Big Data Landscape 2018

For more big data know-how see:
qaware.de/news/big-data-landscape

ALGORITHMS

mahout, Spark + H2O SPARKLING WATER, pandas, R, PredictionIO, Spark MLib, TensorFlow

ANALYTICS FRONTENDS

Apache Zeppelin, jupyter, pentaho, R Studio, tableau, birst, looker

MICRO ANALYTICS SERVICES:
Subsitutes for reporting servers.

Dashboards

RazorFlow, Droptiles

Charting Libraries

Rickshaw, D3, NVD3, InfoVis, Highcharts, dc.js, Flot, DataTables, ChartJS

Data Science | Interactive Analysis | Reporting & Dashboards

User Interface

Analytics

Data Warehouse

Processing

Data Lake

Ingestion

Data Sources

POLYGLOTT PERSISTENCE & ANALYTICS

NewSQL: SQL meets NoSQL.

HIVE, APACHE DRILL, Vitess, TAJO, citusdata, CRATE.IO, Phoenix, presto, Postgres-XL, Spark SQL, Kylin, SNAPPYDATA, Impala

Index Machines: Fast aggregation and search.

Solr, elasticsearch.

In-Memory Databases: Fast access.

redis, AEROSPIKE, APACHE GEODE, Apache Arrow, Apache 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, ALLUXIO

NoSQL DB

KUDU, APACHE HBASE, cassandra

Efficient data serialization formats:

- Integated compression
- Column-oriented storage
- Predicate pushdown

protobuf Protocol Buffers, AVRO, Parquet, Thrift

Data Workflows

Jenkins, beam, 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

HERON, concord, Flink, STORM, Spark Streaming, samza

QAware GmbH
+49 (0) 89 23 23 15 – 0
info@qaware.de
qaware.de

QAWARE SOFTWARE ENGINEERING