Implications of the HPC-ABDS High Performance Computing Enhanced Apache Big Data Stack for workflows

Geoffrey Fox^a, Judy Qiu^a, Shantenu Jha^b, Supun Kamburugamuve^a and Andre Luckow^b ^a School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA ^b RADICAL, Rutgers University, Piscataway, NJ 08854, USA

1. Summary

We have introduced the software stack HPC-ABDS (High Performance Computing enhanced Apache Big Data Stack) to motivate an approach to high performance data analytics. The initial richness of the ABDS includes (cloud) workflow as well as Platform as a Service and system and heterogeneity management. We see these capabilities as relevant for the HPC workflow community but not yet fully exploited even though the distinction between HPPC workflows and data-intensive applications/analysis decrease. This paper argues the case for inclusion of HPC-ABDS approaches and components in any future system architecture, including integration of DevOps for systems management in the next-generation HPC workflow systems.

2. Introduction to HPC-ABDS

In previous work [1-4], we introduced the software stack HPC-ABDS shown online [5] and in the Figure with over 350 entries. These were combined with an application analysis [6-8] and used to motivate an approach to high performance data analytics including identification of a benchmarking set [9, 10]. We divided the stack into 21 architecture layers covering Message and Data Protocols, Distributed Coordination, Security & Privacy, Monitoring, Infrastructure Management, DevOps, Interoperability, File Systems, Cluster & Resource management, Data Transport, File management, NoSQL, SQL (NewSQL), Extraction Tools, Object-relational mapping, In-memory caching and databases, Inter-process Communication, Batch Programming model and Runtime, Stream Processing, High-level Programming, Application Hosting and PaaS, Libraries and Applications, Workflow and Orchestration. Further details of the stack can be found in an online course [11] that includes a section with approximately one slide (and associated lecture video) for each entry in Figure. The software in the figure comes from a variety of sources but we highlight the Apache contribution as it not only has contributed many key packages, it has established a very effective approach to software that has resonated with the cloud and big data commercial revolutions.

We suggest that HPC at all levels should carefully examine this software stack and adopt ABDS systems where relevant; this approach is likely to lead to a more sustainable software ecosystem where HPC can spend its scarce (relative to commercial world) resources on those areas where HPC has special expertise or requirements. We do not necessarily argue that ABDS is better than current HPC solutions but rather that adopting ABDS will in the long run lead to software environments that are cheaper and easier to maintain and have broader scope. We now briefly discuss some of the areas where HPC and ABDS have important overlaps and opportunities for integration.

Lower layers where HPC can make a major impact include scheduling where Apache technologies like Yarn and Mesos need to be compared with the sophisticated HPC approaches such as Slurm and Pilot jobs. Storage is another important area where HPC distributed and parallel storage environments need to be reconciled with the "data parallel" storage seen in HDFS in many ABDS systems. Further important issues are at the higher layers with data management, communication, (high layer or basic) programming, analytics and orchestration. These are areas where there is rapid commodity/commercial innovation and we briefly discuss them in order below. Much science data analysis is centered on files but we expect movement to the common commodity approaches of Object stores, SQL and NoSQL where latter has a proliferation of systems with different characteristics – especially in the data abstraction that varies over row/column, key-value, graph and documents. Note recent developments at the programming layer including Apache Hive and Drill, which offer high-layer access models such as SQL implemented on scalable NoSQL data systems. Maybe Drill can be generalized to offer traditional science interfaces such as FITS and HDF on ABDS data stores. The communication layer includes Publish-subscribe technology used in many approaches to streaming data as well the HPC communication technologies (MPI) which are much faster than most default Apache approaches but can be added [12] to some systems like Hadoop whose modern

version is modular and allows plug-ins for HPC stalwarts like MPI and sophisticated load balancers. The programming layer includes both the classic batch processing typified by Hadoop or Spark and streaming by Storm. The latter seems very relevant to the processing of the streaming observational data such as that from light sources and telescopes. The programming offerings differ in approaches to data model (key-value, array, pixel, bags, and graph), fault tolerance and communication. The trade-offs here have major performance implications.

