Lesson 4 : Hadoop Reporting and Visualization using OBIEE 11g

Mark Rittman, CTO, Rittman Mead

Oracle Openworld 2014, San Francisco

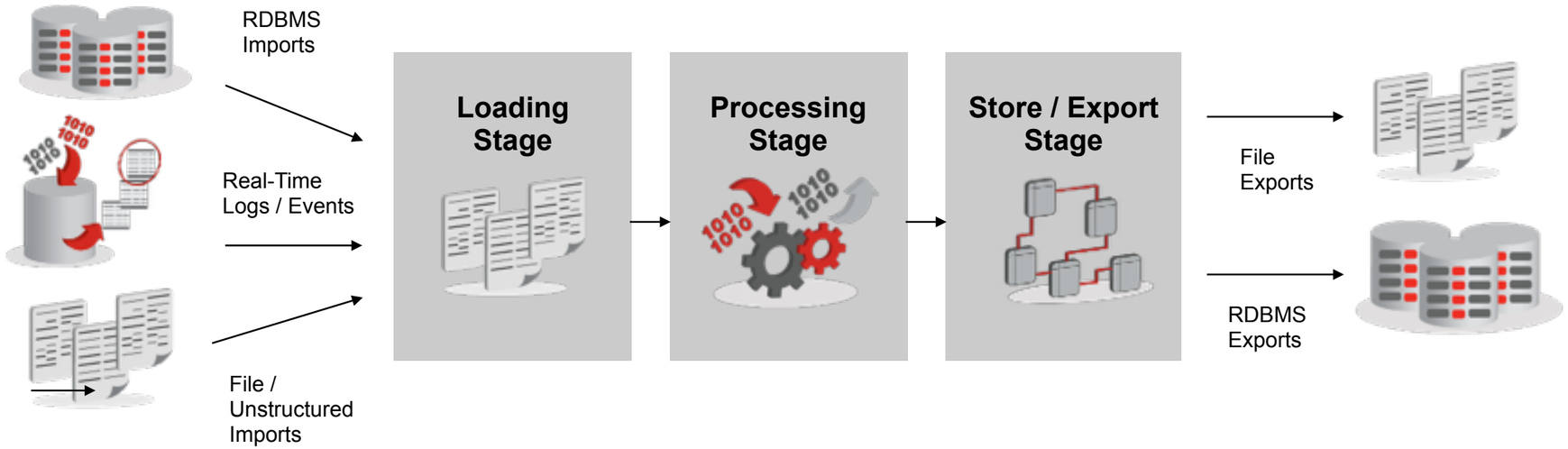
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Moving Data In, Around and Out of Hadoop

- Three stages to Hadoop data movement, with dedicated Apache / other tools
  - **Load** : receive files in batch, or in real-time (logs, events)
  - **Transform** : process & transform data to answer questions
  - **Store / Export** : store in structured form, or export to RDBMS using Sqoop



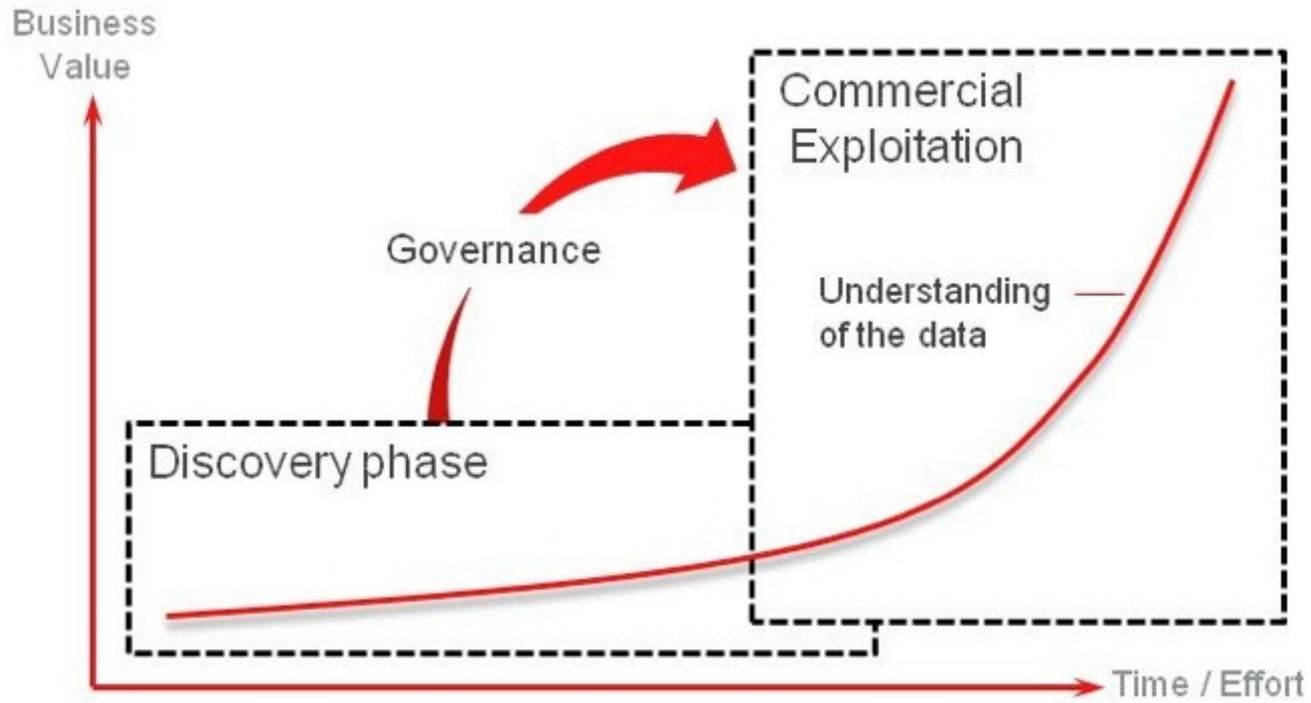
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Discovery vs. Exploitation Project Phases

- Discovery and monetising steps in Big Data projects have different requirements
- Discovery phase
  - Unbounded discovery
  - Self-Service sandbox
  - Wide toolset
- Promotion to Exploitation
  - Commercial exploitation
  - Narrower toolset
  - Integration to operations
  - Non-functional requirements
  - Code standardisation & gover

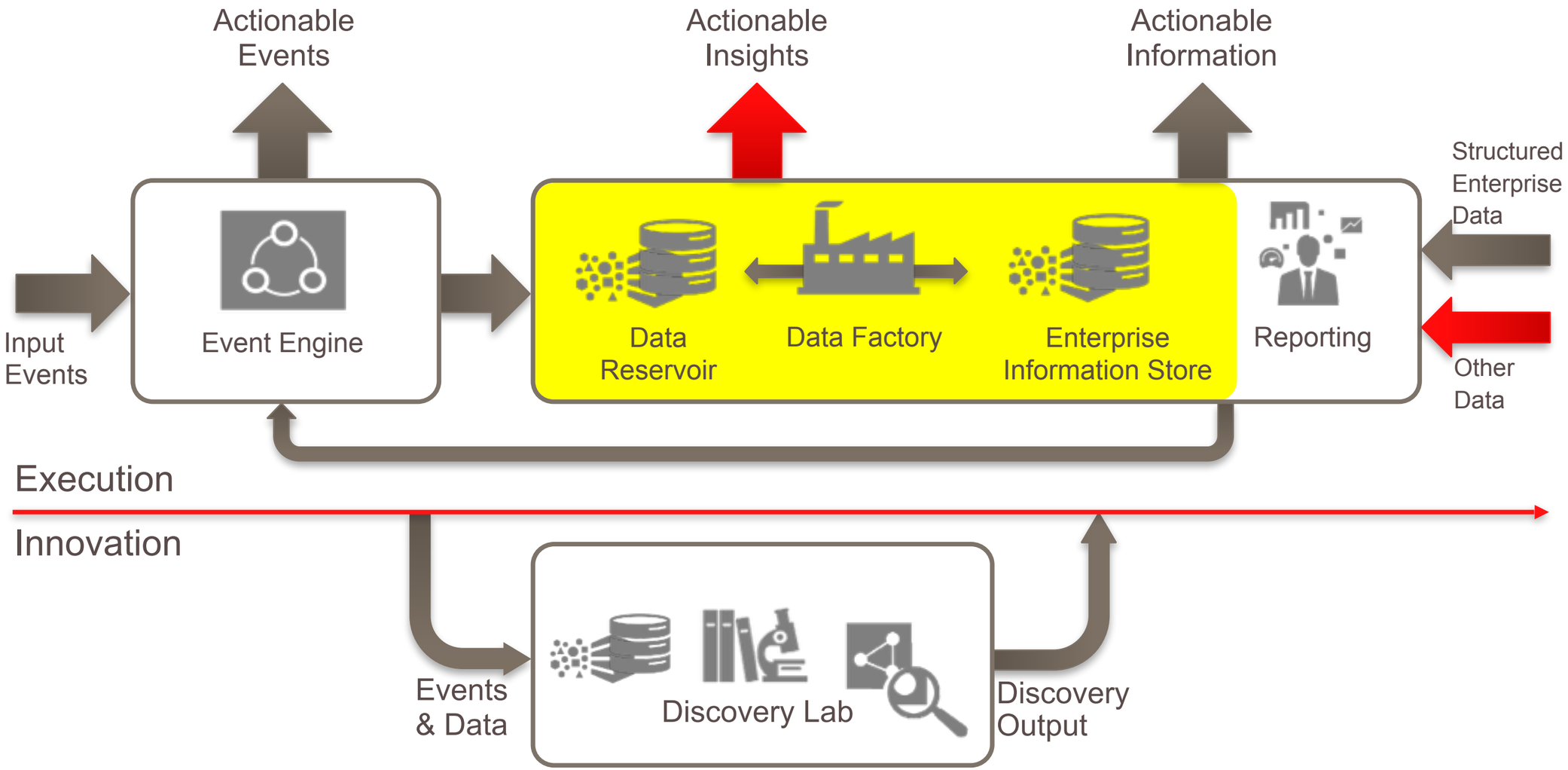


T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

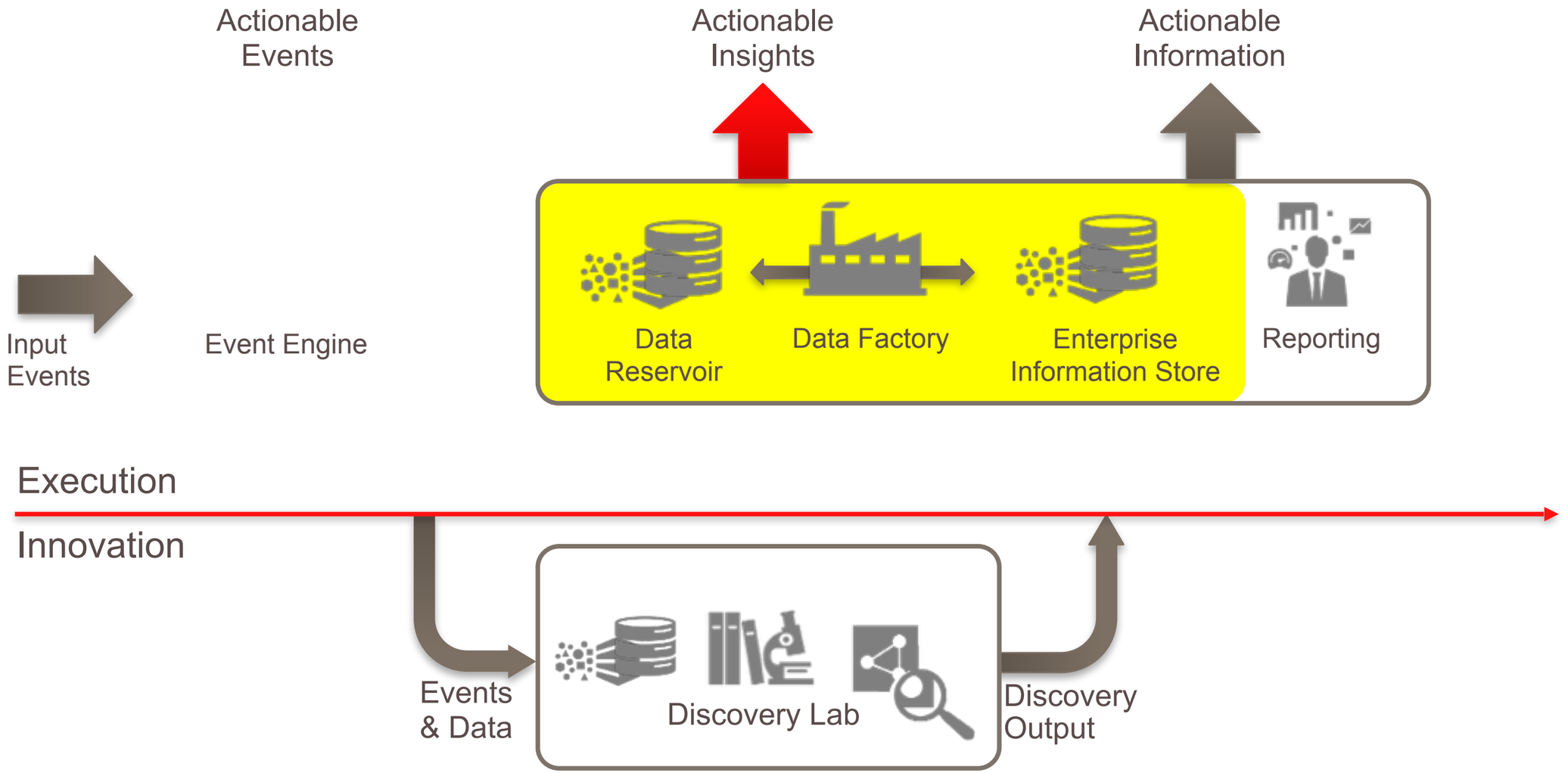
E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Information Solution

