

Extraction , Load & PreProcessing

Chukwa

Suro

Tika

Morphline

Sqoop

NiFi

Flume

Scribe

HIHO

Kafka

Samza

Kestrel

Processing and Analytics MR

Kangaroo

Apache MapReduce

Pangool

Corona

Apex

Pydoop

Pachyderm

RHadoop

Cascading

Crunch

Storm

Spark

Tez

SIMR

JAQL

Processing and Analytics Abstractions

Pig

PigPen

DataFu

Twill

Giraph

SparkStreaming

Flink

Trident

Hama

TinkerPop

Mahout

H2O

Oryx

WEKA

Sparkling Water

MADlib

SQL Abstractions & NoSQL

HCatalog

Hive

Drill

Impala

Phoenix

Presto

Tajo

HAWQ

MRQL

Splout

SQL

PivotalR

Kylin

Columnar

HBase

Cassandra

Kudu

Hypertable

Kiji

Accumulo

Document Store

ArangoDB

MongoDB

RethinkDB

Key Value

Voldemort

Redis

Aerospike

OpenTSDB

Graph

TitanDB

ArangoDB

Neo4j

BI Tools

Tableau

QlikView

Cognos

Datameer

Excel

Teradata

Pentaho

SpagoBI

SAS

Talend

MSBI

Jedox

Palo

Apache Zeppelin

Applications and Frameworks

Spring XD

Akka

NodeJS

Jumbune

Talend

Cask DAP

Elastic Search

Nutch

Solr

Kibana

D3

Banana

Grafana

Storage

LustreFS

CephFS

GlusterFS

Apache HDFS

QuantcastFS

GridGain

Tachyon

Coordination, Workflow

ZooKeeper

Luigi

Oozie

Chronos

Falcon

Azkaban

Security

Sentry

Knox

Ranger

Resource Provisioning & Deployment

Bigtop

Mesos

Cloudbreak

Deploop

Buildoop

Yarn

Puppet

Chef

Salt

Myriad

HOYA

Helix

Whirr

Ambari

HUE

Infrastructure & Monitoring

Cloud Based

OpenStack

Nimbus

CloudStack

As Services

BDAAS

PAAS

DAAS

IAAS

SAAS

Providers

AWS

Azure

Google Cloud

Storage

DAS

NAS

SAN

Virtualization

ESX

XEN

KVM

HyperV

OpenVZ

Docker

Monitoring

Ganglia

Nagios

Splunk

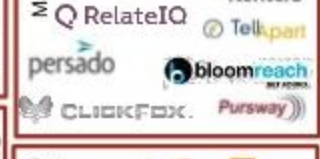
Infrastructure



Analytics



Applications



Cross Infrastructure / Analytics



Data Sources

