Koexistenz?
Oracle und Hadoop gemeinsam nutzen.

Matthias Jung
DOAG Regio 10.09.2019

info@ordix.de www.ordix.de

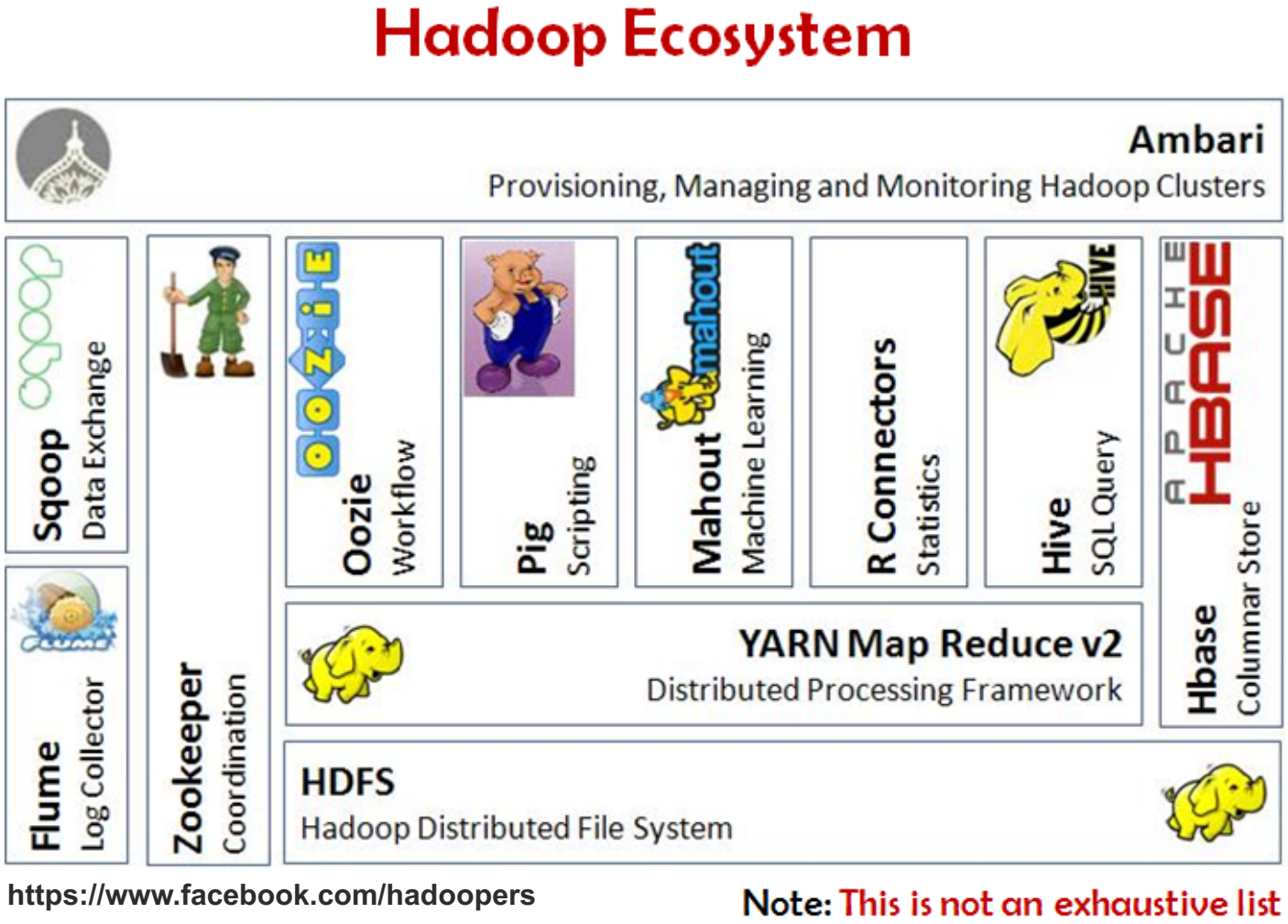
Agenda

ORDIX AG

- Das Hadoop-Ökosystem
- Oracle Big Data Connectors (Auswahl)
- Demonstration: SQL Connector für HDFS

