Advantages of integrating Oracle and Hadoop

• Best of two worlds:
  • Oracle: Online Transactional System
  • Hadoop: Large-Scalable Data Warehouse

• Hybrid systems:
  • Move (read-only) data from Oracle to Hadoop
  • Query Hadoop data from Oracle (using Oracle APIs)
  • Also possible: query Oracle from Hadoop

• Increase scalability and lower ratio cost/performance
  • Hadoop data formats and engines for high performance analytics
  • ….without need of changing the end-user apps connecting to Oracle
Offloading from Oracle to Hadoop

• Step1: Offload **data** to Hadoop
  - Oracle database
  - Table data export
  - Apache Sqoop
  - Hadoop cluster
  - Data formats: **Parquet**, **Avro**

• Step2: Offload **queries** to Hadoop
  - Offload interface: DB LINK, External table
  - SQL engines: **Impala**, **Hive**
Making data sources transparent to end-user

- Hybrid views on Oracle
  - Recent (read-write) data in Oracle
  - Archive data in Hadoop
  - Advantages: Hadoop performance with unchanged applications (Oracle APIs)

```sql
create view hybrid_view as
  select * from online_table where date > '2016-10-01'
union all
  select * from archival_table@hadoop where date <= '2016-10-01'
```

Split point has to be updated after each successful data offload