Cloudera Impala

Cloudera Impala is Cloudera's open source massively parallel processing (MPP) SQL query engine for data stored in a computer cluster running Apache Hadoop.[1]

1 Description

Cloudera Impala is a query engine that runs on Apache Hadoop. The project was announced in October 2012 with a public beta test distribution[2][3] and became generally available in May 2013.[4]

The Apache-licensed Impala project brings scalable parallel database technology to Hadoop, enabling users to issue low-latency SQL queries to data stored in HDFS and Apache HBase without requiring data movement or transformation. Impala is integrated with Hadoop to use the same file and data formats, metadata, security and resource management frameworks used by MapReduce, Apache Hive, Apache Pig and other Hadoop software.

Impala is promoted for analysts and data scientists to perform analytics on data stored in Hadoop via SQL or business intelligence tools. The result is that large-scale data processing (via MapReduce) and interactive queries can be done on the same system using the same data and metadata – removing the need to migrate data sets into specialized systems and/or proprietary formats simply to perform analysis.

Features include:

- Supports HDFS and Apache HBase storage
- Reads Hadoop file formats, including text, LZO, SequenceFile, Avro, RCFile, and Parquet
- Supports Hadoop security (Kerberos authentication)
- Fine-grained, role-based authorization with Sentry[5]
- Uses metadata, ODBC driver, and SQL syntax from Apache Hive

In early 2013, a column-oriented file format called Parquet was announced for architectures including Impala.[6] In December 2013, Amazon Web Services announced support for Impala.[7] In early 2014, MapR added support for Impala.[8]

2 See also

- Dremel — similar tool from Google
- Apache Drill — similar open source project inspired by Dremel

3 References

[1] "Cloudera Impala". Retrieved 14 March 2014.

[2] Larry Digna (October 24, 2012). "Cloudera aims to bring real-time queries to Hadoop, big data". *Between the lines blog*. ZDNet. Retrieved January 20, 2014.

[3] Andrew Brust (October 25, 2012). "Cloudera's Impala brings Hadoop to SQL and BI". *ZDNet*. Retrieved January 20, 2014.

[4] Marcel Kornacker, Justin Erickson (May 1, 2013). "Cloudera Impala 1.0: It's Here, It's Real, It's Already the Standard for SQL on Hadoop". Retrieved April 10, 2014.

[5] Sentry

[6] "Parquet: Columnar Storage for Hadoop". *Project web site*. 2013. Retrieved January 20, 2014.

[7] "Announcing Support for Impala with Amazon Elastic MapReduce". Amazon.com. December 12, 2013. Retrieved January 20, 2014.

[8] "Impala for MapR". MapR.com. February 2, 2014. Retrieved April 10, 2014.

4 External links

- Cloudera Impala commercial web site
- Impala GitHub project source code
- Impala Project Page project web site

5 Text and image sources, contributors, and licenses

5.1 Text

- **Cloudera Impala** *Source:* http://en.wikipedia.org/wiki/Cloudera_Impala?oldid=631605375 *Contributors:* Billhpike, Dlohcierekim, FrescoBot, W Nowicki, Elicollins, BG19bot, Imkevinyang, Filedelinkerbot, AB3L and Anonymous: 4

5.2 Images

5.3 Content license