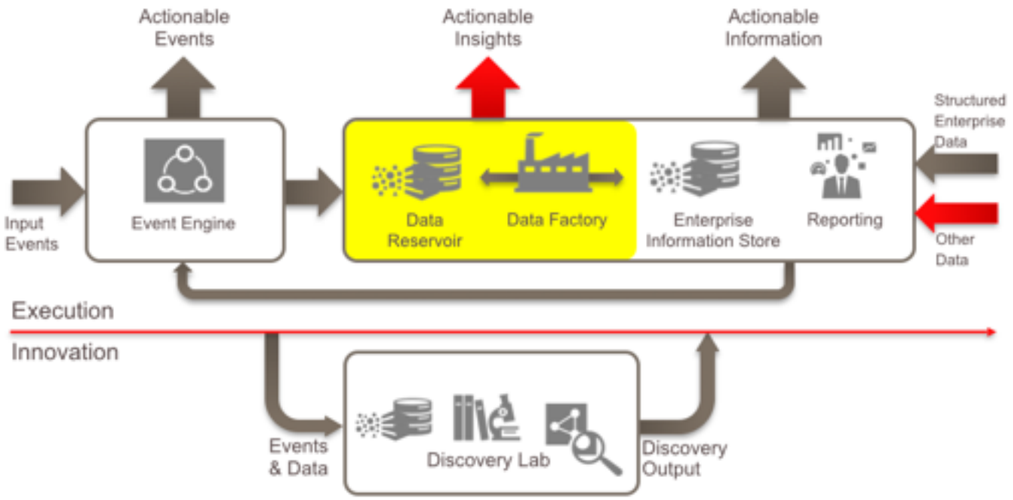


Information Solution



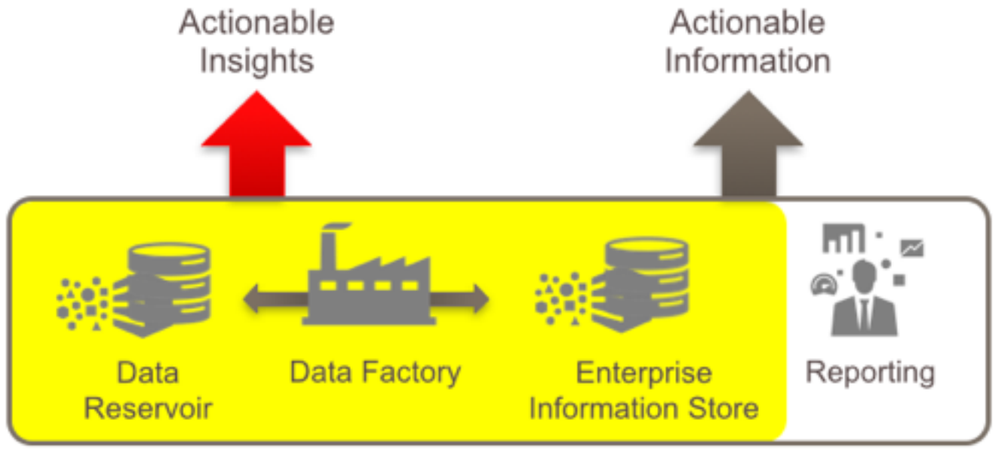
Design Pattern : Information Solution

- Specific solution based on Big Data technologies requiring broader integration to the wider Information Management estate
  e.g. ETL pre-processor for the DW or affordably store a lower level of grain
  - Non-functional requirements more critical in this solution
  - Scalable integration to IM estate
    an important factor for success
  - Analysis may take place in Reservoir
    or Reservoir only used as an aggregator



Options for Sharing Hadoop Output with Wider Audience

- During the discovery phase of a Hadoop project, audience are likely technical
  - Most comfortable with data analyst tools, command-line, low-level access to the data
- During the exploitation phase, audience will be less technical
  - Emphasis on graphical tools, and integration with wider reporting toolset + metadata
- Three main options for visualising and sharing Hadoop data
  1. Coming Soon - Oracle Big Data Discovery (Endeca on Hadoop)
  2. OBIEE reporting against Hadoop direct using Hive/Impala, or Oracle Big Data SQL
  3. OBIEE reporting against an export of the Hadoop data, on Exalytics / RDBMS

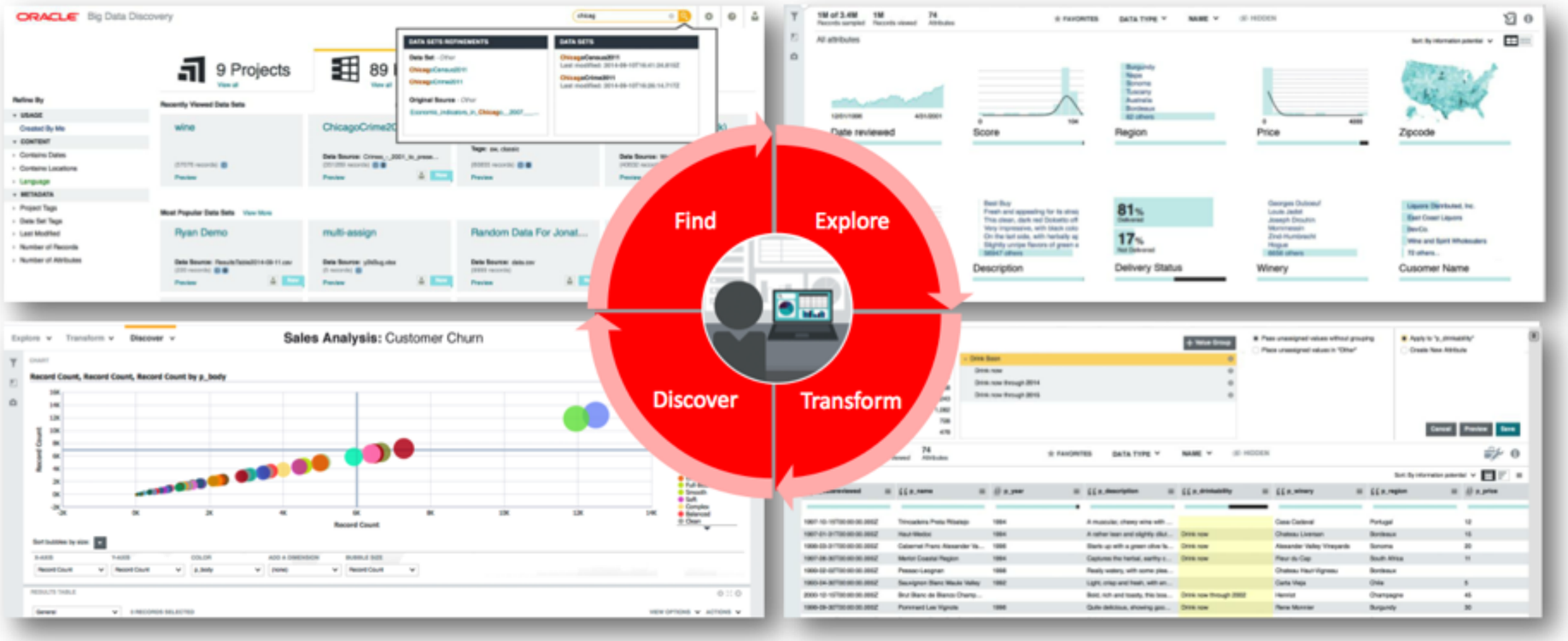


T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

Oracle Big Data Discovery

- Launched at Oracle Openworld 2014 as “The Visual Face of Hadoop”
- Combined Endeca Server search + analytical technology with Spark data transformation



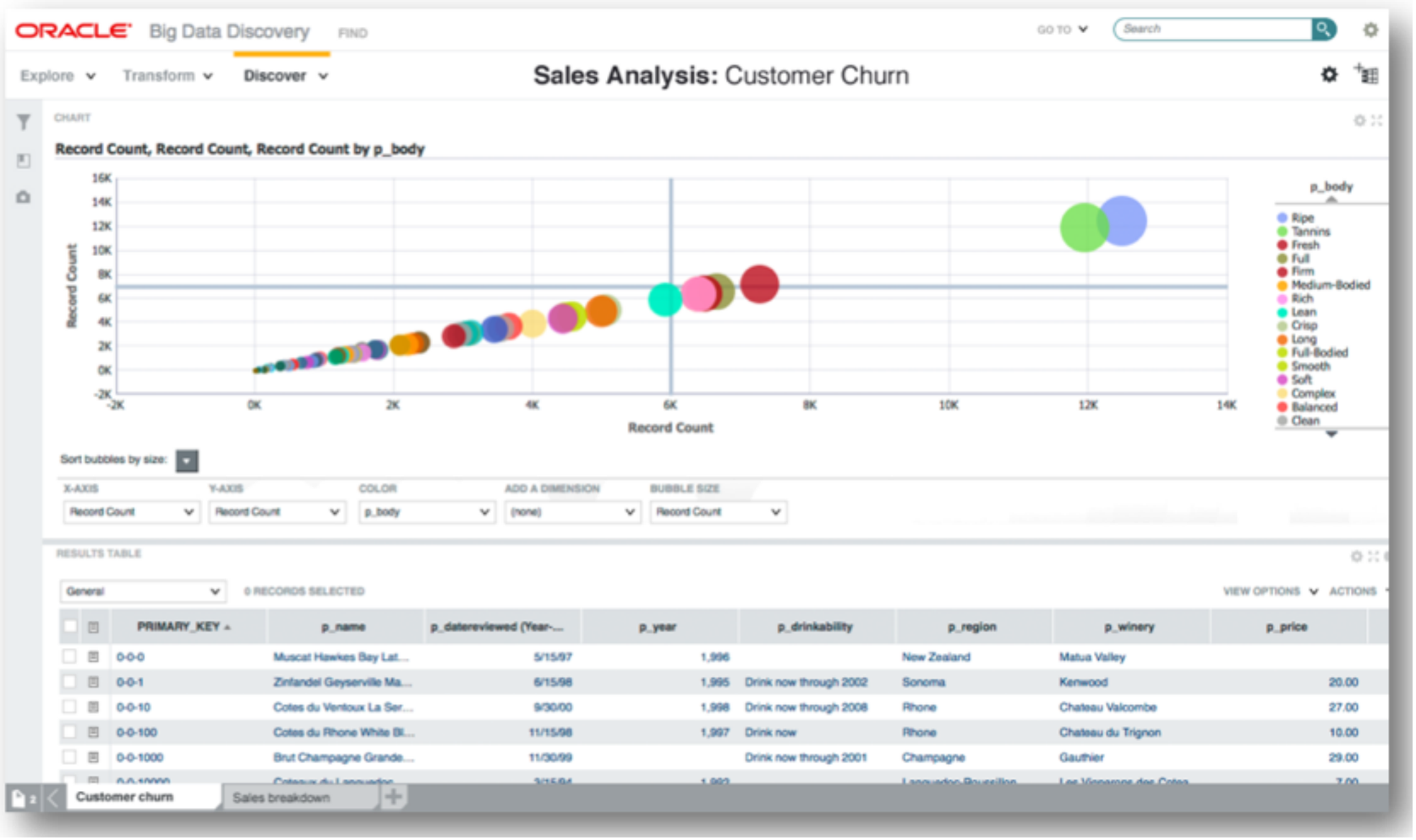
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Interactive Analysis & Exploration of Hadoop Data

- **Mash up** different data sets for deeper perspectives
- **Drag and drop** from a rich library of **interactive visualizations** to compose discovery dashboards
- Filter through data with powerful **search** and intuitive **guided navigation**
- **Publish** blended data sets back to Hadoop
- **Share** projects, bookmarks and snapshots with team members for collaboration

