|  |  |
| --- | --- |
| Figure 2: Kaleidoscope of (Apache) Big Data Stack (ABDS) and HPC Technologies | |
| |  | | --- | | **Cross-Cutting Functions** | | 1) **Message and Data Protocols:** Avro, Thrift, Protobuf | | 2) **Distributed Coordination:** Google Chubby,  Zookeeper, Giraffe, JGroups | | 3) **Security & Privacy:** InCommon, Eduroam  OpenStack Keystone, LDAP, Sentry, Sqrrl, OpenID, SAML OAuth | | 4) **Monitoring:** Ambari, Ganglia, Nagios, Inca |   **21 layers**  **Over 300 Software Packages**  **9 March**  **2015** | |  | | --- | | **17) Workflow-Orchestration:** ODE, ActiveBPEL, Airavata, Pegasus, Kepler, Swift, Taverna, Triana, Trident, BioKepler, Galaxy, IPython, Dryad, Naiad, Oozie, Tez, Google FlumeJava, Crunch, Cascading, Scalding, e-Science Central, Azure Data Factory, Google Cloud Dataflow, NiFi (NSA) | | **16) Application and Analytics:** Mahout , MLlib , MLbase, DataFu, R, pbdR, Bioconductor, ImageJ, Scalapack, PetSc, Azure Machine Learning, Google Prediction API, Google Translation API, mlpy, scikit-learn, PyBrain, CompLearn, Caffe, Torch, Theano, H2O, IBM Watson, Oracle PGX, GraphLab, GraphX, IBM System G, GraphBuilder(Intel), TinkerPop, Google Fusion Tables, CINET, NWB, Elasticsearch | | **15B) Application Hosting Frameworks:** Google App Engine, AppScale, Red Hat OpenShift, Heroku, Aerobatic, AWS Elastic Beanstalk, Azure, Cloud Foundry, Pivotal, IBM BlueMix, Ninefold, Jelastic, Stackato, appfog, CloudBees, Engine Yard, CloudControl, dotCloud, Dokku, OSGi, HUBzero, OODT, Agave, Atmosphere  **15A) High level Programming:** Kite, Hive, HCatalog, Tajo, Shark, Phoenix, Impala, MRQL, SAP HANA, HadoopDB, PolyBase, Presto, Google Dremel, Google BigQuery, Amazon Redshift, Drill, Pig, Sawzall, Google Cloud DataFlow, Summingbird | | **14B) Streams:** Storm, S4, Samza, Granules, Google MillWheel, Amazon Kinesis, LinkedIn Databus, Facebook Puma/Ptail/Scribe/ODS, Azure Stream Analytics  **14A) Basic Programming model and runtime**, **SPMD, MapReduce:** Hadoop, Spark, Twister, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi | | **13) Inter process communication Collectives, point-to-point, publish-subscribe:** MPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, **Public Cloud:** Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs | | **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, Galera Cluster, SciDB, Rasdaman, Apache Derby, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, N1QL, BlinkDB | | **11B) NoSQL:** Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, HBase, Google Bigtable, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Facebook Tao, Titan:db, Jena, Sesame  **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, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs | | **8) File systems:** HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS  **Public Cloud:** Amazon S3, Azure Blob, Google Cloud Storage | | **7) Interoperability:** Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis | | **6) DevOps:** Docker, Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive | | **5) IaaS Management from HPC to hypervisors:** Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds,  **Networking:** Google Cloud DNS, Amazon Route 53 | |