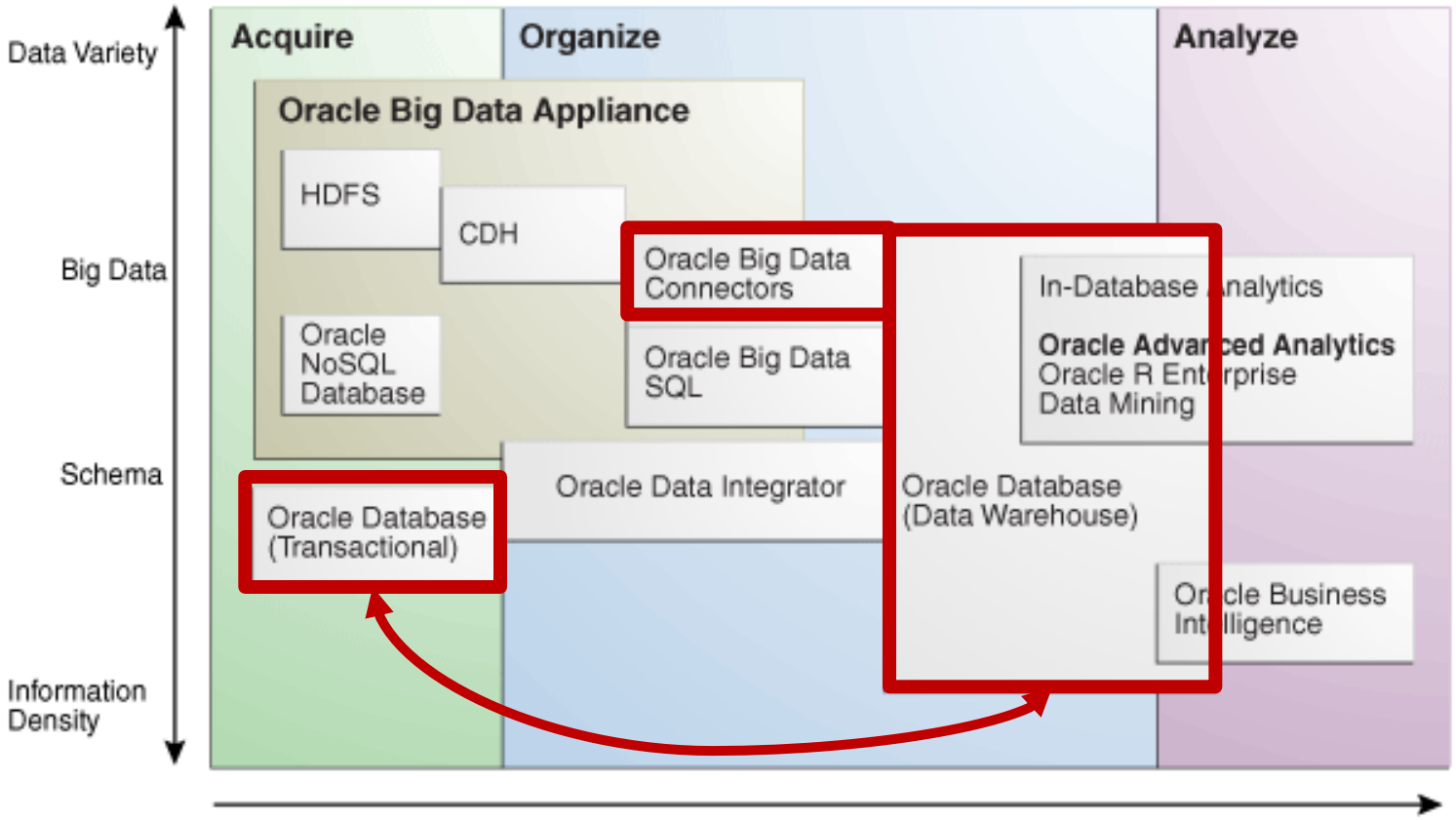
Das Hadoop-Ökosystem

ORDIX AG



Das Oracle Big Data Ökosystem

ORDiX AG



https://docs.oracle.com/cd/E55905_01/doc.40/e55814/concepts.htm#BIGUG119

Worum geht es?