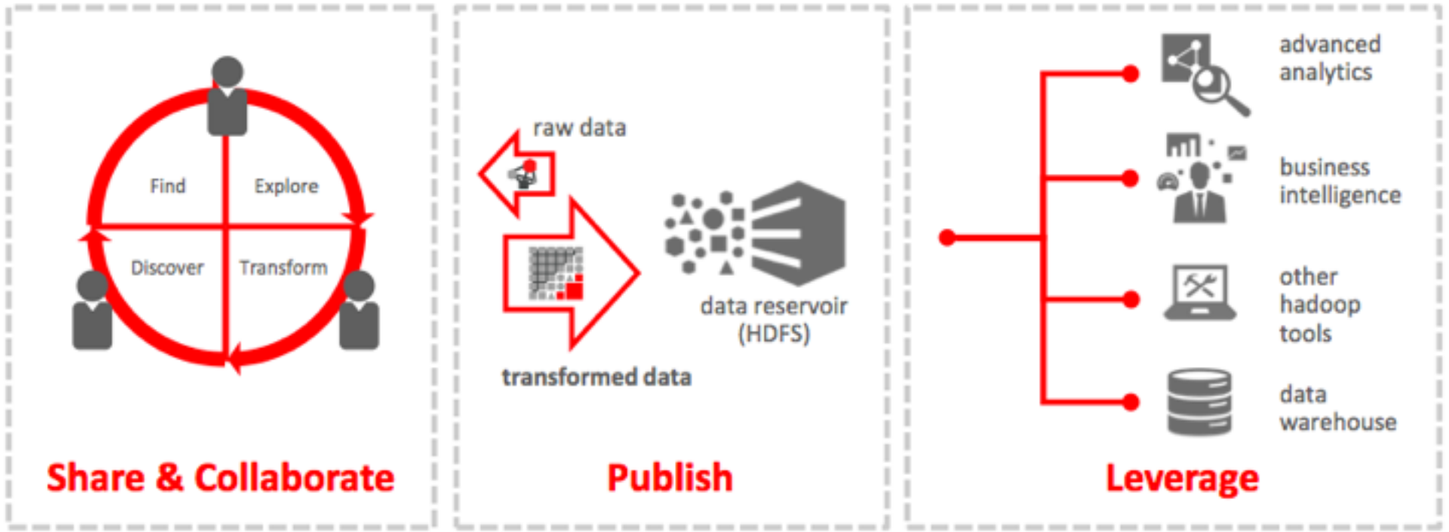


T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Share and Collaborate on Big Data Discovery Projects



- **Share and collaborate** with the team
  - Share projects, bookmarks and snapshots then collaborate and iterate
- **Publish** back to Hadoop
  - Transforms and enrichments may be applied to original data sets in Hadoop
  - Publish blended data sets back to HDFS
- **Leverage** results in other tools
  - Publish data to Hadoop in format optimized for advanced analytic tools (e.g. ORAAH)
  - Hadoop compliant BI tools (e.g. OBIFS) can burst out to the masses
  - Leverage any native Hadoop tooling (e.g. Pig, Hive, Impala, Python, etc)
  - Integrate BDD data sets with DWH to secure, govern and optimize for query performance (e.g. Oracle Big Data SQL)

Oracle Big Data Discovery plays well with the **big data ecosystem**

T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Oracle Business Analytics and Big Data Sources

- OBIEE 11g can also make use of big data sources
  - OBIEE 11.1.1.7+ supports Hive/Hadoop as a data source
  - Oracle R Enterprise can expose R models through DB functions, columns
  - Oracle Exalytics has InfiniBand connectivity to Oracle BDA
- Endeca Information Discovery can analyze unstructured and semi-structured sources
  - Increasingly tighter-integration between OBIEE and Endeca

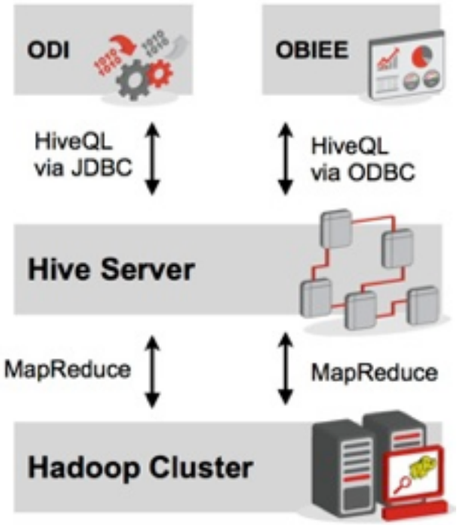
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

New in OBIEE 11.1.1.7 : Hadoop Connectivity through Hive

- MapReduce jobs are typically written in Java, but Hive can make this simpler
- Hive is a query environment over Hadoop/MapReduce to support SQL-like queries
- Hive server accepts HiveQL queries via HiveODBC or HiveJDBC, automatically creates MapReduce jobs against data previously loaded into the Hive HDFS tables
- Approach used by ODI and OBIEE to gain access to Hadoop data
- Allows Hadoop data to be accessed just like any other data source



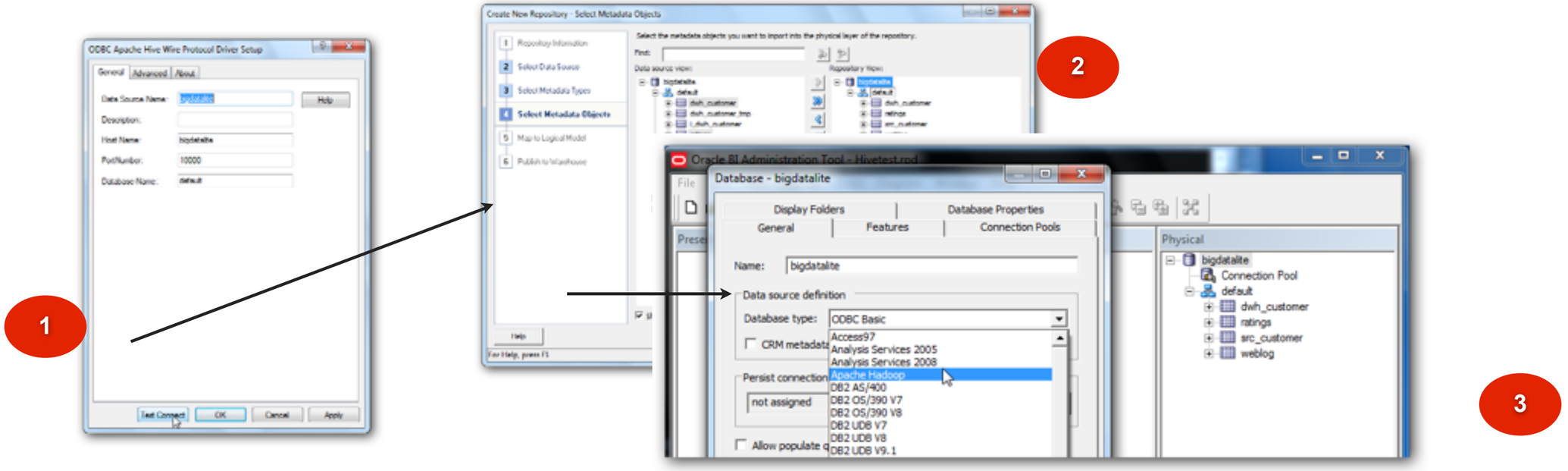
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Importing Hadoop/Hive Metadata into RPD

- HiveODBC driver has to be installed into Windows environment, so that BI Administration tool can connect to Hive and return table metadata
- Import as ODBC datasource, change physical DB type to Apache Hadoop afterwards
- Note that OBIEE queries cannot span >1 Hive schema (no table prefixes)



T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

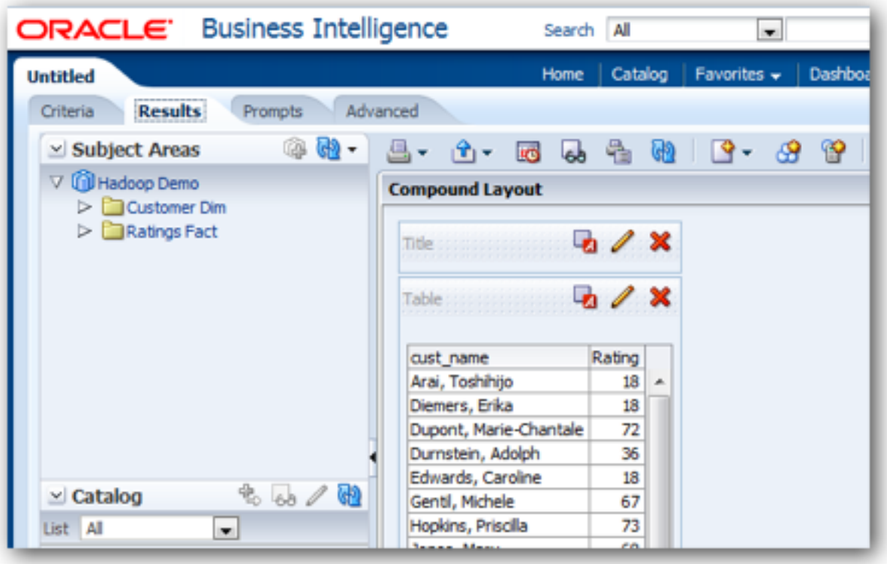
rittmanmead
INTEGRATED ANALYTICS

Set up ODBC Connection at the OBIEE Server

- OBIEE 11.1.1.7+ ships with HiveODBC drivers, need to use 7.x versions though (only Linux supported)
- Configure the ODBC connection in odbc.ini, name needs to match RPD ODBC name
- BI Server should then be able to connect to the Hive server, and Hadoop/MapReduce

```
[ODBC Data Sources]
AnalyticsWeb=Oracle BI Server
Cluster=Oracle BI Server
SSL_Sample=Oracle BI Server
bigdatalite=Oracle 7.1 Apache Hive Wire Protocol

[bigdatalite]
Driver=/u01/app/Middleware/Oracle_BI1/common/ODBC/
 Merant/7.0.1/lib/ARhive27.so
Description=Oracle 7.1 Apache Hive Wire Protocol
ArraySize=16384
Database=default
DefaultLongDataBuffLen=1024
EnableLongDataBuffLen=1024
EnableDescribeParam=0
Hostname=bigdatalite
LoginTimeout=30
MaxVarcharSize=2000
PortNumber=10000
RemoveColumnQualifiers=0
StringDescribeType=12
TransactionMode=0
UseCurrentSchema=0
```



T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Dealing with Hadoop / Hive Latency Option 1 : Impala

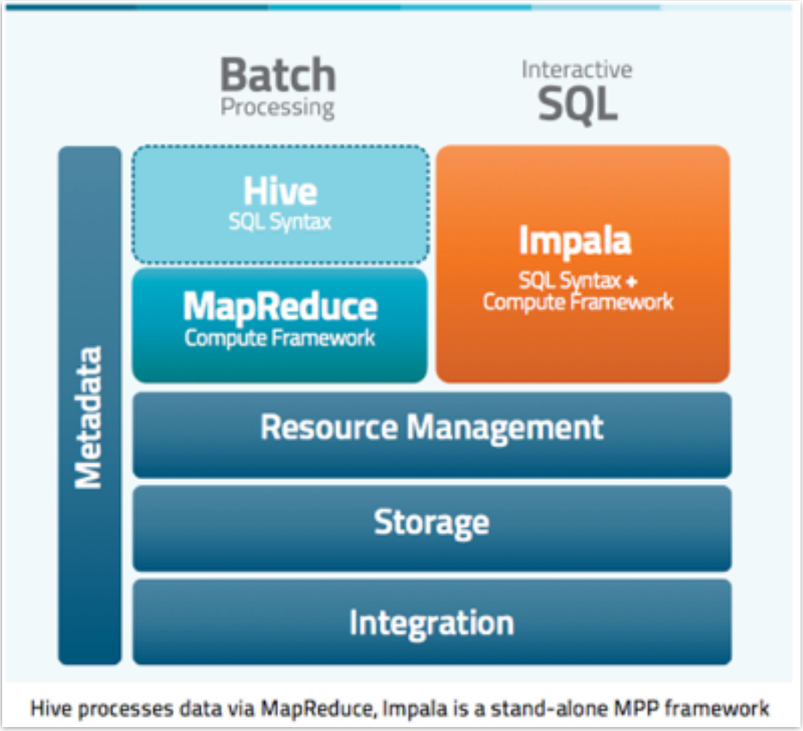
- Hadoop access through Hive can be slow - due to inherent latency in Hive
- Hive queries use MapReduce in the background to query Hadoop
- Spins-up Java VM on each query
- Generates MapReduce job
- Runs and collates the answer
- Great for large, distributed queries ...
- ... but not so good for “speed-of-thought” dashboards

T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

Dealing with Hadoop / Hive Latency Option 1 : Use Impala

- Hive is slow - because it's meant to be used for batch-mode queries
- Many companies / projects are trying to improve Hive - one of which is Cloudera
- Cloudera Impala is an open-source but commercially-sponsored in-memory MPP platform
- Replaces Hive and MapReduce in the Hadoop stack
- Can we use this, instead of Hive, to access Hadoop?
  - It will need to work with OBIEE
  - Warning - it won't be a supported data source (yet…)



