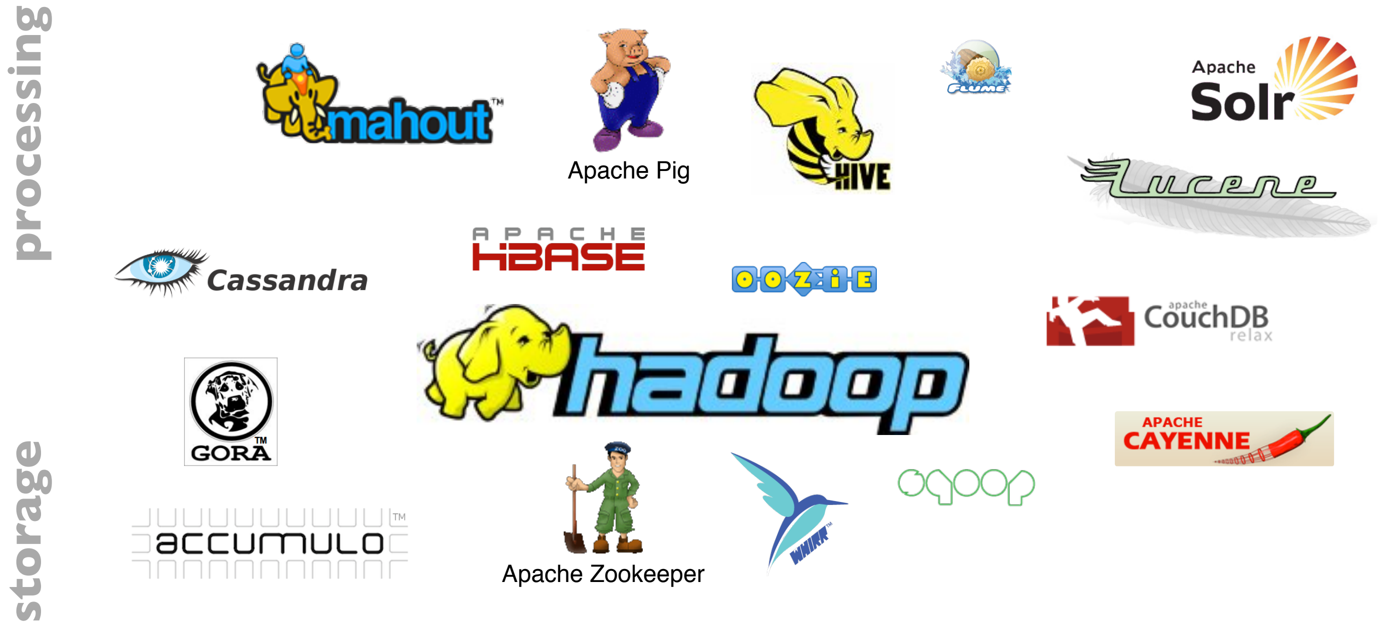
Apache Big Data

projects.apache.org



Accumulo - a sorted, distributed key/value store
http://accumulo.apache.org/

Cassandra - column-oriented database
http://cassandra.apache.org/

Cayenne - object-relational mapping (ORM)
http://cayenne.apache.org/

CouchDB - NoSQL document-oriented datastore
http://couchdb.apache.org/

Gora - in-memory data model & persistence
http://gora.apache.org/

Hadoop - a distributed computing platform:

- HDFS - distributed file system for Hadoop
- MapReduce - parallel computation on clusters

http://hadoop.apache.org/

HBase - column-oriented database on top of Hadoop
http://hbase.apache.org/

Hive - data warehouse with SQL-like access
http://hive.apache.org/

Flume - collection & import of log and event data
http://flume.apache.org/

Lucene - indexing and search
http://lucene.apache.org/

Mahout - library for machine learning & data mining
http://mahout.apache.org/

Pig - high-level programming language for Hadoop
http://pig.apache.org/

Oozie - workflow management for Hadoop
http://oozie.apache.org/

Solr - Lucene-based enterprise search platform
http://lucene.apache.org/solr/

Sqoop - imports data from RDBMS into Hadoop
http://sqoop.apache.org/

Whirr - cloud-agnostic deployment of clusters
http://whirr.apache.org/

Zookeeper - configuration & coordination
http://zookeeper.apache.org/

Apache Big Data

incubator.apache.org

Ambari - deployment, configuration & monitoring of Hadoop clusters
http://incubator.apache.org/ambari/

V V V

Blur - search platform for searching massive amounts of data in a cloud computing environment
http://incubator.apache.org/blur/

V **V** V

Chukwa - log collection & analysis framework for Hadoop clusters
http://incubator.apache.org/chukwa/

V V V

Crunch - a Java library for writing, testing & running pipelines of MapReduce jobs
http://incubator.apache.org/crunch/

V V V

Drill - interactive analysis of large-scale dataset in seconds
http://incubator.apache.org/drill/

V V **V**

HCatalog - schema & data type sharing over Pig, Hive & MapReduce
http://incubator.apache.org/hcatalog/

V V V

Kafka - distributed publish-subscribe messaging system
http://incubator.apache.org/kafka/

V V V

Mesos - a cluster manager providing resource sharing & isolation across cluster applications
http://incubator.apache.org/mesos/

V V V

S4 - platform for processing continuous unbounded streams of data
http://incubator.apache.org/s4/

V V **V**

Tashi - infrastructure for service providers to build applications harnessing cluster computing resources to efficiently access repositories of rich data
http://incubator.apache.org/tashi/

V V V

legend

V V V
focus on volume

V **V** V
focus on variety

V V **V**
focus on velocity