**Orchestration & Workflow** Oozie, ODE, Airavata and OODT (Tools)   
 NA: Pegasus, Kepler, Swift, Taverna, Trident, ActiveBPEL, BioKepler, Galaxy

**Data Analytics Libraries:**

**Machine Learning   
Mahout , MLlib , MLbase  
CompLearn (NA)**

**Linear Algebra   
Scalapack, PetSc (NA)**

**Statistics, Bioinformatics  
R, Bioconductor (NA)**

**Imagery ImageJ (NA)**

**In memory distributed databases/caches: GORA** (general object from NoSQL), **Memcached (NA),** **Redis**(NA) (key value), **Hazelcast (NA)**, **Ehcache** (NA);

**MRQL  
(SQL on Hadoop, Hama, Spark)**

**Hive   
(SQL on Hadoop)**

**Pig (Procedural Language)**

**Shark  
(SQL on Spark, NA)**

**Hcatalog  
Interfaces**

**Impala (NA)**

**Cloudera**

**(SQL on Hbase)**

**Swazall  
(Log Files  
Google NA)**

**High Level (Integrated) Systems for Data Processing**

**Parallel horizontally scalable Data Processing**

**Giraph  
~Pregel**

**Tez  
(DAG)**

**Spark  
(Iterative  
MR)**

**Storm**

**S4  
Yahoo**

**Samza  
LinkedIn**

**Hama  
(BSP**)

**Hadoop  
(Map  
Reduce)**

**Pegasus  
on Hadoop  
(NA)**

**NA:Twister  
Stratosphere  
Iterative MR**

**Graph**

**Batch**

**Stream**

**Pub/Sub Messaging Netty(NA)/ZeroMQ(NA)/ActiveMQ/QPid/Kafka**

**ABDS Inter-process Communication**

**Hadoop, Spark Communications MPI(NA)  
& Reductions Harp Collectives(NA)**

**HPC Inter-process Communication**

**Mesos, Yarn, Helix, Llama(Cloudera) Condor, Moab, Slurm, Torque(NA) ……..**

**ABDS Cluster Resource Management**

**HPC Cluster Resource Management**

**ABDS File Systems User Level HPC File Systems (NA)**

**HDFS, Swift, Ceph FUSE(NA) Gluster, Lustre, GPFS, GFFS   
 Object Stores POSIX Interface Distributed, Parallel, Federated**

**iRODS(NA)**

**Interoperability Layer Whirr / JClouds OCCI CDMI (NA)  
DevOps/Cloud Deployment Puppet/Chef/Boto/CloudMesh(NA)**

**Cross Cutting  
Capabilities**

**ZooKeeper, JGroups**

**Security & Privacy**

Monitoring Ambari Ganglia, Nagios, Inca (NA)

**Distributed Coordination**

Message **Protocols**

**Thrift, Protobuf (NA)**

**Apache Big Data Stack (ABDS) with HPC Integration/Enhancement**

**SQL**

**MySQL  
 (NA)**

**SciDB  
(NA)  
Arrays, R,Python**

**Phoenix  
(SQL on HBase)**

**UIMA  
(Entities)  
(Watson)**

**Tika  
(Content)**

**Extraction Tools**

**Cassandra  
(DHT)**

**NoSQL: Column**

**HBase  
(Data on  
 HDFS)**

**Accumulo  
(Data on   
HDFS)**

**Solandra  
(Solr+  
Cassandra)  
+Document**

**Azure  
Table**

**NoSQL: Document**

**MongoDB  
 (NA)**

**CouchDB**

**Lucene  
Solr**

**Riak  
~Dynamo**

**NoSQL: Key Value (all NA)**

**Dynamo  
Amazon**

**Voldemort  
~Dynamo**

**Berkeley  
DB**

**Neo4J  
 Java Gnu   
(NA)**

**NoSQL: General Graph**

**RYA RDF on  
Accumulo**

**NoSQL: TripleStore RDF SparkQL**

**AllegroGraph  
Commercial**

**Sesame (NA)**

**Yarcdata  
 Commercial  
 (NA)**

**Jena**

**ORM Object Relational Mapping:** **Hibernate**(NA), **OpenJPA** and  **JDBC** **Standard**

**File  
Management**

**IaaS System Manager Open Source Commercial Clouds  
OpenStack, OpenNebula, Eucalyptus, CloudStack, vCloud, Amazon, Azure, Google**

**Bare  
Metal**

**Green layers are Apache/Commercial Cloud (light) to HPC (darker) integration layers**

**NA – Non Apache   
projects  
  
Qiu/Jha/Fox/  
Kamburugamuva  
March 9 2014**

**Data Transport BitTorrent, HTTP, FTP, SSH Globus Online (GridFTP)**



