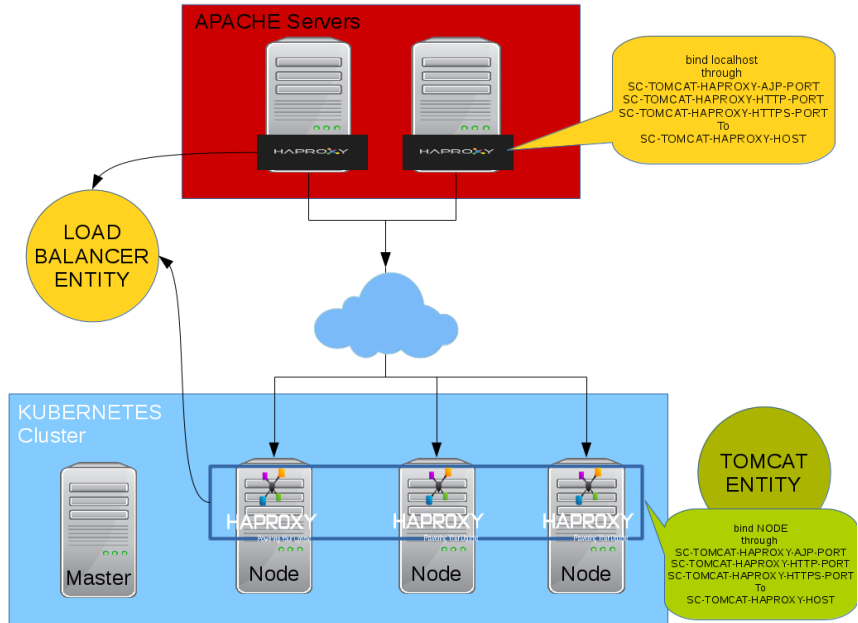
# Architecture Choice : Solution 1 HAProxy Load Balancer

## • PROS

- Already used with Tomcat proof of concept.
- Fast to understand that a node is up again.

## • CONS

- Not able to understand if the service it is available or not.
- Slow to switch from one broken node to another one.



Credit: Jose Daniel Arco @CERN



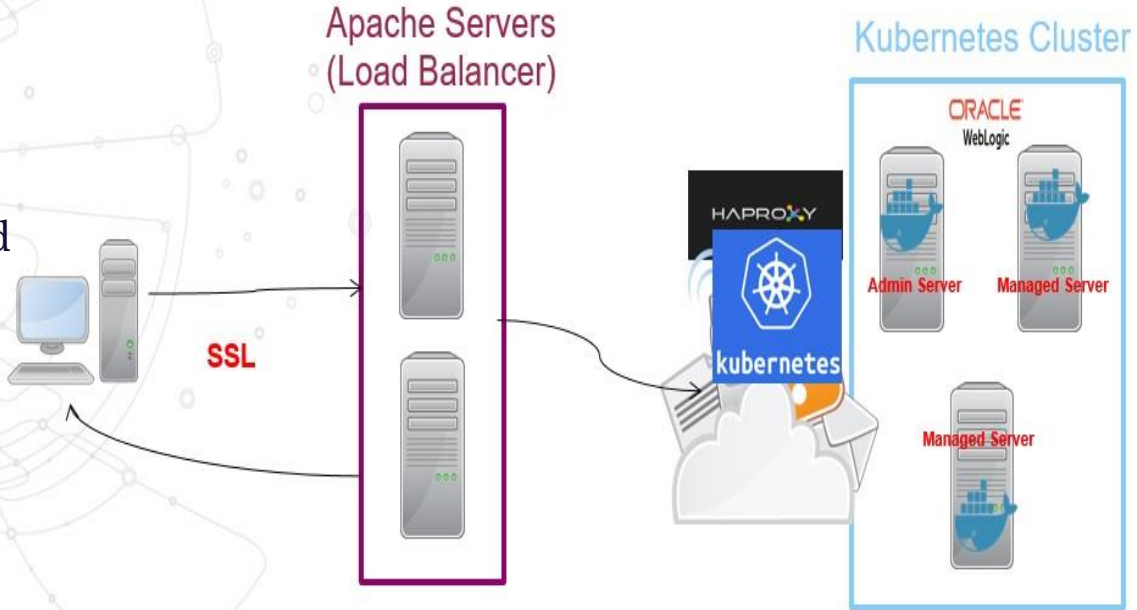
# Final Architecture : Solution 2 Apache Load Balancer

- **PROS**

- Use of the WebLogic proxy plugin.
- Already good experience and well-tested.