 Cross- Cutting [1] Worklow Orchestration: ODE, Active MPEL, Airvand, Pegasus, Kepler, Swift, Tavera, Triana, Trident, BioKepler, Galaxy, Python, Dryad, Naiad, Oozie, Tez, Google Flumelava, Crunch, Cascading, Scalding, e-Science Central, Aure Data Factory, Google Cloud Dataflow, NiFi (NSA), Jitterbit, Talend, Pentabo, Apatr, Docker Compose [1] Amestage [1	Kala' have a f (A sould) B's Data f(a al. (ADDf)) and HDC Task salas's		
Cutting Functions Trident, BioKepler, Galaxy, IPython, Dryad, Naida, Oozie, Tez, Google FlumeUvaz, Crunch, Cascading, Scalding, e-Science Central, Aurre Data Factory, Google Cloud Dataflow, NFi (NSA), Jitterbit, Talend, Protocols 10 Message and Data Protocols To Application and Analytics: Mahout, MLlib MLbase, Dataf'u, R. pbdR, Bioconductor, ImageJ, Open CV, Scalapack, Pets A. Aure Machine Learning, Google Prediction API & Translation API, mlpy, scikit-learn. PyBrain, CompLearn, DAAL (Intel), Caffe, Torch, Theano, DL4; IE20, IBW Watson, Oracle POX, GraphLab, GrapkX, TBM System G, GraphBuilder(Hutt), TinkerPor, Google Fusion Tables, CINET, NBB, Elasticsearch, Kibana, Logstash, Graylog, Splunk, Tableau, D3js, Itreeis, Pottos, CA, Google Chubby, Stackato, approg. CloudBese, Engine Yard, CloudCountrol, dotCloud, Dokku, OSGi, HUBZero, OODT, Agave, Atmosphere 15A) High Hevel Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoenis, Impala, MRQL, SAP HANA, HadoopDB, PolyBase, Pivotal HD/Hawq, Presto, Google Demel, Google BigQuery, Amazon Redshift, Briangy 13S) Security & Privacy: Informan Table Streams: Storm, S4, Sarnza, Granules, Google Damel, Google BigQuery, Amazon Redshift, Briangy: Toren 14A) Basic Programming model and runitme, SPMD MagReduce: Hadoop, Spark, Twister, MR-MPI, Stratoophere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem 20 Nonitoring: Monitoring: Matai, Ganglia, Nagios, Inc. 10 Inter process communication Collectives, point-to-point, publish-subscribe: WPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbiMQ, NaradBarokering, OPIA, Karfa, Kestrel, JMS, AMQP, Storm, MOTT, Maronethe Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azare Oueues, Event Hubs SAML OAuth <		Kaleidoscope of (Apache) Big Data Stack (ABDS) and HPC Technologies	
Functions Scalading, e-Science Central, Azure Data Factory, Google Cloud Dataflow, NiFi (NSA), Jitterbit, Talend, Pentabo, Apatar, Docker Compose 1) Message and Data Protocols: 16) Application and Analytics: Mahout, MLlib , MLbase, DataFu, R, pbdR, Bioconductor, ImageJ, OpenCV, Scalapack, PetSc, Azure Machine Learning, Google Prediction API & Translation API, mipy, scikit-learn, PBrain, Complacm, DAAL(Intel), Caff, Croch, Theano, DL4, H2O, IBM Wason, Oracle POK, GraphLab, GraphX, IBM System G, GraphBuilder(Intel), TinkerPop, Google Proion Tables, CINET, NWB, Blastisecarch, Kihan, Logstah, Graylo, Splunk, Tableau, D3; Kitreeja, Potree, DC.js 2) Distributed Coordination. 