EVALUATION OF

ORACLE BIG DATA INTEGRATION TOOLS

Harun Urhan

supervised by
Zbigniew Baranowski
CONTENT

- Overview of Hadoop
- Oracle Data Integrator: Application Adapters for Hadoop
- Oracle SQL Connectors for HDFS
BIG DATA? HADOOP?

BIG DATA & HADOOP
HOW? WHAT IS IT?
BIG DATA? HADOOP?

LOOK GUYS...

IT'S SO BIG!!!!!!
BIG DATA? HADOOP?

- high scalability
- distributed storage
- “big data” processing
ORACLE DATA INTEGRATOR: APPLICATION ADAPTERS FOR HADOOP
ORACLE DATA INTEGRATOR: APPLICATION ADAPTERS FOR HADOOP

- graphical
- drag & drop
- realtime task progress tracking
ORACLE DATA INTEGRATOR:
APPLICATION ADAPTERS FOR HADOOP

• graphical
• drag & drop
• realtime task progress tracking
RESULTS
RESULTS

- configured for and tested on CERN infrastructure
- performance measured - 3 MB/s of data transfer
- in parallel mode - overall throughput scales
RESULTS

- configured for and tested on CERN infrastructure
- performance measured - 3 MB/s of data transfer
- in parallel mode - overall throughput scales

- non centralized
- learning curve
- Sqoop is CPU bound (hadoop side)
ORACLE SQL CONNECTORS FOR HDFS

- query HDFS
- with SQL
- from Oracle
ORACLE SQL CONNECTORS FOR HDFS

- query HDFS
- with SQL
- from Oracle
RESULTS
RESULTS

- installed and tested on CERN infrastructure
- tested with different configurations (ex: compression, parallelism)
- performance measured - 20 MB/s per process
RESULTS

- ✔ installed and tested on CERN infrastructure
- ✔ tested with different configurations (ex: compression, parallelism)
- ✔ performance measured - 20 MB/s per process

- ⚠ bottleneck on client side - CPU bound
- ⚠ only text file format support
SUMMARY

- Data integration cycle successfully run on CERN infrastructure
- OSCH is promising but architectural enhancements would be better
- ODI is a complete tool and ready to use
Thank you