Options for Connecting Hadoop to Oracle

There’s a comprehensive comparison matrix made by Tanel

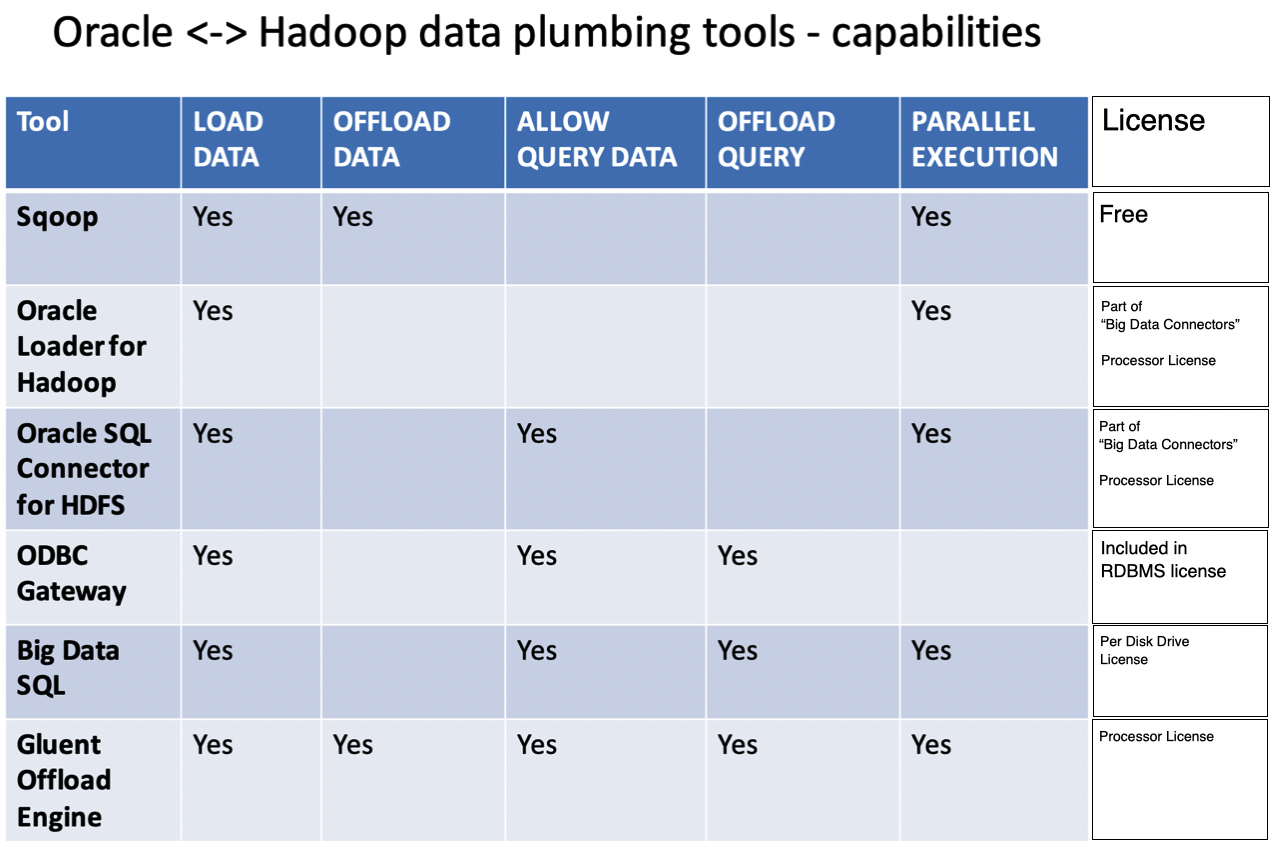
<http://www.ooug.org/wp-content/uploads/2016/05/Tanel_Poder_Connecting_Hadoop_and_Oracle.pdf>