150 Application Hosting Frameworks: Google App Engine, AppScale, Red Hat OpenShift, Heroku, Aerobatic, AWS Elastis Beanstalk, Azure, Cloud Poundry, Pivotal, IBM BlueMix, Ninefold, Lelastic, Stackato, appfog, CloudBees, Engine Yard, CloudCourol, dotCloud, Dokku, OSGi, HUBzero, OODT, Agave, Atmosphere 15A) High level Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoenix, Impala, MRQL, SAP HANA, HadoopDB, PolyBase, Pivotal HD/Hawq, Presto, Google MilWheel, Amazon Kinesis, LinkedIn Databus, Frivacy: InCommon, Eduroam 148) Streams: Storm, S4, Samza, Granules, Google MilWheel, Amazon Kinesis, LinkedIn Databus, Frivacy: 10, Diere process communication Collectives, point-to-point, publish-subscribe: MPL, Harp, Netty, ZeroMQ, ActiveAQ, RabbiMQ, NaradaBrokering, QPM, Kafka, Kestel, JMS, AMQP, Storm, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azare Queues, Event Hubs StatD Acath 11) Inter process communication Collectives, point-to-point, publish-subscribe: MPL, Harp, Netty, Yard, ActiveAQ, RabbiMQ, NaradaBrokering, QPM, Kafka, Kestrel, JMS, AMQP, Storm,			
Pentaho, Apatar, Docker Compose 10) Apsigned 10) Application and Analytics: Mahout, MLib, MLase, DataFu, R., pbdR, Bioconductor, ImagJ, OpenCV, Scalapack, PetSc, Azure Machine Learning, Google Prediction API & Translation API, mlpy, scikit-learn. PyBrain, CompLearn, DAAL(Intel), Caffe, Torch, Theano, DL4; IL2O, IBW Watson, Oracle PCX, Scalapack, PetSc, Azure Machine Learning, Google Prediction, API & Translation API, mlpy, scikit-learn. PyBrain, CompLearn, DAAL(Intel), Caffe, Torch, Theano, DL4; IL2O, IBW Watson, Oracle PCX, GraphEuider(Hutc), TinkerPo, Google Pusion Tables, CINET, NWB, Elasticsearch, Kihana, Logstash, Graylog, Splunk, Tableau, D3,is, Itrree is, Porte, DC.;s 153) Application Hosting Frameworks: Google Applicine, AppScale, Red Hat OpenShirk, Heroku, Acrobatic, AWS Elastic Beanstalk, Azure, Cloud Foundry, Pivotal, IBM BlueMix, MIGL, SAP HANA, HadoopDB, PolyBase, Pivotal ID/Hawa, Presto, Google Dremel, Google BigQuery, Amazon Redshirt, Drini, Kyoto Cabinet, Pig, Sawzall, Google Cloud DataFlow, Summingbird 30 Security & Jayawas, Storm, S4, Samza, Granules, Google MIWheel, Amazon Kinesis, LinkedIn Databus, Facebook Puma/Itail/Scribe/DDS, Azure Stream Analytics, Floe InCommon, Eduroam 140) Base Programming model and runime. SPMD, MapReduce: Hadoop, Spark, Twister, MR-MPI, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem Monitoring: 16) Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbiMQ, NaradaBrokering, QPI, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, Marionet Collective, Public Cloud: Anazaon SNS, Lambda, Google PUB, Sub, Azare Queese, Svent Hubs SAML LOAth 10) Inter proces	Cutting		
 Message and Data Malyties: Mahour, MLilb, ML hase, DataFu, R., ploff, Birconductor, Imagel, OpenCV, Scalapack, Petics, Azure Machine Learning, Google Prediction API & Translation API, https://www.plot.ib.abcaraphy. 1004 Message and Pata Market Anter Machine Learning, Google Prediction API & Translation API, https://www.plot.ib.abcaraphy. 1004 Distributed Coordination: Www.B. Elasticsearch, Kihana, Logatash, Graylog, Spluth, Tableau, D3; Nitreejs, Potree, DC.;s Distributed Coordination: Statesarch, Kihana, Logatash, Crubel Southy, Pivotal, IBM Bulewhis, Ninefold, Jelastic, Stackato, appfog, CloudBees, Engine Yard, CloudControl, dotCloud, Dokku, OSGi, HUBzero, OODT, Agave, Atmosphere May Distributed Coordination: Statesarch, Xihana, Cloud Foundry, Pivotal, IBM Bulewhis, Ninefold, Jelastic, Stackato, appfog, CloudBees, Engine Yard, CloudControl, dotCloud, Dokku, OSGi, HUBzero, OODT, Agave, Atmosphere Statosphere, Pivota Silb High