CS516 Project-0

Chengen Xie

Jan 26th 2015

**1010data**

https://www.1010data.com/uploads/files/1010data\_Robin\_Bloor,\_Ph\_D,\_Big\_Data\_Analytics\_white\_paper.pdf

* High-performance “big data” database
* Interactive analysis capability with financial and risk management functions, as well as time series analysis and text analysis
* Fast to implement
* Also can be a stand-alone database

**Apache Hive**

http://en.wikipedia.org/wiki/Apache\_Hive

* Using map-reduce
* Acceleration with indexing
* Base on SQL and translate into Hadoop for execution and using Hadoop ecosystem
* Built-in and external tools for data mining
* Lack support for “transactions and materialized views”
* Not strictly follow SQL standard

**Apache Drill**

<http://en.wikipedia.org/wiki/Apache_Drill> <https://www.mapr.com/blog/top-10-reasons-using-apache-drill-now-part-mapr-distribution-including-hadoop#.VMcB_4es7D0>

* Open source
* Data-intensive distributed computing system and support large dataset
* Base on Google’s Big Query and support SQL queries
* Provides interactive analysis

**Apache Storm**

<http://www.sqlstream.com/wp-content/uploads/2014/02/SQLstream_STORM_datasheet.pdf>

<https://storm.apache.org/about/integrates.html>

Inherent parallelism with low latency

Good integration with other queuing system

Simple API

Fault-tolerant

**AWS RedShift**

<http://en.wikipedia.org/wiki/Amazon_Redshift>

* Cloud computing support with massive parallel processing technology and therefore can handle large dataset
* Can use analytical and BI tools

**Azure SQL Database**

<http://en.wikipedia.org/wiki/SQL_Azure>

* Allow making relational queries
* Cloud-based and data storage
* Backend by Microsoft SQL

**Cloudant**

<http://en.wikipedia.org/wiki/Cloudant>

* Open source
* Mapreduce

**couchdb**

<http://couchdb.apache.org>

* Accessibility through web and html
* [distributed scaling](http://en.wikipedia.org/wiki/CAP_theorem)
* fault-tolerant
* Using JS for Mapreduce
* Open source

**Database.com**

<http://www.salesforce.com/platform/database/>

* Cloud-based

**Big Query**

<http://en.wikipedia.org/wiki/BigQuery>

* incorporated with Mapreduce

**EnterpriseDB**

<http://en.wikipedia.org/wiki/EnterpriseDB>

* Time-proof reliable technology foundation PostgreSQL
* Open source with good developer community
* Connections with many common programming languages

**Firebird**

<http://en.wikipedia.org/wiki/Firebird_(database_server>)

http://www.itexto.net/devkico/?p=398

* Scalable databse
* Low hardware requirement
* Free