**Message Protocols**

**Thrift, Protobuf**

Security & Privacy

**Cross Cutting Capabilities**

**Distributed Coordination ZooKeeper,**

**Groups**

Monitoring Ambari, Ganglia, Nagios, Inca

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

**Data Analytics Libraries**

**Machine Learning: Mahout, MLlib, MLbase, CompLear**

**Linear Algebra: Scalapack, PetSc**

**Statistics: Bioinformatics R, Bioconductor**

**Imagery: ImageJ**

**Kaleidoscope of Apache Big Data Stack (ABDS) and HPC Technologies**

**LEGEND**

**Indicates**

**Non-Apache   
projects**

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

**ABDS Cluster Resource Management**

**Mesos, Yarn, Helix, Llama Condor, Moab, Slurm, Torque**

**ABDS File Systems User Level HPC File**

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

**Interoperability Layer: Whirr / JClouds OCCI CDMI  
DevOps/Cloud Deployment Puppet/Chef/Boto/CloudMesh**

**IaaS System Manager (Open Source): Commercial Clouds:**

**Bare  
Metal**

**Data Transport: BitTorrent, HTTP, FTP, SSH Globus Online**

**HPC Cluster Resource Management**

**vCloud**

**CloudStack**

**Eucalyptus**

**OpenNebula**

**OpenStack**

**Amazon**

**Azure**

**Google**

**MRQL**

**Hive**

**Pig**

**Shark**

**Impala Cloudera**

**Swazall**

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

**Giraph~Pregel**

**Tez**

**Spark**

**Storm**

**S4, Yahoo**

**Samza, LinkedIn**

**Hama**

**Hadoop**

**Pegasus on Hadoop**

**Twister Iterative MR**

**Graph**

**Batch**

**Stream**

**Parallel Horizontally Scalable Data Processing**

**ABDS Inter-process Communication**

**HPC Inter-process Communication**

**Pub/Sub Messaging Netty/ZeroMQ/ActiveMQ/QPid/Kafka**

**Hadoop, Spark Communications & Reductions MPI Harp Collectives**

**Hcatalog Interfaces**

**Stratosphere**

**In memory distributed databases/caches: GORA, Memcached, Redis,** **Hazelcast**, **Ehcache**

**MySQL**

**SciDB, Arrays, R, Python**

**Phoenix**

**Cassandra**

**HBase**

**Accumulo**

**Solandra, +Document**

**NoSQL: Document**

**MongoDB**

**CouchDB**

**Lucene Solr**

**Riak**

**Amazon**

**Voldemort**

**BerkeleyDB**

**Azure Table**

**Neo4J, Java Gnu**

**NoSQL: General Graph**

**RYA RDF on Accumulo**

**AllegroGraph Commercial**

**Sesame**

**Yarcdata Commercial**

**Jena**

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

**Extraction Tools**

**SQL**

**UIMA**

**Tika**

**NoSQL: Column**

**NoSQL: Key Value**

**NoSQL: TripleStore RDF SparkQL**

**File Management**

**iRODS**