ORDiX AG

- Interaktion zwischen einer „normalen“ Oracle DB und Hadoop (HDFS)

- Worum geht es nicht:
  - Oracle Big Data Appliance / Oracle Big Data SQL
  - Oracle GoldenGate
  - Oracle Database Gateway for ODBC
  - Oracle Data Integrator (ODI)
  - Oracle Advanced Analytics
  - Sonstige ETL Produkte

Agenda

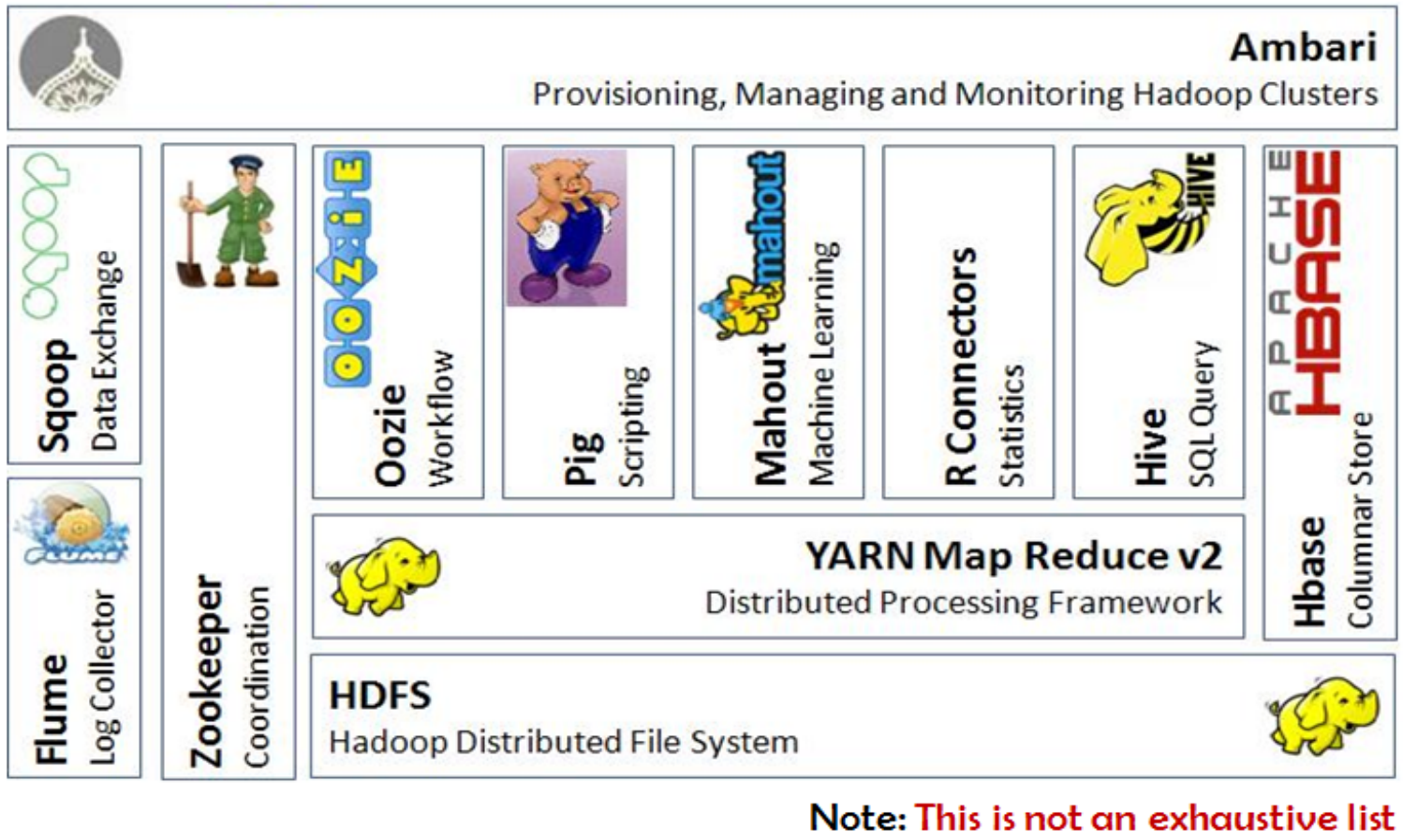
ORDIX AG

- Das Hadoop-Ökosystem
- Oracle Big Data Connectors (Auswahl)
- Demonstration: SQL Connector für HDFS