Hive processes data via MapReduce, Impala is a stand-alone MPP framework

T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

Demo

Using Hive to Provide Data for OBIEE11g

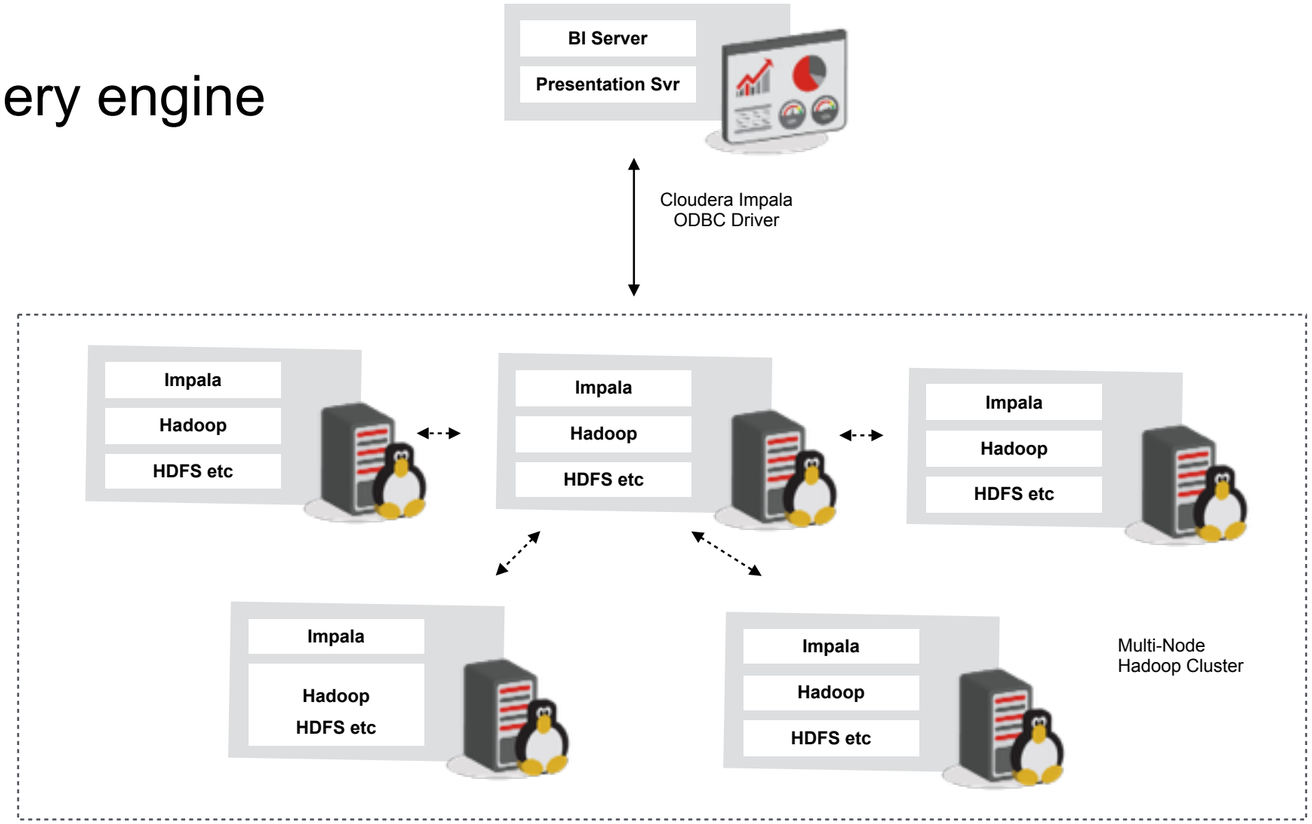
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

How Impala Works

- A replacement for Hive, but uses Hive concepts and data dictionary (metastore)
- MPP (Massively Parallel Processing) query engine that runs within Hadoop
  - Uses same file formats, security, resource management as Hadoop
- Processes queries in-memory
- Accesses standard HDFS file data
- Option to use Apache AVRO, RCFile, LZO or Parquet (column-store)
- Designed for interactive, real-time SQL-like access to Hadoop



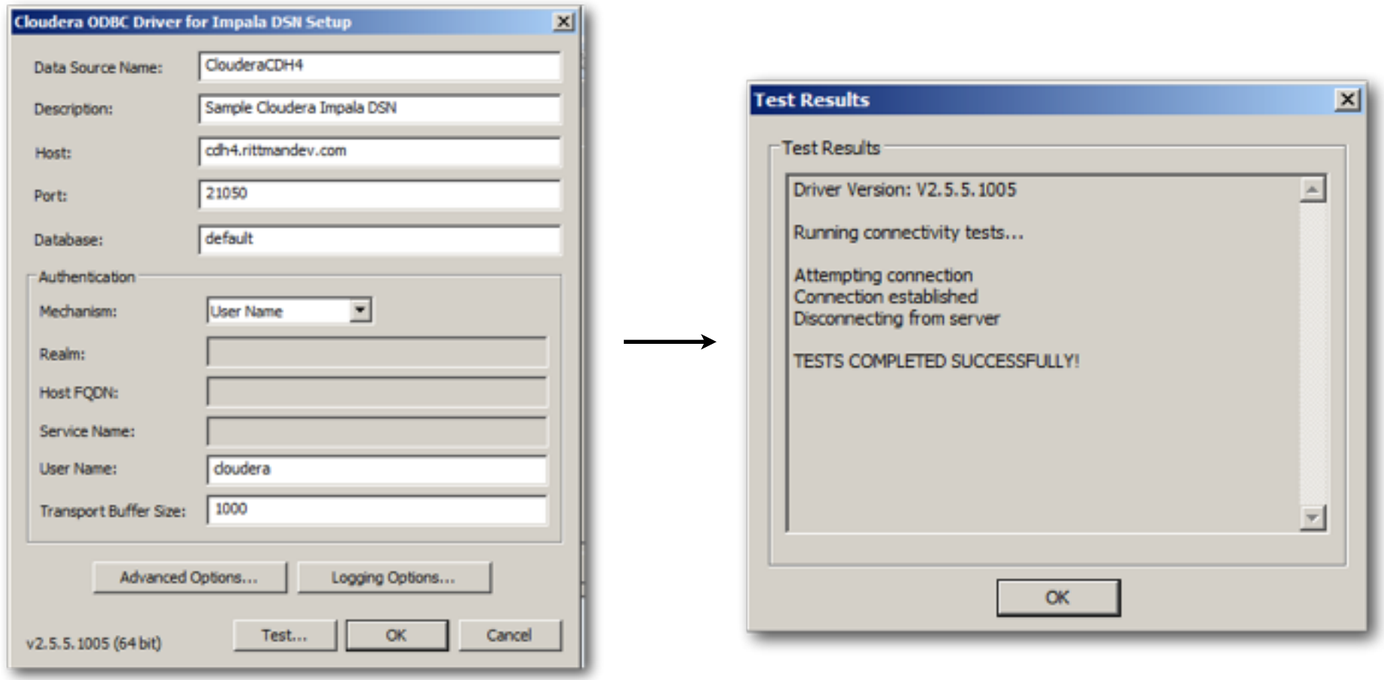
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Connecting OBIEE 11.1.1.7 to Cloudera Impala

- Warning - unsupported source - limited testing and no support from MOS
- Requires Cloudera Impala ODBC drivers - Windows or Linux (RHEL etc/SLES) - 32/64 bit
- ODBC Driver / DSN connection steps similar to Hive



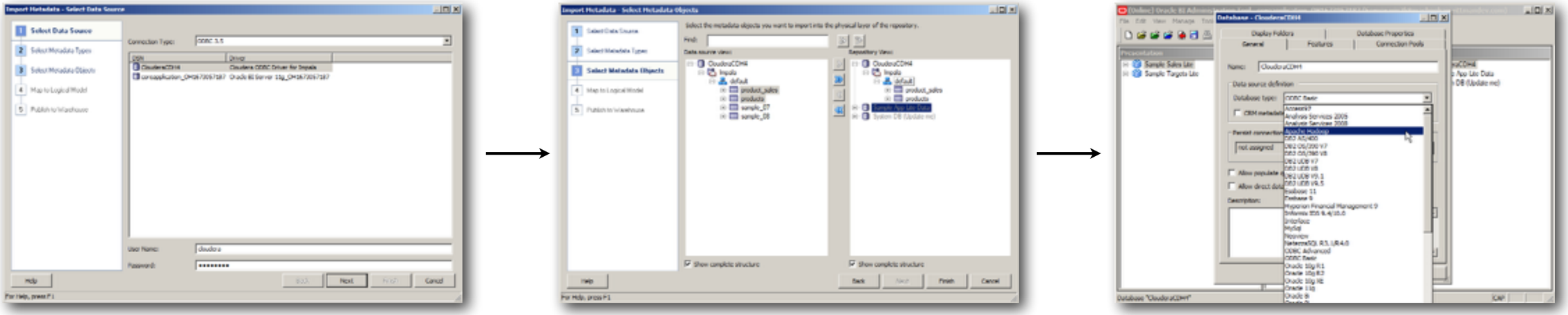
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Importing Impala Metadata

- Import Impala tables (via the Hive metastore) into RPD
- Set database type to “Apache Hadoop”
  - Warning - don’t set ODBC type to Hadoop- leave at ODBC 2.0
  - Create physical layer keys, joins etc as normal



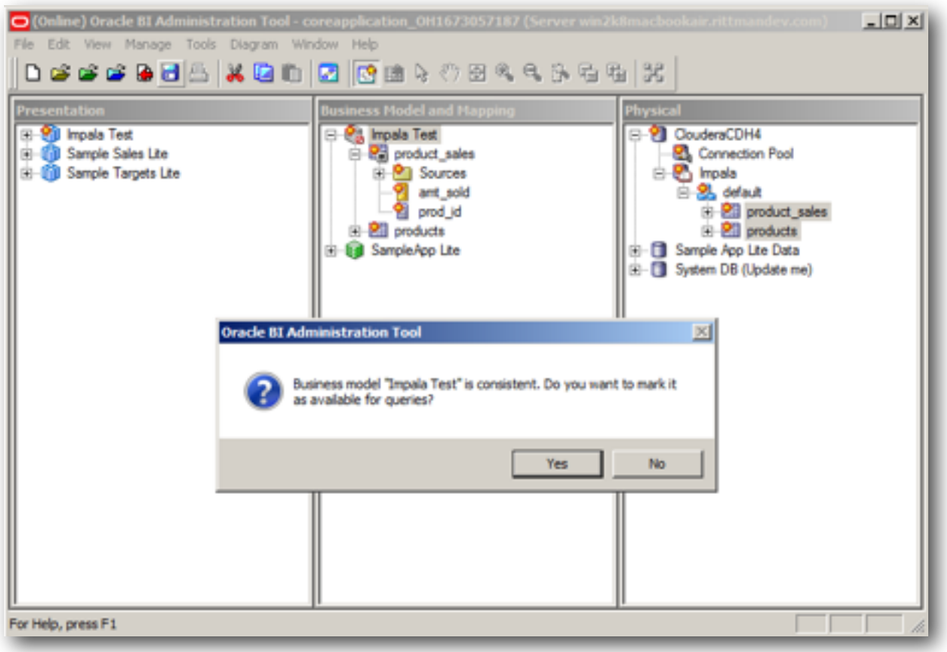
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

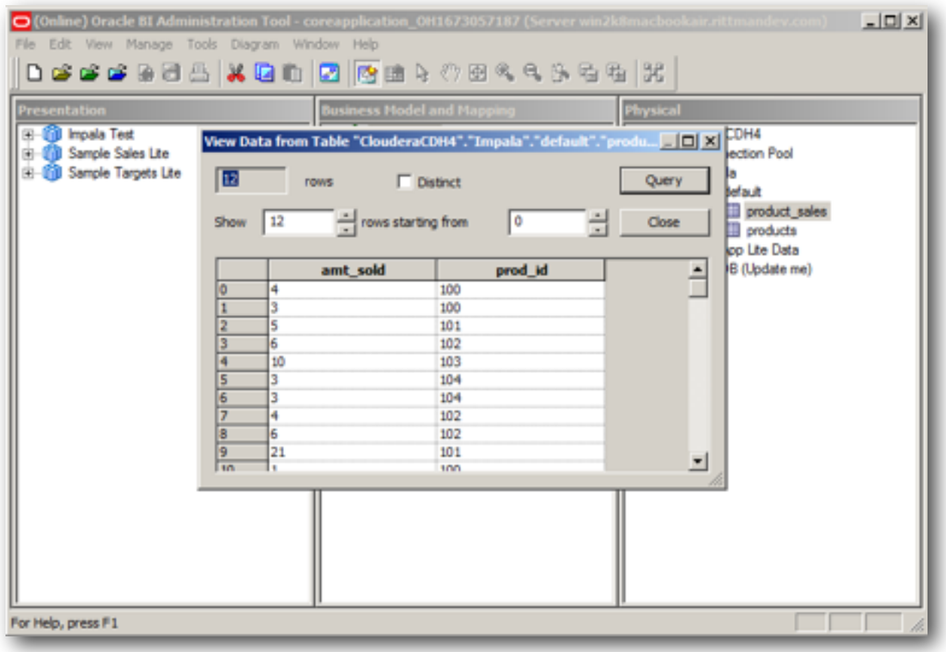
E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Importing RPD using Impala Metadata

- Create BMM layer, Presentation layer as normal
- Use “View Rows” feature to check connectivity back to Impala / Hadoop





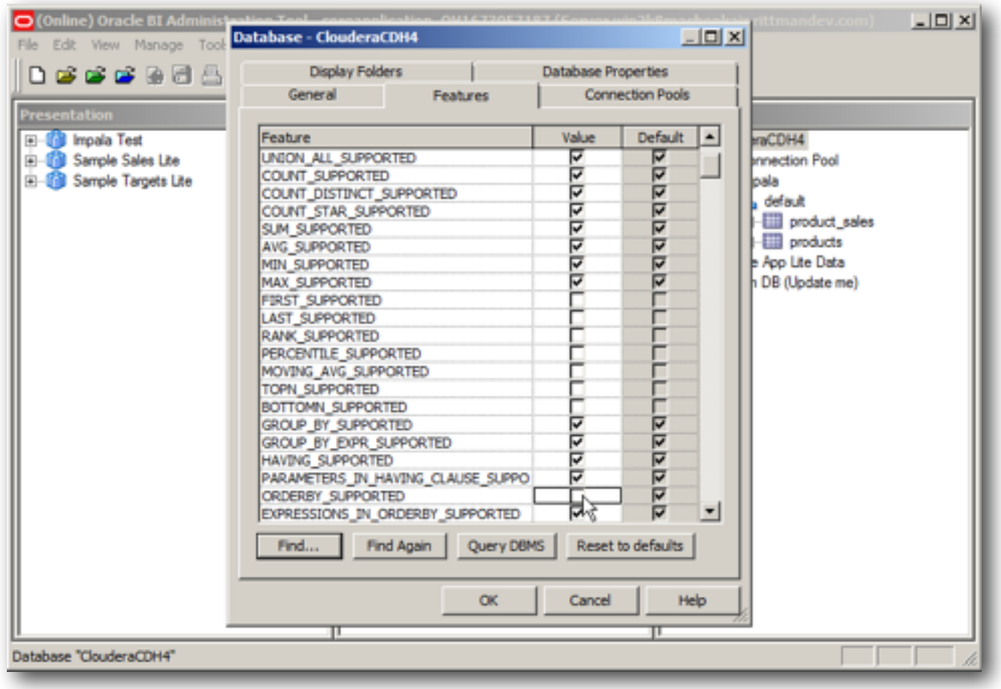
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Impala / OBIEE Issue with ORDER BY Clause