which is also available on video <https://www.youtube.com/watch?v=gSECYOi-9eA>

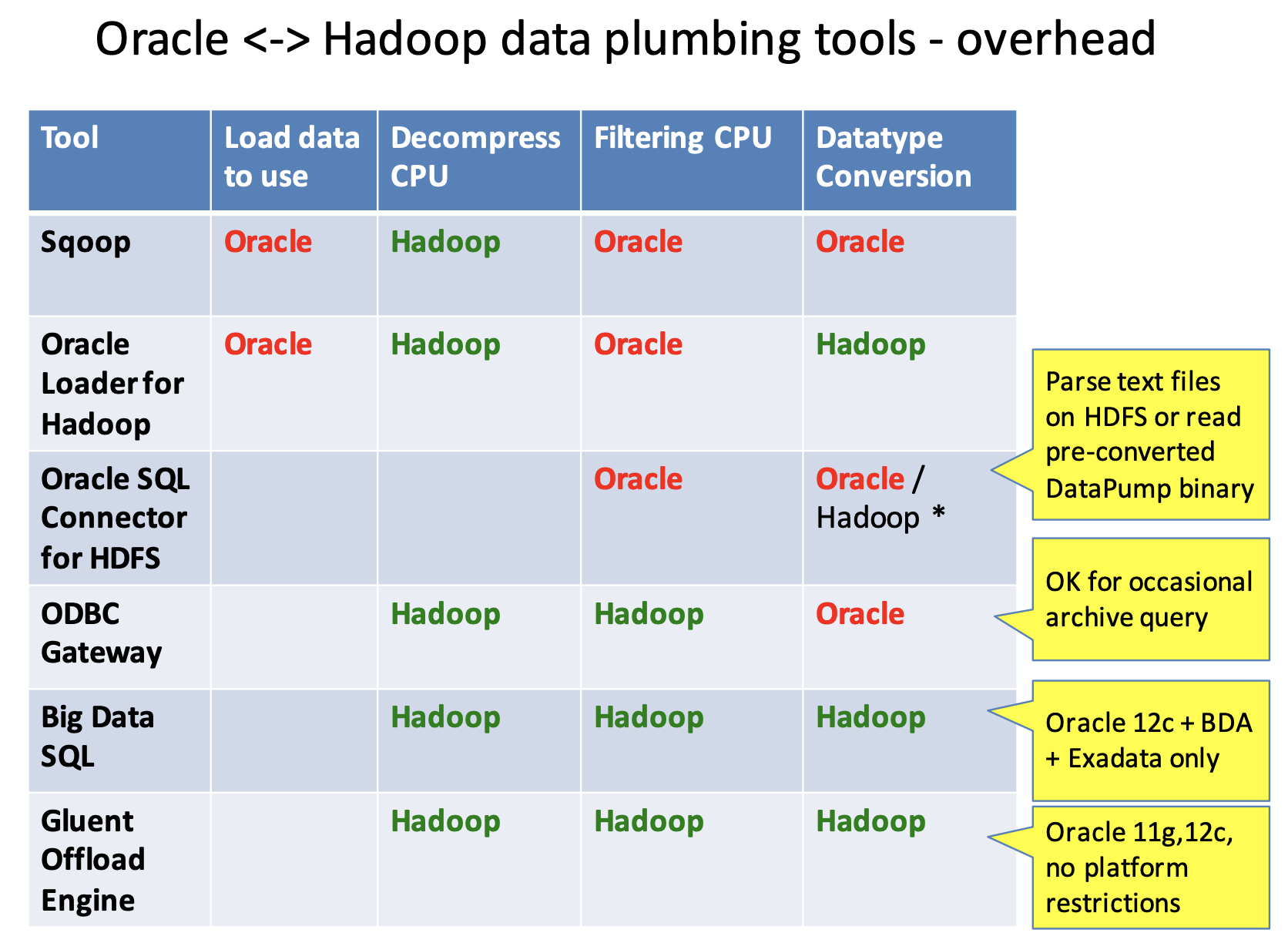
Below is the matrix comparison from Tanel’s deck. I added the License column for completeness.