Big Data Connectors

ORDIX AG

- **Oracle Dat**
- **Oracle Loa**
- Oracle R A
- Oracle XQ
- Oracle Dat
- **Oracle SQ**



https://www.facebook.com/hadoopers

Voraussetzungen

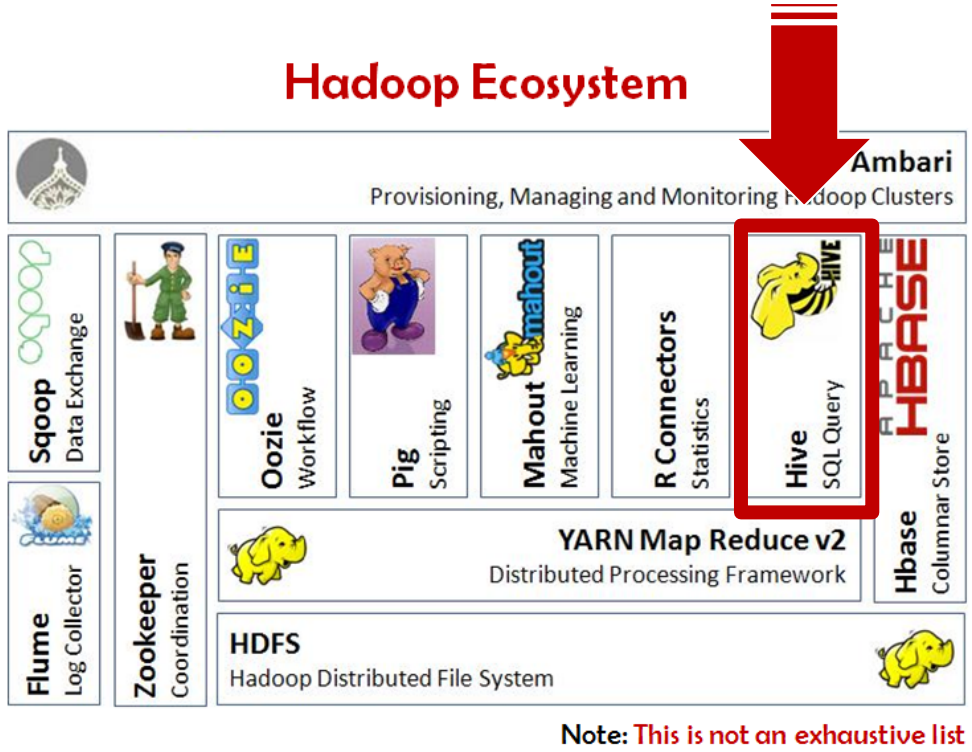
ORDiX AG

https://www.bigdata-insider.de/cloudera-und-hortonworks-vereinigen-ihre-datenplattformen-a-817696/

Oracle Datasource for Apache Hadoop

ORDIX AG

- Hive QL- / Spark SQL-Zugriff auf Oracle-Tabellen
- Oracle dient als Quelle
- Daten in Hadoop können mit Oracle Daten „gejoined“ werden
- Auf Seiten der Oracle DB ist nicht viel zu tun:
  - Bereitstellen der Tabellenstruktur
  - User-Account



