Big Data Landscape 2016

ALGORITHMS

mahout Spark + H₂O SPARKLING WATER **pandas**

R PredictionIO Spark MLlib

TensorFlow

ANALYTICS FRONTENDS

Apache Zeppelin jupyter pentaho R Studio

+ableau birst looker

MICRO ANALYTICS SERVICES:
Subsitutes for reporting servers.

Dashboards

RazorFlow Droptiles

Charting Libraries

Rickshaw	dc.js
D3	Flot
NVD3	DataTables
InfoVis	ChartJS
Highcharts	

Data Science	Interactive Analysis	Reporting & Dashboards

User Interface

POLYGLOTT PERSISTENCE & ANALYTICS

NewSQL: SQL meets NoSQL.

HIVE APACHE DRILL Vitess TAJO

CRATE.IO Phoenix presto

Spark SQL Kylin SNAPPYDATA Impala

Index Machines: Fast aggregation and search.

Solr elasticsearch.

In-Memory Databases: Fast access.

redis AEROSPIKE

APACHE GEODE Apache Arrow Ignite

Time Series Databases

CHRONIX InfluxDB OpenTSDB Heroic

druid KairosDB Atlas Prometheus

Graph Databases

neo4j ArangoDB OrientDB Spark GraphX

STRUCTURED DATA LAKE:
The eternal memory.

Distributed Filesystem

HDFS TACHYON

NoSQL DB

KUDU APACHE HBASE cassandra

Efficient data serialization formats:

- Integated compression
- Column-oriented storage
- Predicate pushdown

protobuf Protocol Buffers AVRO Parquet

Thrift

Analytics

Data Warehouse

Processing

Data Lake

Ingestion

Data Sources

Data Workflows

Jenkins

OOZIE nifi

Azkaban Luigi

ETL Jobs

cascading Pig

Spark talend* open data solutions

Massive Parallelization

Spark

hadoop

Data Logistics

kafka FLUME

linkedin / databus

Stream Processing

Flink STORM

Spark Streaming samza

QAWARE
SOFTWARE ENGINEERING