- **CONS**

- Outside of containers world.



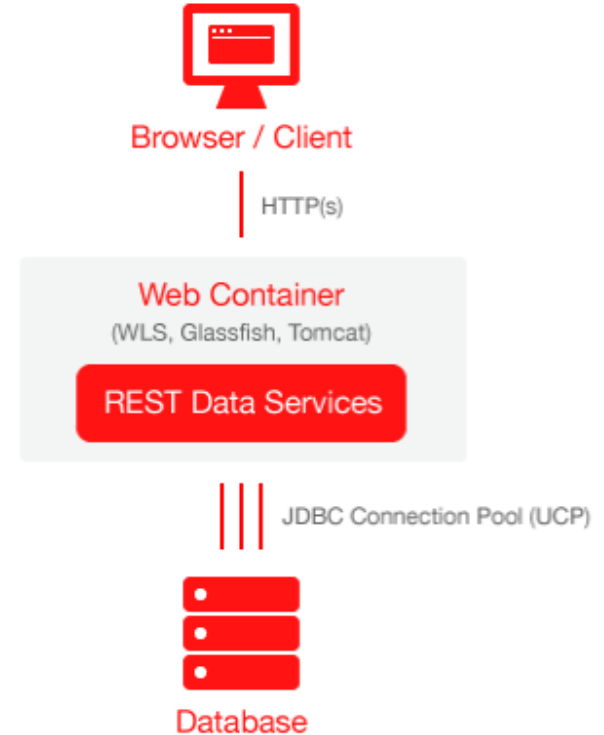


# CERN WebLogic 12.2.1 images

- Different images stored in CERN gitlab registry.
- Integration of WebLogic 12.2.1 with CERN tools to monitor and manage it.
- Different templates to define :
  - SSL.
  - Session Replication.
  - Identity Keystore and Trust Keystore.
  - Dynamic Cluster.
  - Machines, Servers.

# Oracle REST Data Service

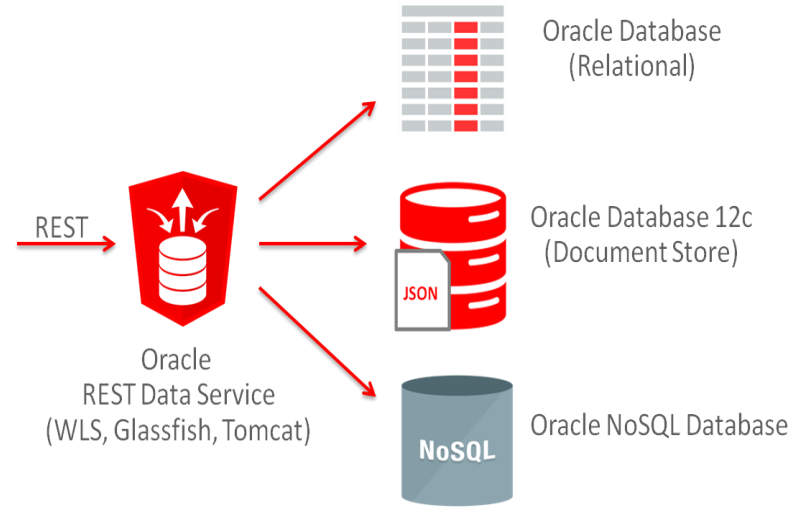
- **Polyglot clients.**
- **Standard way to talk with ORDS (JSON).**
  - The client do not talk directly with JDBC driver.
- **Stateless.**
  - Easy to recover.
  - Simple.



Credit: <https://blogs.oracle.com/24>

# Oracle REST Data Service

- **Running in production.**
- **Security enforced by WebLogic.**
- **Evaluating auth and authz via OAuth2.**
- **Phase out OHS.**