https://www.facebook.com/hadoopers

Oracle Datasource for Apache Hadoop

ORDIX AG

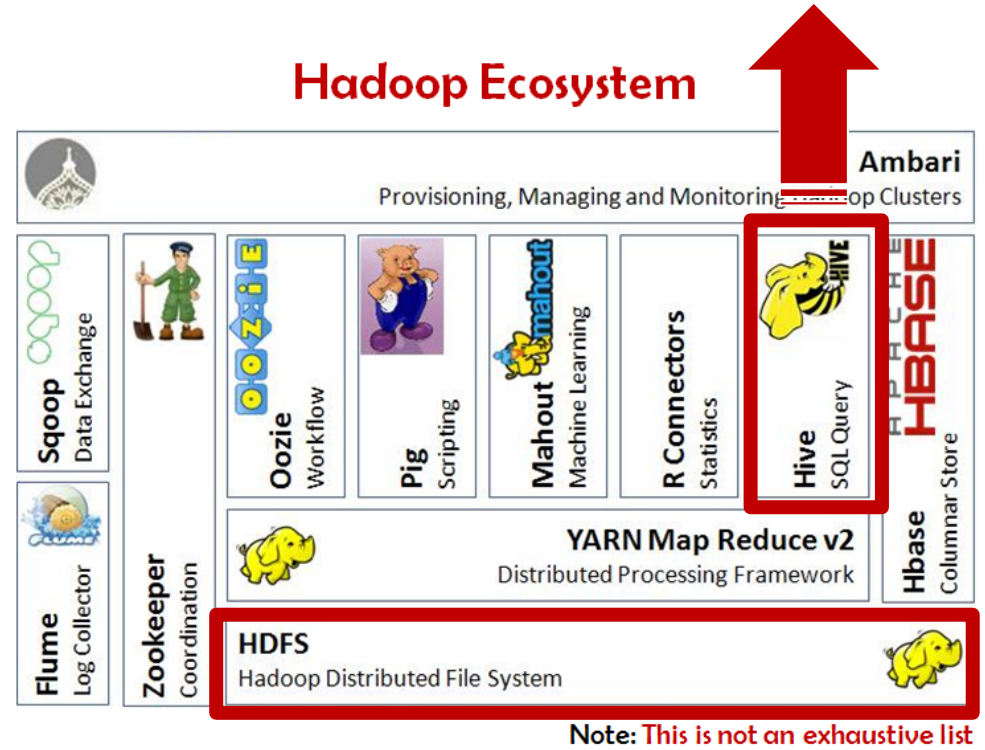
- Beispiel:

```
hive> CREATE EXTERNAL TABLE ora_tab
            STORED BY 'oracle.hcat.osh.OracleStorageHandler' TBLPROPERTIES (
            'mapreduce.jdbc.url' = 'jdbc:oracle:thin:@//oraclehost:1521/prod',
            'mapreduce.jdbc.username' = 'scott',
            'mapreduce.jdbc.password' = 'tiger',
            'mapreduce.jdbc.input.table.name' = 'employee',
            'oracle.hcat.osh.useMonitor'='true',
            'oracle.hcat.osh.fetchSize'='10000',
            'oracle.hcat.osh.useOracleParallelism'='true'
);
```

Oracle Loader for Hadoop

ORDIX AG

- „ETL“-Tool, welches in der Lage ist verschiedene Hadoop-Quellen in eine Oracle DB zu laden
- Hadoop ist die Quelle
- Formate
  - Text Files (z.B. aus HDFS)
  - Hive-Tabellen
    - Parquet
    - JSON
- Zusatzfunktionen
  - Datentransformationen (Typen)
  - Kompression (in der DB)
- Online- / Offline-Modus



https://www.facebook.com/hadoopers

Online Database Mode

ORDIX AG

- Ablauf
  - OLH verbindet sich zur Laufzeit mit der Zieldatenbank
  - Metadaten der Zieltabelle werden gelesen
  - Daten werden aus Hadoop gelesen und direkt in die Zieltabelle geschrieben
- Vorteile
  - Einfach zu konfigurieren
  - Einfach zu betreiben
- Nachteile
  - Verbindung von Hadoop zur Datenbank notwendig
  - Höhere Last auf Oracle als im Offline Mode
  - Nicht so schnell wie Offline mit Data Pump-Dateien

Offline Database Mode

ORDIX AG

- Ablauf
  - Metadaten der Zieltabelle werden aus Datei gelesen
  - Daten werden aus Hadoop gelesen
  - Daten werden in HDFS-Datei geschrieben (Data Pump-Datei)
- Vorteile
  - Zur Laufzeit keine Verbindung zur Datenbank notwendig
  - Offline mit Data Pump-Dateien ist am schnellsten
  - Geringste Last auf der Oracle-Datenbank
- Nachteile
  - Aufwendiger zu konfigurieren
  - Datei Transfer zusätzlich notwendig

Oracle Loader for Hadoop

ORDIX AG

- Beispiel:
  (https://blogs.oracle.com/bigdataconnectors/how-to-load-oracle-tables-from-hadoop-tutorial-part-2-hello-world)