- Although checking rows in the BI Administration tool worked, any query that aggregates data in the dashboard will fail
- Issue is that Impala requires LIMIT with all ORDER BY clauses
  - OBIEE could use LIMIT, but doesn't for Impala at the moment (because not supported)
- Workaround - disable ORDER BY in Database Features, have the BI Server do sorting
  - Not ideal - but it works, until Impala supported



T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead INTEGRATED ANALYTICS

Demo

Using Impala to Provide Data for OBIEE11g

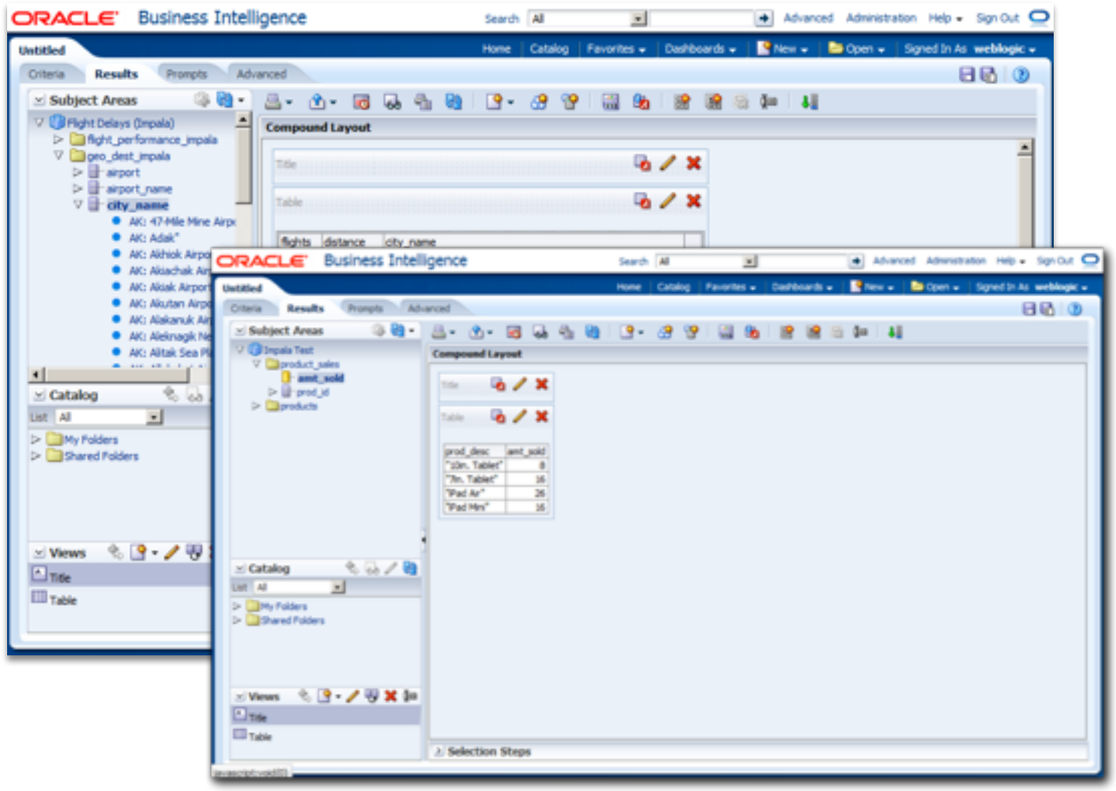
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

So Does Impala Work, as a Hive Substitute?

- With ORDER BY disabled in DB features, it appears to
- But not extensively tested by me, or Oracle
- But it's certainly interesting
- Reduces 30s, 180s queries down to 1s, 10s etc
- Impala, or one of the competitor projects (Drill, Dremel etc) assumed to be the real-time query replacement for Hive, in time
  - Oracle announced planned support for Impala at OOW2013 - watch this space



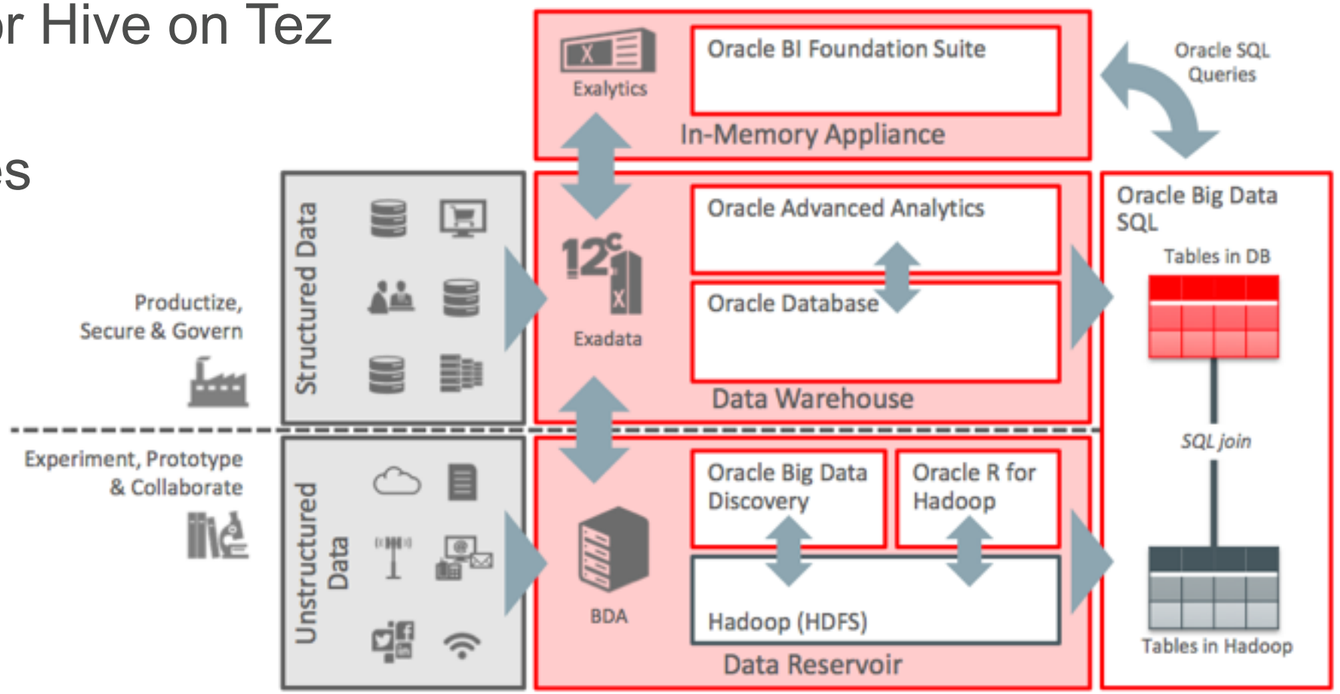
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Dealing with Hadoop / Hive Latency Option 2 : Big Data SQL

- Preferred solution for customers with Oracle Big Data Appliance is Big Data SQL
- Oracle SQL Access to both relational, and Hive/NoSQL data sources
- Exadata-type SmartScan against Hadoop datasets
  - Response-time equivalent to Impala or Hive on Tez
  - No issues around HiveQL limitations
  - Insulates end-users around differences between Oracle and Hive datasets



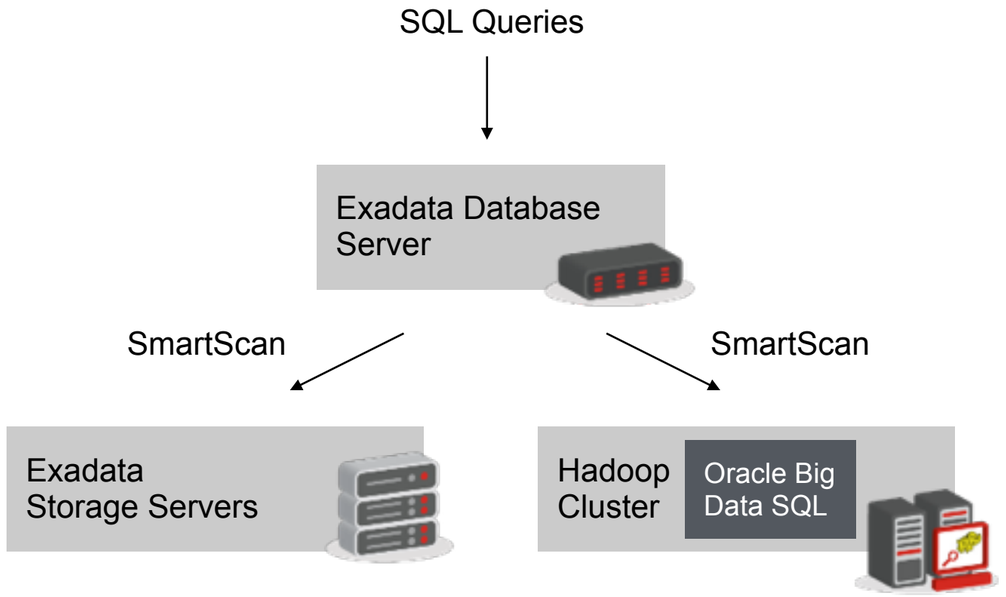
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Oracle Big Data SQL

- Part of Oracle Big Data 4.0 (BDA-only)
  - Also requires Oracle Database 12c, Oracle Exadata Database Machine
- Extends Oracle Data Dictionary to cover Hive
- Extends Oracle SQL and SmartScan to Hadoop
- Extends Oracle Security Model over Hadoop
  - Fine-grained access control
  - Data redaction, data masking
- Uses fast c-based readers where possible (vs. Hive MapReduce generation)
- Map Hadoop parallelism to Oracle PQ
- Big Data SQL engine works on top of YARN
  - Like Spark, Tez, MR2

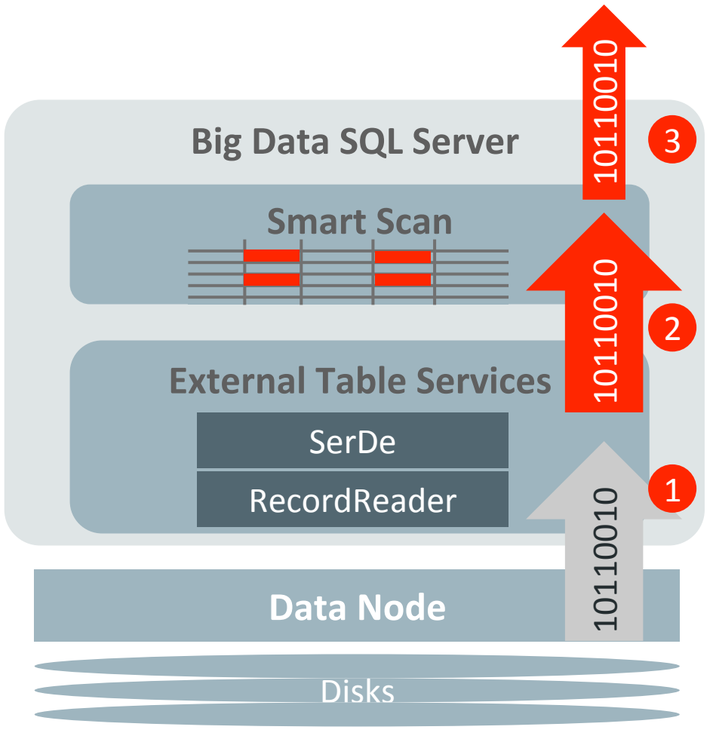


T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

Big Data SQL Server Dataflow

1. Read data from HDFS Data Node
   - Direct-path reads
   - C-based readers when possible
   - Use native Hadoop classes otherwise
2. Translate bytes to Oracle
3. Apply SmartScan to Oracle bytes
   - Apply filters
   - Project columns
   - Parse JSON/XML
   - Score models



