|  |  |
| --- | --- |
| Kaleidoscope of (Apache) Big Data Stack (ABDS) and HPC Technologies January 14 2015 | |
| |  | | --- | | Cross-Cutting Functions | | 1) Message and Data Protocols: Avro, Thrift, Protobuf | | 2) Distributed Coordination: Zookeeper, Giraffe, JGroups | | 3) Security & Privacy: InCommon, OpenStack Keystone, LDAP, Sentry, Sqrrl | | 4) Monitoring: Ambari, Ganglia, Nagios, Inca  **21 layers**  **289 Software Packages** | | |  | | --- | | **17) Workflow-Orchestration:** Oozie, ODE, ActiveBPEL, Airavata, Pegasus, Kepler, Swift, Taverna, Triana, Trident, BioKepler, Galaxy, IPython, Dryad, Naiad, Tez, Google FlumeJava, Crunch, Cascading, Scalding, e-Science Central, Azure Data Factory, NiFi (NSA) | | **16) Application and Analytics:** Mahout , MLlib , MLbase, DataFu, mlpy, scikit-learn, CompLearn, Caffe, R, Bioconductor, ImageJ, pbdR, Scalapack, PetSc, Azure Machine Learning, Google Prediction API, Google Translation API, Torch, Theano, H2O, Google Fusion Tables, Oracle PGX, GraphLab, GraphX, CINET, NWB, Elasticsearch, IBM System G, IBM Watson, GraphBuilder(Intel), TinkerPop | | **15A) High level Programming:** Kite, Hive, HCatalog, Databee, Tajo, Pig, Phoenix, Shark, MRQL, Impala, Presto, Sawzall, Drill, Google BigQuery (Dremel), Google Cloud DataFlow, Summingbird, SAP HANA, IBM META, HadoopDB, PolyBase  **15B) Frameworks:** Google App Engine, AppScale, Red Hat OpenShift, Heroku, AWS Elastic Beanstalk, IBM BlueMix, Ninefold, Aerobatic, Azure, Jelastic, Cloud Foundry, CloudBees, Engine Yard, CloudControl, appfog, dotCloud, Pivotal, OSGi, HUBzero, OODT | | **14A) Basic Programming model and runtime**, **SPMD, MapReduce:** Hadoop, Spark, Twister, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus **14B) Streams:** Storm, S4, Samza, Google MillWheel, Amazon Kinesis, LinkedIn Databus, Facebook Scribe/ODS, Azure Stream Analytics | | **13) Inter process communication Collectives, point-to-point, publish-subscribe:** Harp, MPI, Netty, ZeroMQ, ActiveMQ, RabbitMQ, QPid, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, Azure Event Hubs, Amazon Lambda  **Public Cloud:** Amazon SNS, Google Pub Sub, Azure Queues | | **12) In-memory databases/caches:** Gora (general object from NoSQL), Memcached, Redis (key value), Hazelcast, Ehcache, Infinispan | | **12) Object-relational mapping:** Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC | | **12) Extraction Tools:** UIMA, Tika | | **11C) SQL(NewSQL):** Oracle, DB2, SQL Server, SQLite, MySQL, PostgreSQL, SciDB, Apache Derby, Google Cloud SQL, Azure SQL, Amazon RDS, rasdaman, BlinkDB, N1QL, Galera Cluster, Google F1, Amazon Redshift, IBM dashDB | | **11B) NoSQL:** HBase, Accumulo, Cassandra, Solandra, MongoDB, CouchDB, Lucene, Solr, Berkeley DB, Riak, Voldemort, Neo4J, Yarcdata, Jena, Sesame, AllegroGraph, RYA, Espresso, Sqrrl, Facebook Tao, Google Megastore, Google Spanner, Titan:db  **Public Cloud:** Azure Table, Amazon Dynamo, Google DataStore | | **11A) File management:** iRODS, NetCDF, CDF, HDF, OPeNDAP, FITS, RCFile, ORC, Parquet | | **10) Data Transport:** BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop | | **9) Cluster Resource Management**: Mesos, Yarn, Helix, Llama, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Google Omega, Facebook Corona | | **8) File systems:** HDFS, Swift, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS, Haystack, f4  **Public Cloud:** Amazon S3, Azure Blob, Google Cloud Storage | | **7) Interoperability:** Whirr, JClouds, OCCI, CDMI, Libcloud, TOSCA, Libvirt | | **6) DevOps:** Docker, Puppet, Chef, Ansible, Boto, Cobbler, Xcat, Razor, CloudMesh, Heat, Juju, Foreman, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic | | **5) IaaS Management from HPC to hypervisors:** Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, VMware ESXi, vSphere, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, VMware vCloud, Amazon, Azure, Google and other public Clouds,  **Networking:** Google Cloud DNS, Amazon Route 53 | |

