Oracle Big Data SQL
Hands-on Lab

Marty Gubar - Big Data PM

Big Made Simple
Database Strategy for Big Data

Conventional view of Data Management

- SQL
- DATA

Emerging view of Data Management

- Office
- SQL
- ORACLE
- NoSQL
Big Data SQL Configurations

Engineered Systems

cloudera

Commodity Servers

cloudera

Mixed Deployment

cloudera

Mixed Deployment

cloudera

Oracle Cloud*

* Coming Soon!
CREATE TABLE movielog (click VARCHAR2(4000))
ORGANIZATION EXTERNAL (TYPE ORACLE_HIVE
DEFAULT DIRECTORY DEFAULT_DIR
ACCESS PARAMETERS
(
com.oracle.bigdata.tablename logs
com.oracle.bigdata.cluster mycluster
))
REJECT LIMIT UNLIMITED;

• New types of external tables
  – ORACLE_HIVE (leverage hive metadata)
  – ORACLE_HDFS (specify metadata)

• Access parameters used to describe how to identify sources and process data on the hadoop cluster
Oracle Big Data SQL

Massively Parallel SQL Query across Oracle, Hadoop and NoSQL

Offload Query to Data Nodes

Hadoop & NoSQL

Oracle Database 12c

Small data subset quickly returned

Offload Query to Exadata Storage Servers
Big Data SQL: Another Hadoop Processing Engine

Storage Layer
- Filesystem (HDFS)
- NoSQL Databases (Oracle NoSQL DB, HBase)

Resource Management (YARN, cgroups)

Processing Layer
- MapReduce and Hive
- Spark
- Impala
- Search (Big Data SQL)

Meta data Store
Big Data SQL Performance Features

IO Reduction Features Deliver Compound Results

User Query: 100 TB
Partition Pruning: 10 TB
Storage Indexing: 1 TB
Predicate Pushdown: 100 GB
Scenario: On-line Movie Streaming Site

• Data sources
  – Recommendations in NoSQL DB
  – User behavior in HDFS
  – Data Warehouse in Oracle Database

• Gain value from all data!
List of workshops

• Part 1 - Configuring Oracle Big Data SQL
• Part 2 - Create Oracle Table Over Application Log
• Part 3 - Leverage the Hive Metastore to Access Data in Hadoop & Oracle NoSQL Database
• Part 4 - Applying Oracle Database Security Policies Across the Big Data Platform
List of workshops

• Part 5 - Using Oracle Analytic SQL Across All Your Data
• Part 6 - Introduction to SQL Pattern Matching
• Part 7 - Checking the Pattern Matching Process
• Part 8 - Creating a More Useful Data Set
• Part 9 - Other Useful 12c Analytical SQL Features
Ravello accelerates the move from data-center to public cloud

Run data-center based **VMware** or **KVM** workloads on public cloud

*without any changes*

Same VMs, networking, storage

No app changes → Data-center to cloud *in hours*, and *NOT months*
Connecting

• You will be assigned an IP Address
• Use VNC:
  – vnc <ip-address>:1
  – password: welcome1
• or SQL Developer
  – ip-address:1521:orcl
  – moviedemo / welcome1
Starting the Hands-on Lab

SQL Developer

Command Prompt

Oracle Big Data Lite Virtual Machine

Big Data SQL HQL

Hands-on Lab Document
For More Information

• Get started with the software using Oracle Big Data Lite: http://www.oracle.com/technetwork/database/bigdata-appliance/oracle-bigdatalite-2104726.html
  – (simply google “Big Data Lite”)
  – Link to HOL is on that page

• Oracle.com:

• Data Warehouse Insider Blog: https://blogs.oracle.com/datawarehousing/
Integrated Cloud
Applications & Platform Services