T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

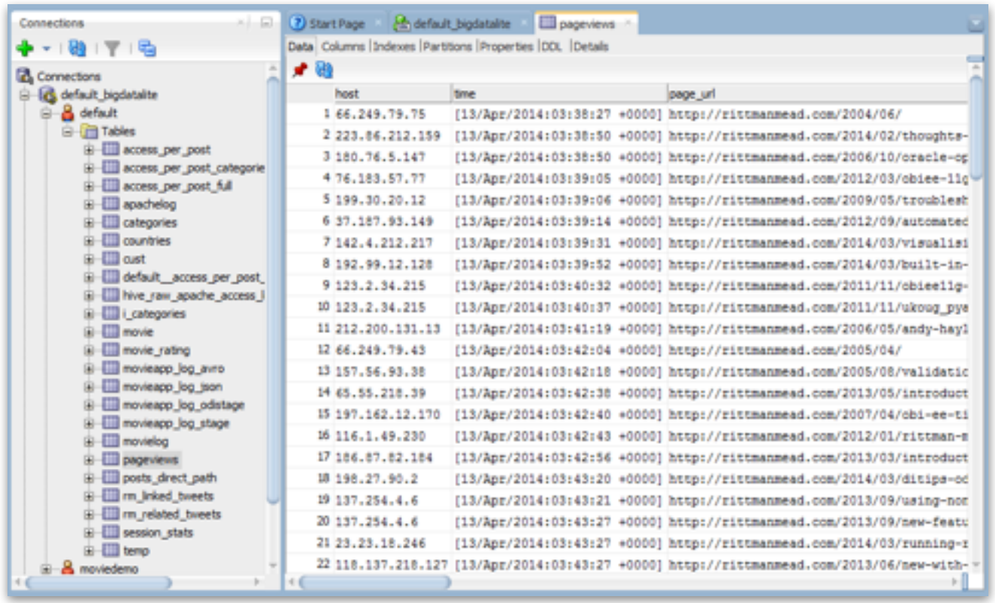
E : info@rittmanmead.com
W : www.rittmanmead.com

View Hive Table Metadata in the Oracle Data Dictionary

- Oracle Database 12c 12.1.0.2.0 with Big Data SQL option can view Hive table metadata
  - Linked by Exadata configuration steps to one or more BDA clusters
- DBA_HIVE_TABLES and USER_HIVE_TABLES exposes Hive metadata
- Oracle SQL*Developer 4.0.3, with Cloudera Hive drivers, can connect to Hive metastore

```
SQL> col database_name for a30
SQL> col table_name for a30
SQL> select database_name, table_name
  2  from dba_hive_tables;

DATABASE_NAME                  TABLE_NAME
------------------------------ ------------------------------
default                        access_per_post
default                        access_per_post_categories
default                        access_per_post_full
default                        apachelog
default                        categories
default                        countries
default                        cust
default                        hive_raw_apache_access_log
```



T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

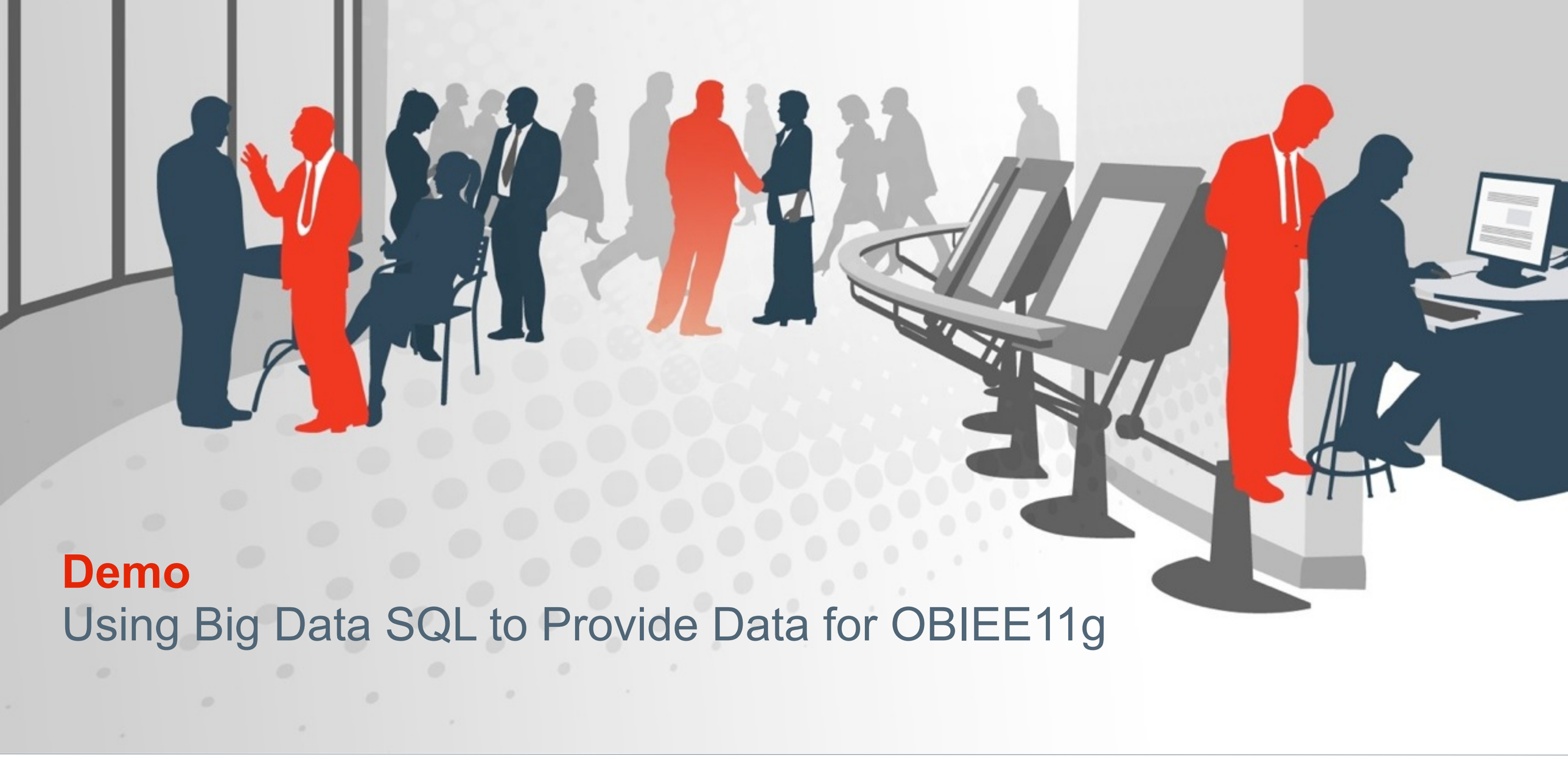
E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Hive Access through Oracle External Tables + Hive Driver

- Big Data SQL accesses Hive tables through external table mechanism
  - **ORACLE_HIVE** external table type imports Hive metastore metadata
  - **ORACLE_HDFS** requires metadata to be specified
- Access parameters **cluster** and **tablename** specify Hive table source and BDA cluster

```
CREATE TABLE access_per_post_categories(
  hostname varchar2(100),
  request_date varchar2(100),
  post_id varchar2(10),
  title varchar2(200),
  author varchar2(100),
  category varchar2(100),
  ip_integer number)
organization external
(type oracle_hive
 default directory default_dir
 access parameters(com.oracle.bigdata.tablename=default.access_per_post_categories));
```



Demo

Using Big Data SQL to Provide Data for OBIEE11g

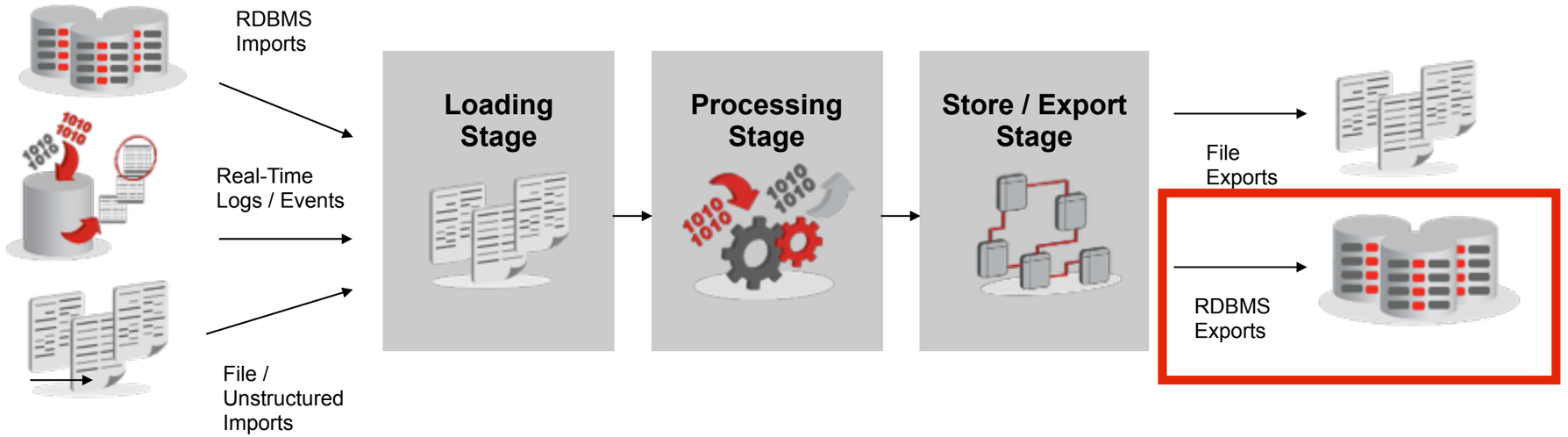
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Alternative to Direct Against Hadoop : Export to Data Mart

- In most cases, for general reporting access, exporting into RDBMS makes sense
- Export Hive data from Hadoop into Oracle Data Mart or Data Warehouse
- Use Oracle RDBMS for high-value data analysis, full access to RBDMS optimisations
- Potentially use Exalytics for in-memory RBDMS access



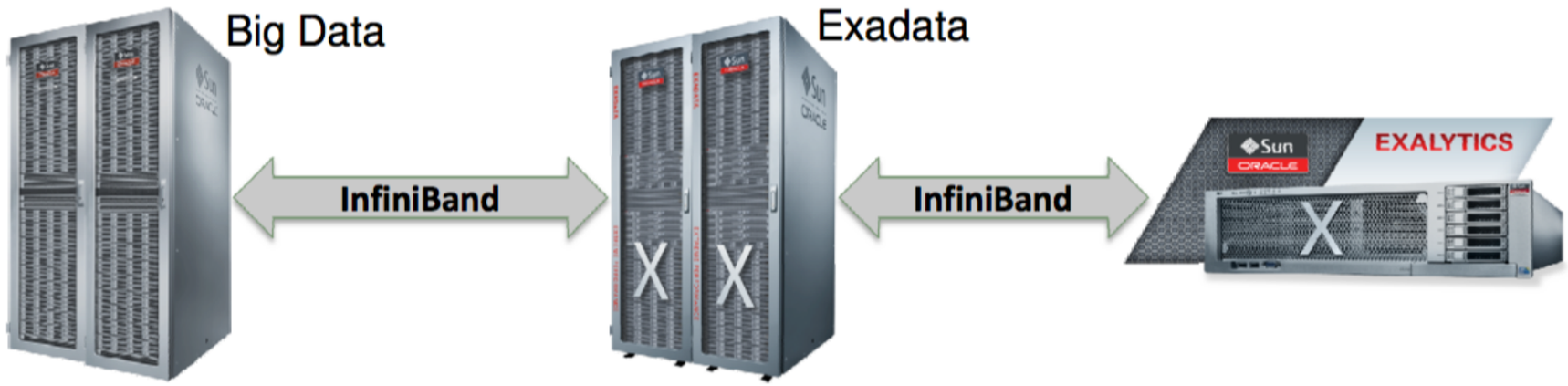
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Using the Right Server for the Right Job

- Hadoop for large scale, high-speed data ingestion and processing
- Oracle RDBMS and Exadata for long-term storage of high-value data
- Oracle Exalytics for speed-of-though analytics in TimesTen and Oracle Essbase



Optimised for Hadoop Processing

Optimised platform for acquiring, organising and loading unstructured data into Oracle RBDMS

Optimised for DW/OLTP Workload

Extreme performance via pre-integrated, optimised Oracle DB platform

Optimised for Analytics Workload

In-memory, speed-of-thought dashboards, reports & analytics

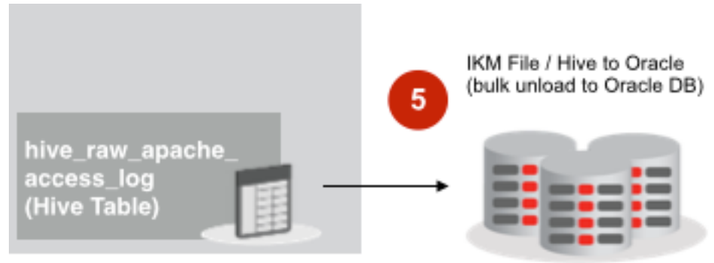
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