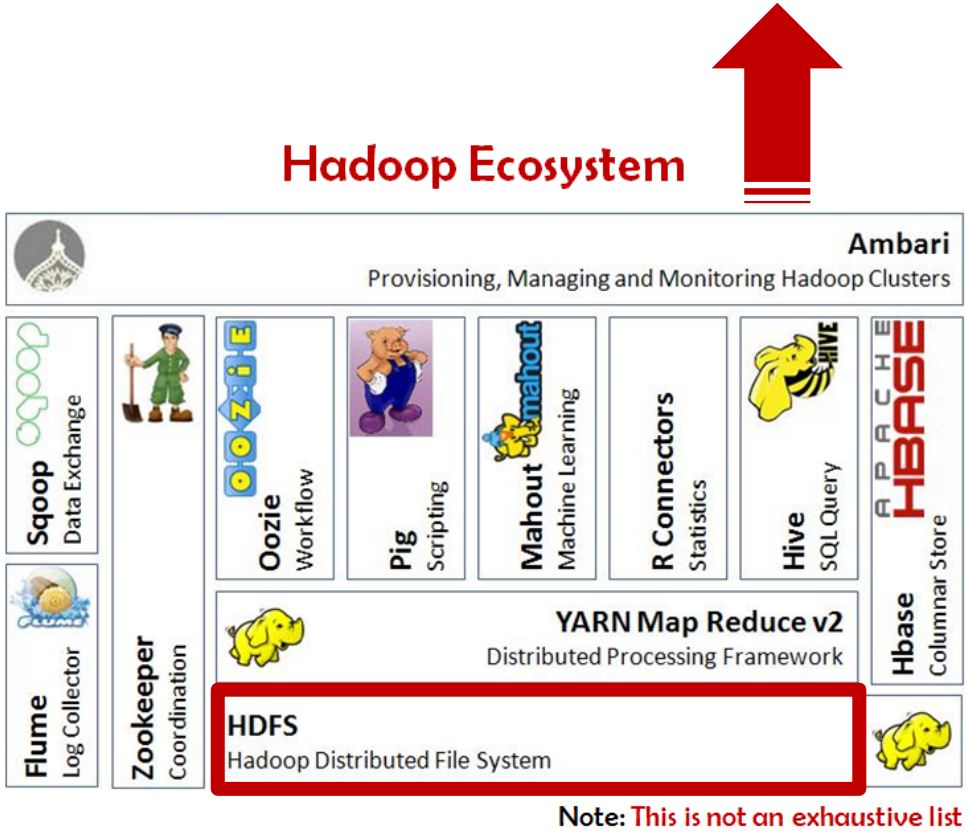
```
bash> $HADOOP_HOME/bin/hadoop jar $OLH_HOME/jlib/oraloader.jar
oracle.hadoop.loader.OraLoader
  -D oracle.hadoop.loader.jobName=OLHP_fivdti_dtext_jdbc_0_722
  -D oracle.hadoop.loader.loaderMapFile=file:/tmp/loaderMap_fivdti.xml
  -D mapred.reduce.tasks=0
  -D mapred.input.dir=/user/olh_performance/fivdti/56000000_90
  -D mapred.output.dir=/user/oracle/olh_test/results/fivdti/722
  -conf /tmp/oracle_connection.xml
  -conf /tmp/dtextInput.xml
  -conf /tmp/jdbcOutput.xml
```

(Online: Direkt über JDBC in die DB)

Oracle SQL Connector for HDFS

ORDIX AG

- SQL Interface auf HDFS-Dateien
- HDFS dient als Quelle für die Abfrage (SELECTs)
- „Read-access“
- Zugriff wird über „external tables“ realisiert
- Kann zum Laden von Daten in die DB genutzt werden



https://www.facebook.com/hadoopers

Agenda

ORDIX AG

- Das Hadoop-Ökosystem
- Oracle Big Data Connectors (Auswahl)
- Demonstration: SQL Connector für HDFS

SQL Connector for HDFS

ORDIX AG

- Installation
- Hadoop Client auf der Datenbank installieren
  ```
  [Cloudera YUM Repository installieren]
  yum install hadoop-client
  ```
- HDFS-Konfiguration von HDFS Name Node kopieren
  ```
  /etc/hadoop/conf/hdfs-site.xml
  /etc/hadoop/conf/core-site.xml
  ```
- Zugriff des Hadoop-Clients testen
  ```
  hadoop fs –ls /user/oracle/
  ```

SQL Connector for HDFS

ORDIX AG

- Oracle SQL Connector installieren
  https://www.oracle.com/technetwork/database/database-technologies/bdc/big-data-connectors/downloads/index.html

```
bash> unzip orahdfs-3.8.2.zip
```