Ievel Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoenix, Impala, MRQL, SAP HANA, HadoopDB, PolyBase, Pivotal HD/Haw, Presto, Google Dimel, Google BigQuery, Amazon Redshift, Brill, Kyoto Cabinet, Pig, Savzall, Google Cloud DataTow, Summingbires, LinkedIn Databus, Frivacy: InCommon, Editors, Phys. Sawzall, Google Cloud DataTow, Summingbires, LinkedIn Databus, Frivacy: InCommon, Editors, Phys. Cloud And runtime, SPMD, MapReduce: Hadoop, Span, Twister, MR-MPI, Stratosphere (Apache Flink, Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapdCaph, Tatem Matornet Collective, Public Cloud: Amazon SNS, Lambda, Google Publich, Galois, Medusa-GPU, MapGraph, Tatem Matornet Galabases/acches: Goot Legneral Object From NoSQL). Mencached, Redis, LMDB (key value), Hazelcast, Ehache, Infinispan Monitorrig: 10 Diget-relational mapping: Hibernate, OpenIPA, EclipseLink, DataNucleus, ODBC/JDBC Bigabble, LevelDB, Megastaches: Goot Ligneral My SQL, P	Functions		
and Data Protocols 10, Application and Analytics, Sizuabat, Malabat, Malabat, Sizuabat, Google Prediction Appl. Ref. Translation APJ, mbpg. scikit-learn, PyBrain, CompLearn, DAAL (Intel), Caffe, Torch, Theano, D.J., H2O, BIW Watson, Oracle PGX, Graphalbab, GraphX, MBM System G, GraphBuilder(Intel), TinkerPop, Google Fusion Tables, CINET, WB, Elasticsearch, Kibana, Logstabh, GraphV, Caffe, Torch, Theano, D.J., H2O, BIW Watson, Oracle PGX, Graphalbab, GraphX, Kana, Logstabh, Graylog, Splunk, Tableau, D3js, three js, Porte, DC.js 2) Distributed 15B) Application Hosting Praneworks: Google App Engine, AppScale, Red Hat OpenShith, Heroku, Acrobatic, AWS Elastic Beanstalk, Azure, Cloud Foundry, Pivotal, IBM BlueMix, Ninefold, Jelastic, Stackao, apriog, CloudBees, Engine Yard, CloudControl, dorCloud, Dokku, OSG, HUBZero, OODT, Agave, Atmosphere 3/Socurity & MadoopDB, PolyBase, Pivotal HD/Hawa, Presto, Google Dirmel, Google BigQuery, Amazon Redshift, Drill, Kyoto Cabinet, Pig, Sawzall, Google Cloud DataFlow, Summingbird 3/Socurity & MadoopDB, PolyBase, Pivotal (Google Cloud DataFlow, Summingbird 3/Socurity & Madoraph, Process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroNQ, ActiveMQ, RabbifMQ, NaradaBrokering, QPid, Kafka, Kestrel, JNS, AMQP, Stomp, MQTT, Marionet Collective, Public Cloud: Anaradon SNS, Landbad, Google Publ Sub, Azare Queeue, Sevent Hubs SAML Ovath 4) Immerory database/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Hazeleast, Ehcache, Infinispan 10. Olget-relational mapping: Hibernate, OpenIPA, EclipseLink, DataNucleus, ODBC/JDBC 11. Interp process communication Collective, Public Cloud: Asamz Tabbase, Software	1) Message		
Protools: Open CV, Scalagek, PreSC, AZule Machine Learning, Oodge Prediction Ari & Trainslation Ari, Impy, scikit-learn, Pybrain, Complearn, DAAL(Intel), Cafry Torch, Theano, DLA, H2O, IBM Watson, Oracle PGX, GraphLab, GraphX, IBM System G, GraphBuilder(Intel), TinkerPop, Google Fusion Tables, CINET, NWB, Elasticsearch, Kibana, Logatsab, Grayloz, Splank, Tableau, D3; st, threejs, Potree, DC. js 2) Distributed Coordination Coordination GraphX, IBM System G, Grayloz, Splank, Tableau, D3; st, threejs, Potree, DC. js 2) Distributed Coordination Coordination Acrobatic, AVS Elastic Beanstalk, Azure, Cloud Foundry, Pivotal, IBM BlueMix, Ninefold, Jelastic, Stackato, appfog, CloudBees, Engine