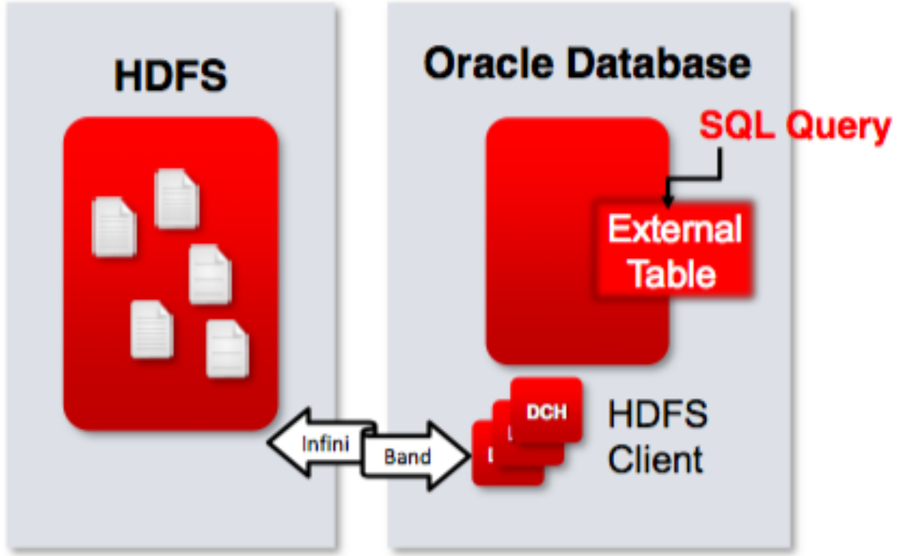
5 Bulk Unload Summary Data to Oracle Database

- Final requirement is to unload final Hive table contents to Oracle Database
- Several use-cases for this:
  - Use Hadoop / BDA for ETL offloading
  - Use analysis capabilities of BDA, but then output results to RDBMS data mart or DW
  - Permit use of more advanced SQL query tools
  - Share results with other applications
- Can use Sqoop for this, or use Oracle Big Data Connectors
  - Fast bulk unload, or transparent Oracle access to Hive



Oracle Direct Connector for HDFS

- Enables HDFS as a data-source for Oracle Database external tables
- Effectively provides Oracle SQL access over HDFS
- Supports data query, or import into Oracle DB
- Treat HDFS-stored files in the same way as regular files
- But with HDFS's low-cost
- … and fault-tolerance

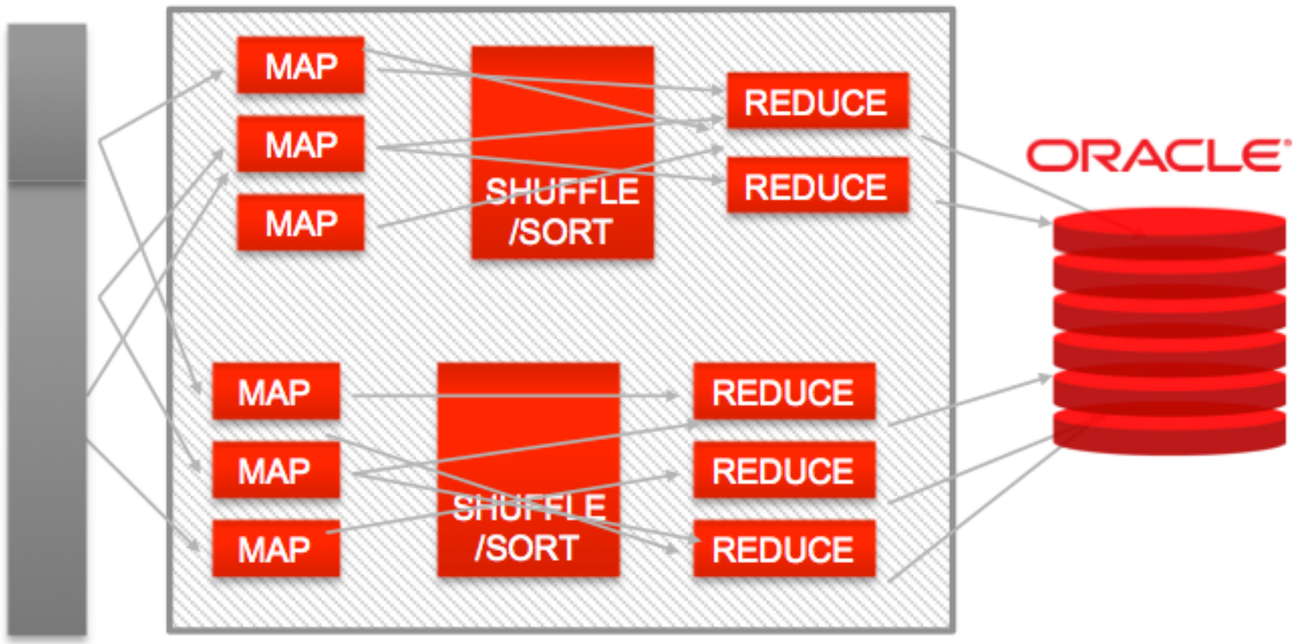


T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

Oracle Loader for Hadoop (OLH)

- Oracle technology for accessing Hadoop data, and loading it into an Oracle database
- Pushes data transformation, “heavy lifting” to the Hadoop cluster, using MapReduce
- Direct-path loads into Oracle Database, partitioned and non-partitioned
- Online and offline loads
- Load from HDFS or Hive tables
- Key technology for fast load of Hadoop results into Oracle DB



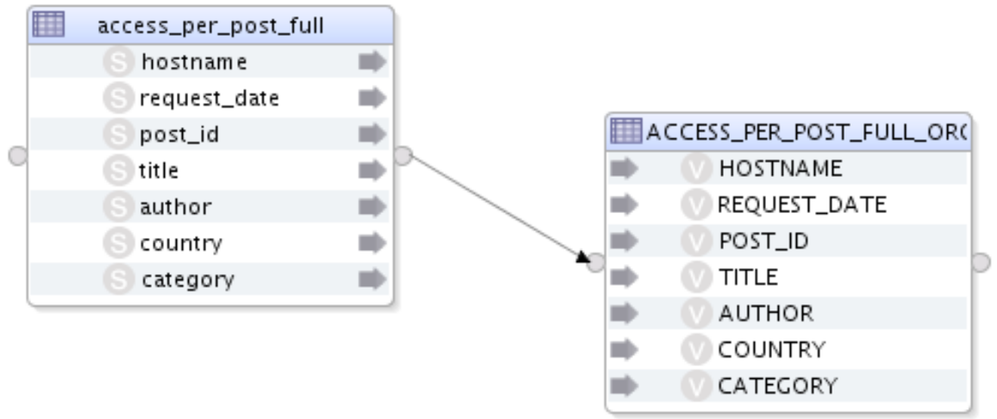
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead INTEGRATED ANALYTICS

IKM File/Hive to Oracle (OLH/ODCH)

- KM for accessing HDFS/Hive data from Oracle
- Either sets up ODCH connectivity, or bulk-unloads via OLH
- Map from HDFS or Hive source to Oracle tables (via Oracle technology in Topology)



T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

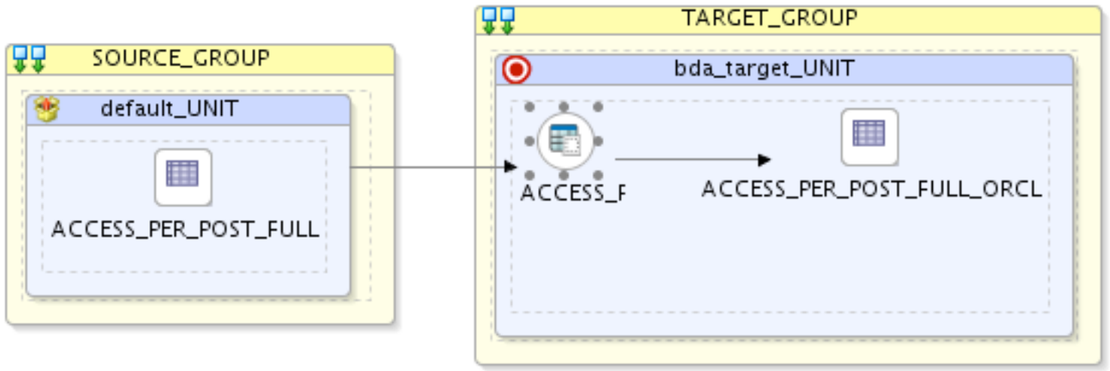
Environment Variable Requirements

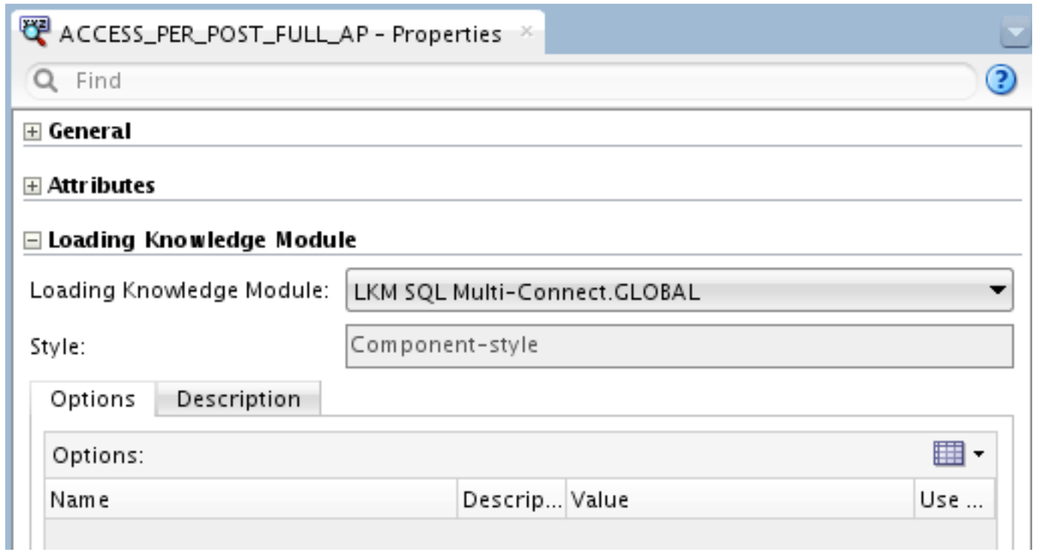
- Hardest part in setting up OLH / IKM File/Hive to Oracle is getting environment variables correct - OLH needs to be able to see correct JARs, configuration files
- Set in /home/oracle/.bashrc - see example below

```
export HIVE_HOME=/usr/lib/hive
export HADOOP_CLASSPATH=/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/*:/etc/hive/conf:$HIVE_HOME/lib/
hive-metastore-0.12.0-cdh5.0.1.jar:$HIVE_HOME/lib/libthrift.jar:$HIVE_HOME/lib/libfb303-0.9.0.jar:$HIVE_HOME/
lib/hive-common-0.12.0-cdh5.0.1.jar:$HIVE_HOME/lib/hive-exec-0.12.0-cdh5.0.1.jar
export OLH_HOME=/home/oracle/oracle/product/oraloader-3.0.0-h2
export HADOOP_HOME=/usr/lib/hadoop
export JAVA_HOME=/usr/java/jdk1.7.0_60
export ODI_HIVE_SESSION_JARS=/usr/lib/hive/lib/hive-contrib.jar
export ODI_OLH_JARS=/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/ojdbc6.jar,/home/oracle/oracle/
product/oraloader-3.0.0-h2/jlib/orai18n.jar,/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/orai18n-
utility.jar,/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/orai18n-mapping.jar,/home/oracle/oracle/
product/oraloader-3.0.0-h2/jlib/orai18n-collation.jar,/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/
oraclepki.jar,/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/osdt_cert.jar,/home/oracle/oracle/product/
oraloader-3.0.0-h2/jlib/osdt_core.jar,/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/commons-
math-2.2.jar,/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/jackson-core-asl-1.8.8.jar,/home/oracle/
oracle/product/oraloader-3.0.0-h2/jlib/jackson-mapper-asl-1.8.8.jar,/home/oracle/oracle/product/
oraloader-3.0.0-h2/jlib/avro-1.7.3.jar,/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/avro-mapred-1.7.3-
hadoop2.jar,/home/oracle/oracle/product/oraloader-3.0.0-h2/jlib/oraloader.jar,/usr/lib/hive/lib/hive-
metastore.jar,/usr/lib/hive/lib/libthrift-0.9.0.cloudera.2.jar,/usr/lib/hive/lib/libfb303-0.9.0.jar,/usr/lib/
hive/lib/hive-common-0.12.0-cdh5.0.1.jar,/usr/lib/hive/lib/hive-exec.jar
```

Configuring the KM Physical Settings

- For the access table in Physical view, change LKM to LKM SQL Multi-Connect
  - Delegates the multi-connect capabilities to the downstream node, so you can use a multi-connect IKM such as IKM File/Hive to Oracle