Here’s the summary:

* You can load and unload data for FREE. This can be achieved by loading/reading the Hadoop data using ODBC Gateway and unload data using Sqoop
  + PRO
    - Free
  + CONS
    - ODBC gateway SQL is a serial session so if you are reading a big Hive table then expect slow performance
    - Complex SQLs could be an issue (nested subqueries, etc.) and should be pushed down to the Hive database as a new table to simplify the Oracle SQL
* Using Oracle “Big Data Connectors” (old way) and “Big Data SQL” (new way)
  + PRO
    - Allows parallel execution and complex SQLs with filter pushdown
  + CON
    - Big Data SQL – per disk drive license (just like Exadata Storage Cells)
* Using Gluent
  + PRO
    - Transparently access Oracle + Hadoop table (hybrid view) with complex SQLs. Same object name is used, that means no app change
    - Allows parallel execution on data Load/Offload with filter pushdown
    - Management tools included to maintain hybrid views and schedule offload
  + CON
    - 3rd party software – per processor license (only on Hadoop side)



Below matrix shows where the overhead is when doing read/load/unload operations:

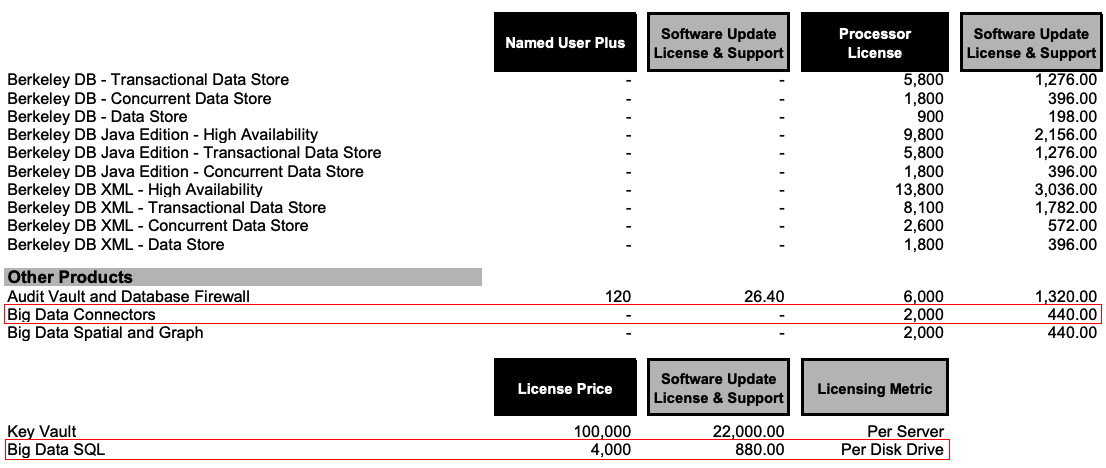


Here are the other references:

* Integration of Oracle and Hadoop at CERN: hybrid databases affordable at scale
  + <https://canali.web.cern.ch/canali/docs/Integration_Oracle_Hadoop_LC_CHEP2016.pdf>
* Oracle Big Data Connectors
  + <https://www.abis.be/resources/presentations/obug20130326bigdata.pdf>
* Oracle and Hadoop, let them talk together
  + <https://www.doag.org/formes/pubfiles/9479832/2017-DA-Laurent_Leturgez-Hadoop_and_Oracle__Let_Them_Talk_Together__-Manuskript.pdf>
* How To Query Hive and Impala from Oracle using ODBC Heterogeneous Gateway
  + <https://community.oracle.com/docs/DOC-1002634>
* Oracle Big Data SQL - copy2hadoop
  + <https://blogs.oracle.com/datawarehousing/data-loading-into-hdfs-part2-data-movement-from-the-oracle-database-to-the-hdfs>
  + <https://docs.oracle.com/bigdata/bds31/BDSUG/copy2bda.htm#BIGUG76738>

Price List:

* <https://www.oracle.com/assets/technology-price-list-070617.pdf>



* Big Data Connectors User’s Guide
  + <https://docs.oracle.com/cd/E27101_01/doc.10/e27365/start.htm>
* Big Data SQL Licensing
  + <https://docs.oracle.com/bigdata/bds31/BDSIG/licensing.htm#BDSIG-GUID-8E911C46-3FAA-44BA-BCEE-FB9EF94E494E>
* Database Gateway and Generic Connectivity (DG4ODBC) Licensing Considerations (Doc ID 232482.1)

x