- Umgebung anpassen

```
export JAVA_HOME=/usr
export OSCH_HOME=/home/oracle/orahdfs-3.8.2
export OSCH_BIN_PATH=$OSCH_HOME/bin
export HADOOP_CLASSPATH=$OSCH_HOME/jlib
```

SQL Connector for HDFS

ORDIX AG

- Oracle DB-Objekte anlegen:

```
# User anlegen
SQL> create user hdfsuser identified by hdfs;
SQL> grant create session, create table, create view to hdfsuser;
SQL> grant execute on sys.utl_file to hdfsuser;

# Binary access
SQL> create directory osch_bin_path as AS '/home/oracle/orahdfs-3-8-2/bin‘;
SQL> SQL GRANT READ, EXECUTE ON DIRECTORY OSCH_BIN_PATH TO hdfsuser;
SQL> create directory external_table_dir as '/home/oracle/hdfs_tables‘;
SQL> GRANT READ, WRITE ON DIRECTORY external_table_dir TO hdfsuser;
```

- Externe Tabelle generieren (Manifest regelt die Details):

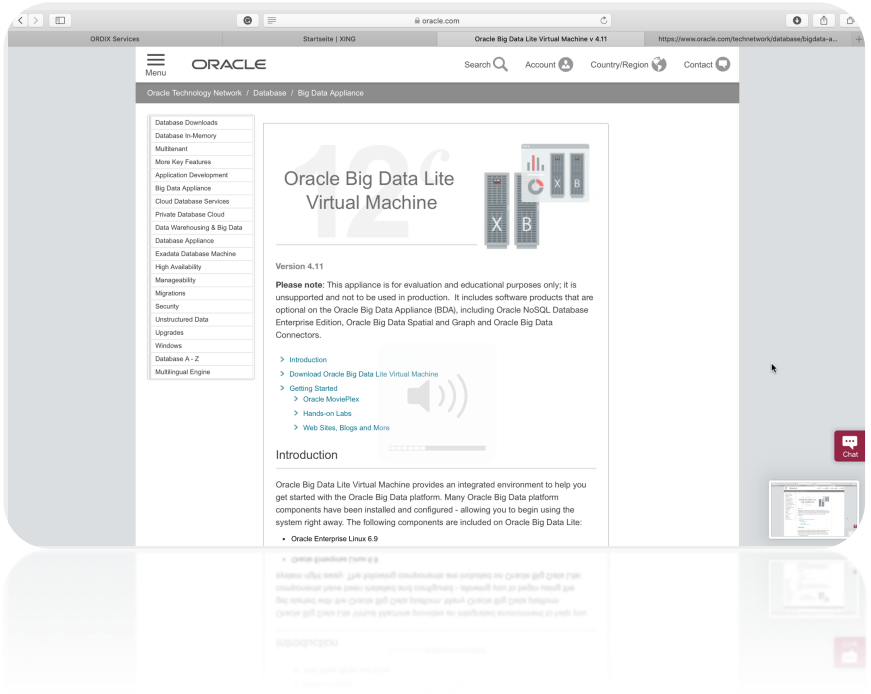
```
hadoop jar /home/oracle/orahdfs-3.8.2/jlib/orahdfs.jar
oracle.hadoop.exttab.ExternalTable -conf manifest.xml -createTable
```

Hinweis

ORDIX AG

- Oracle bietet eine komplette virtuelle Spielwiese an:
  https://www.oracle.com/technetwork/database/bigdata-appliance/oracle-bigdatalite-2104726.html

- Voraussetzung:
  - 2 Cores
  - 5 GB RMAN
  - 50 GB Disk



ORDIX® best practice
einfach. gut. beraten.

Vielen Dank für Ihre Aufmerksamkeit

ORDIX AG
Aktiengesellschaft für
Softwareentwicklung, Schulung,
Beratung und Systemintegration

Zentrale Paderborn
Karl-Schurz-Straße 19a
33100 Paderborn
Tel.: 05251 1063-0
Fax: 0180 1 67349 0

Seminarzentrum Wiesbaden
Kreuzberger Ring 13
65205 Wiesbaden
Tel.: 0611 77840-00

info@ordix.de
www.ordix.de