Yard, CloudContol, dotCloud, Dokku, OSGi, HUBzero, OODT, Agave, Atmosphere 2: A Big Hevel Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoenix, Impala, MRQL, SAP HANA, HadoopDB, PolyBase, Pivotal HD/Hawq, Presto, Google Dremel, Google BigQuery, Amazon Redshift, Drill, Kyoto Cabiner, Pig, Sawazall, Google Cloud DataFlow, Sammingbid 3) Security & Prayse HB) Streams: Storm, S4, Samza, Granules, Google MilWheel, Amazon Kinesis, LinkedIn Databus, Facebook Puma/Piail/Scribe/ODS, Azure Stream, Analytics, Floe HA/ Basic Programming model and runtime. SPND, MapReduce: Hadoop, Spark, Twister, MR-MPI, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem BAMbari, Ganglia, Andari, Totem 10. Inter process communication Collectives, point-to-point, publish-subserbic MPI, Harp, Netty, ZeroMO, ActiveMQ, RabioMo, NaradaBrokering, Opti, KafaX, Kestrel, JMS, AMOP, Shomp, MQTT, Marionette Collective. Public Cloud: Amazon SNS, Lambda			
Avro, Thrift, Schellenth, Pybrain, Completin, DAALUmer, Catte, 10rch, Intendo, D.G., Monto, Databes, CINET, Protobuf PGX, Graphala, GraphX, IBM System G, GraphBuilder(Intel), TinkerPop, Google Fusion Tables, CINET, NWB, Elasticsearch, Kibana, Logstash, Graylog, Splunk, Tableau, D3 js, threejs, Portee, DC, js Schenker, Kibana, Logstash, Graylog, Splunk, Tableau, D3 js, threejs, Portee, DC, js 2) Distributed Schenker, Sch			
 Protohuf Protohuf Pistributed Obstributed Coordination Google App Engine, AppScale, Red Hat OpenShift, Heroku, Aerobatic, AWS Elastic Beanstalk, Azure, Cloud Foundy, Pivotal IBM BlueMix, Ninefold, Jelastic, Stackato, appfog. CloudBees, Engine Yard, CloudControl, dotCloud, Dokku, OSGi, HUBzero, OODT, Agave, Atmosphere ISA) High level Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoenix, Impala, MRQL, SAP HANA, HadoopDB, PolyBase, Pivotal HD/Hawq, Presto, Google Dremel, Google Biguery, Amazon Redshift, Drill, Kyoto Cabinet, Pig. Sawzall, Google Cloud DataFlow, Saurmingbird Stecurity & Pivaay: IROmmon, Educomon, Educomon, Educomon, Strom, SA, Sanza, Granules, Google MillWheel, Amazon Kinesis, LinkedIn Databus, Facebook PumaPitai/Steo/DDS, Azare Stream Analytics, Floe HA) Basic Programming model and runtime, SPMD, MapReduce: Hadoop, Spark, Twister, MR-MPI, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem Ilotre process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbiMQ, NaradBrokering, Qpki, Katka, Kestel, JMS, AMQP, Stomp, MQTT, Magios, Inca Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC Inmemory database/aches: Gora (general object from NoSQL), Memached, Redis, LMDB (key value), Haze(Last, Elacache, Infinispan Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC Extraction Tools: UUMA, Tika Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC Bill NayOL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, KyotoTokyo Cabinet, Tycooon, Tyrant, Mon			