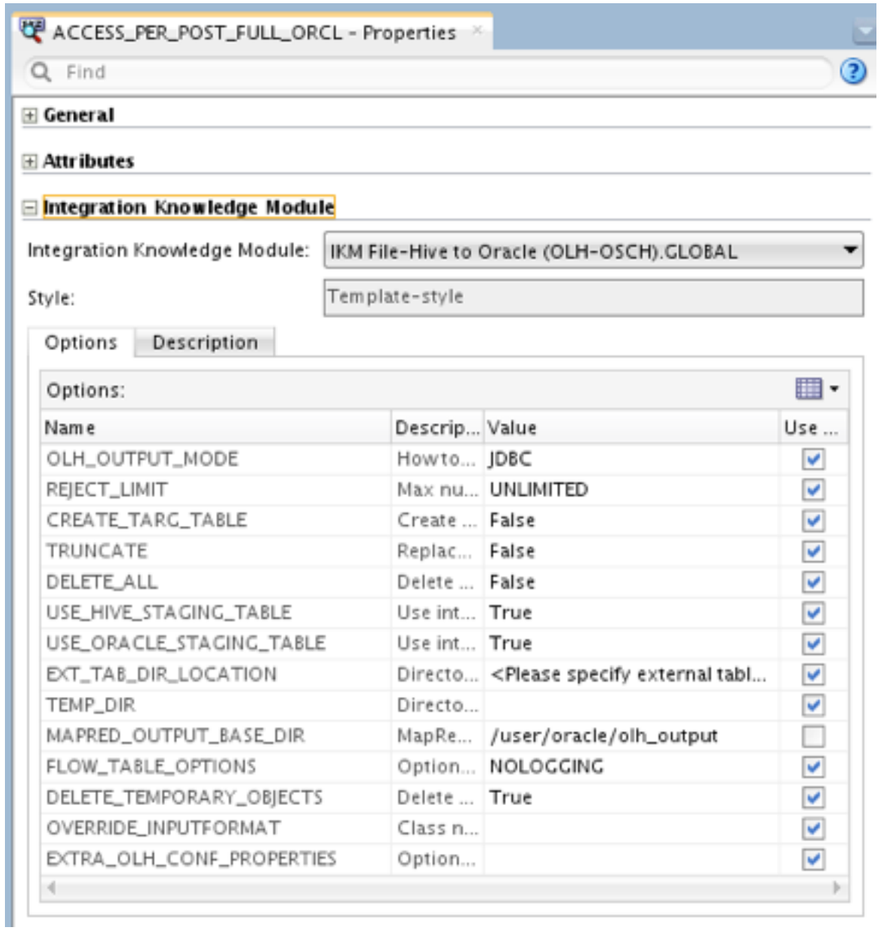
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

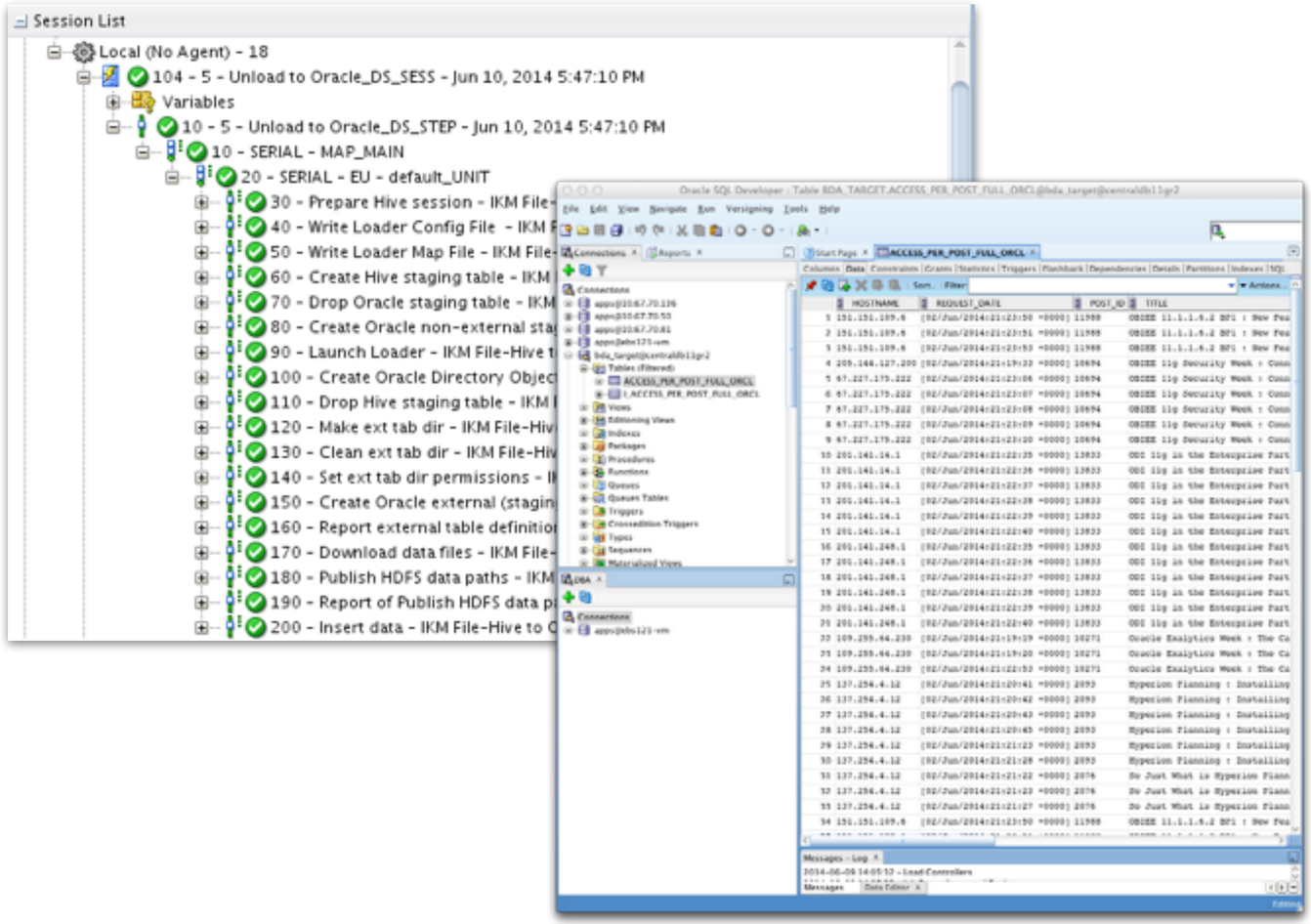
Configuring the KM Physical Settings

- For the target table, select IKM File/Hive to Oracle
  - Only becomes available to select once LKM SQL Multi-Connect selected for access table
  - Key option values to set are:
    - OLH_OUTPUT_MODE (use JDBC initially, OCI if Oracle Client installed on Hadoop client node)
    - MAPRED_OUTPUT_BASE_DIR (set to directory on HFDS that OS user running ODI can access)



Executing the Mapping

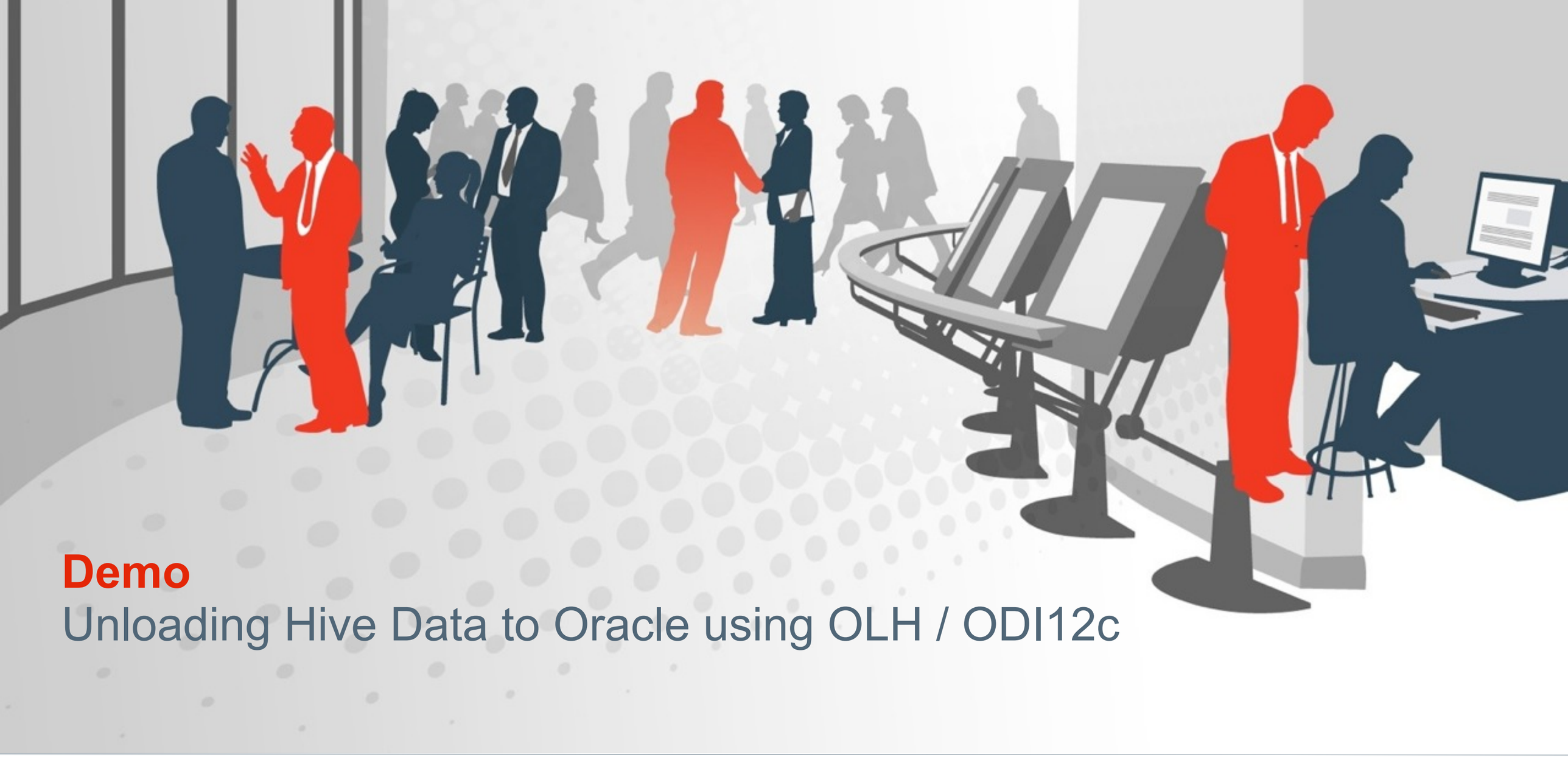
- Executing the mapping will invoke OLH from the OS command line
- Hive table (or HDFS file) contents copied to Oracle table



T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS



Demo

Unloading Hive Data to Oracle using OLH / ODI12c

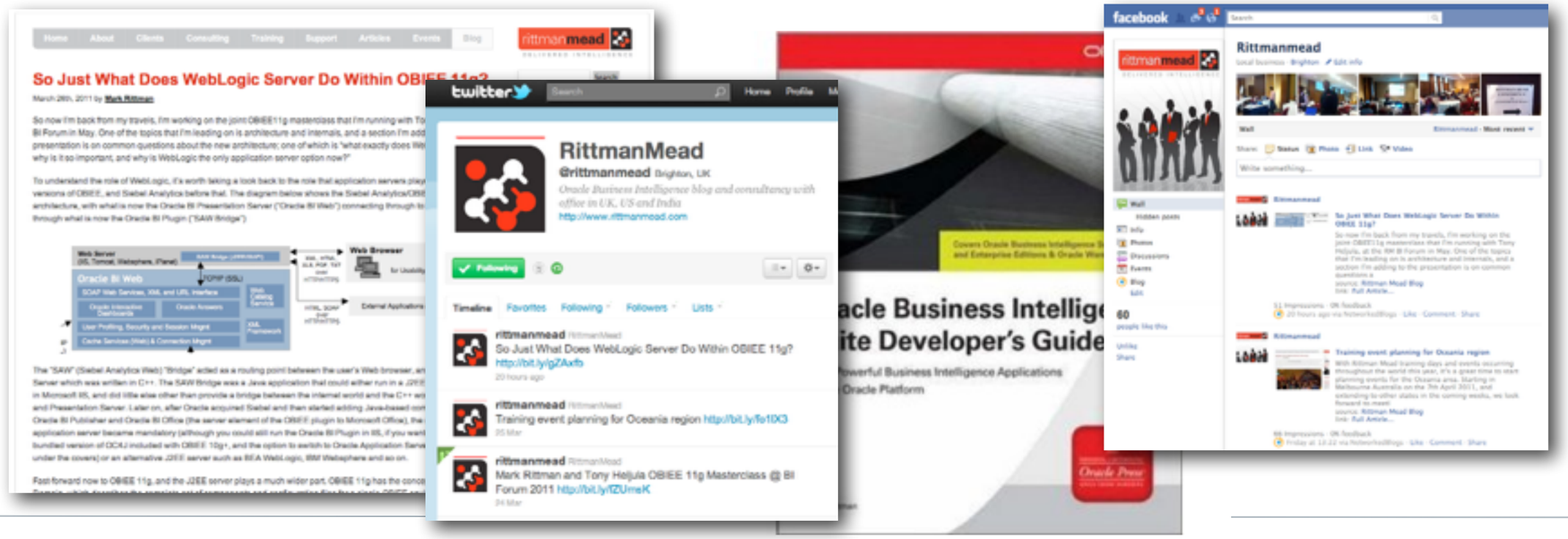
T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Thank You for Attending!

- Thank you for attending this presentation, and more information can be found at http://www.rittmanmead.com
- Contact us at info@rittmanmead.com or mark.rittman@rittmanmead.com
- Look out for our book, “Oracle Business Intelligence Developers Guide” out now!
- Follow-us on Twitter (@rittmanmead) or Facebook (facebook.com/rittmanmead)



T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS

Lesson 4 : Hadoop Reporting and Visualization using OBIEE 11g

Mark Rittman, CTO, Rittman Mead

Oracle Openworld 2014, San Francisco

T : +44 (0) 1273 911 268 (UK) **or** (888) 631-1410 (USA) **or**
+61 3 9596 7186 (Australia & New Zealand) **or** +91 997 256 7970 (India)

E : info@rittmanmead.com
W : www.rittmanmead.com

rittmanmead
INTEGRATED ANALYTICS