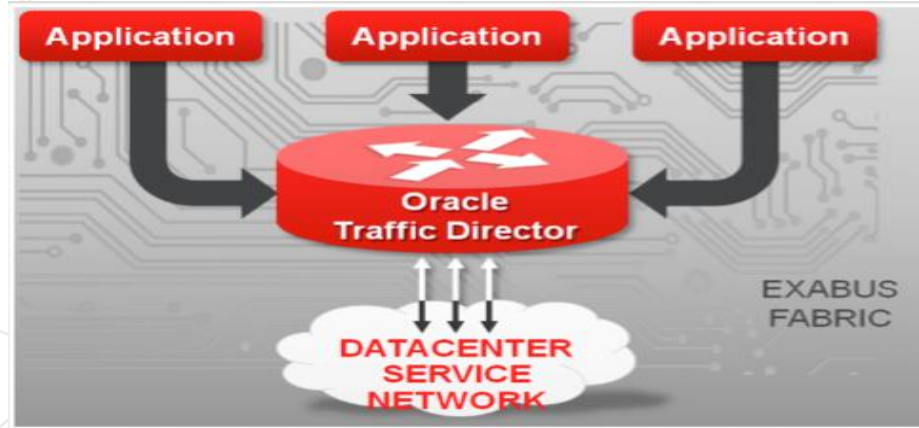


Credit: [www.oracle.com](http://www.oracle.com)



# Oracle Traffic Director

- **Oracle Traffic Director + WebLogic partitions + Dynamic cluster**
  - High Availability
  - Flexibility
    - Need just the definition of one Managed Node.
    - Scale up and down in a transparent way for the users.
  - Easy to install



# Trial Testing Oracle Java Cloud : Part 1

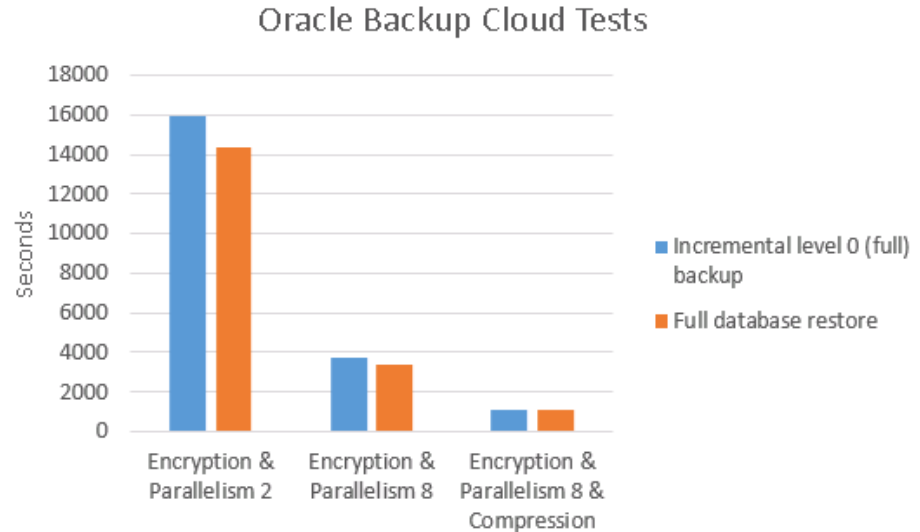
- **Interesting features:**
  - Same consoles.
  - Full services management.
  - Same APIs.
  - Fast response.
  - Full REST API.
  - Automatic patching.
  - Full development platform.
  - Network Access Rules.
- **To Do :**  
Planning real tests.

# Testing Oracle Java Cloud : Part 2

- **Single Sign On**
  - Needed a tunnel between VMs running on Oracle Cloud and server registered under cern.ch domain.
- **Service provisioning automatically**
  - REST API allow you to interact via command line.
- **Evaluate the integration between CERN tools and the Oracle Cloud.**

# Oracle Backup Cloud (Functional Tests)

- **Data center in USA**
  - Legal Issue.
  - Latency.
- **Initial problems with installation behind proxy.**
- **Data center in EU :**
  - Performance Tests.
  - Functional Tests.



# Conclusions and Plans

- **Oracle Cloud - interesting first functionality tests**
- **In-Memory - performance**
- **Data Analytics**
  - Facilitate data exploitation (knowledge discovery, ETL, analytics)
  - Streaming analytics and processing
- **Test Oracle Backup Cloud in Europe.**
- **Test Java Cloud as solution for our dev and test services.**
- **WebLogic :**
  - LDAP and SSO
  - Test Load Balancing and HAProxy with high load.
  - Integrate with IT Departments Services.
    - OpenStack Magnum
- **Test Oracle Container Service Cloud.**

# Public Outreach

- **5 Presentations** at Oracle OpenWorld 2016
- 1 Presentation at AppDev Deploy Summit in Paris 2016.
- Several round table discussion, customer panel and conference session
- 4 reference meetings with industry leading companies



- 8 external presentations (~ last 6 months)
- 2 publications on Spanish newspapers (El Mundo, Expansion)

Onsite workshops with multiple industry leading companies

More details: <http://openlab.web.cern.ch/publications>

# Questions