 2) Distributed (NPS, ElasticSearch, Khaala, Logstash, Uraylog, Splutk, Taoleau, D.3,S. Inteel, S. Pottee, J.C.J.S. (SB, 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 (Sokeurje, J., High level Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoenix, Impala, MRQL, SAP HANA, HadoopDB, PolyBase, Pivotal HD/Hawq, Presto, Google Dremel, Google BigQuery, Amazon Redshift, Diroups (Securiy & Privacy: Inf. Kyoto Cabinet, Pig, Sawzall, Google Cloud DataFlow, Summingbird (JB: Unterparenting model and runtime, SPMD), MapReduce: Hadoop, Spark, Twister, MR-MPI, Statosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MagGraph, Totem (Jarter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZroMQ, ActiveMQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azare Queue, Event Hubs (Li) SQL(NewSQL): Oracle, IDB2, SQL Server, SQLite, MySQL, PostgreSQL, CUBRID, Galera Cluster, SciDR, Rasdama, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google FJ, IBM dashDB, NIQL, BlinkDB (Li) NagoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Blab, Mas, Sart, State, Sart, Strate, Speresso, CouchDB, Couchbase, BM Cloudant, Pivotal Gemfire, HBase, Google Blab, Nigue, Hanzon Kines, Pitter, HDPS, Swift, Haystack, f4, Cinder, Ceph, FICS, RCFile, ORC, Parquet (D) Diat Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (Grid/TP), Flume, Sqoop, Pivotal GPLOAD/GPFDIST (Diata Transport: BitTorrent, HTTP, FTP, SSH, Globus Colnel, G			
 Coordination: Google Coordination: Google Aerobatic, AWS Elastic Beanstalk, Azare, Cloud Foundry, Pivotal, IBM BlueMix, Ninefold, Jelastic, Stackato, appfog. CloudBees, Engine Yard, CloudControl, dotCloud, Dokku, OSGi, HUBzero, OODT, Agave, Atmosphere 15. High level Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoein:, Inpala, MRQL, SAP HANA, HadoopDB, PolyBase, Pivotal HD/Hawq, Presto, Google Dremel, Google BigQuery, Amazon Redshift, Drill, Kyoto Cabinet, Pig. Sawzall, Google Cloud DataFlow, Summingbird 18 Streams: Storm, S4, Samza, Granules, Google MillWheel, Amazon Kinesis, LinkedIn Databus, Facebook Puma/Pail/Scribe/ODS, Azure Stream Analytics, Floe 14A) Basic Programming model and runtime, SPMD, MapReduce: Hadoop, Spark, Twister, MR-MPI, Haopensite, Agave, Anache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem 15. Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty. ZeroNQ, ActiveNQ, RahdBrokering, QHd, KafaK, Kestrel, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs SAML OAuth 10. Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty. ZeroNQ, ActiveNQ, RahdBrokering, QHd, KafaK, Kestrel, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs SAML OAuth 11. Detacleast, Ehcache, Infinispan 12. Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC 12. Bartaction Tools: UIMA, Tika 11. Singu LevelDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrtl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titandb, Jeaa, Sesame Public Cloud: Azure Table, Amazon Dynamer, Accumulo, Cassandra, RYA, Sqrtl, Neo4J, Yarcdata, AllegroGraph, Blazegraph			
Google Chubby, Chubby, Chubby, Chubby, Giraffe, JisAl High level Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoenix, Impala, MRQL, SAP HANA, HadoopDB, PolyBase, Pivotal HD/Hawq, Presto, Google Diremel, Google BigQuery, Amazon Redshift, Drill, Kyoto Cabinet, Pig. Sawzall, Google Cloud DataFlow, Summingbird 13) Security & Privacy: InCommon, Eduroam OpenStack 14B Streams: Storm, S4, Samza, Granules, Google Direl, Google BigQuery, Amazon Redshift, Drill, Kyoto Cabinet, Pig. Sawzall, Google Cloud DataFlow, Summingbird 14D Streams: Storm, S4, Samza, Granules, Google Dirbhow, Summingbird 14B Streams: Storm, S4, Samza, Granules, Google MillWheel, Amazon Kinesis, LinkedIn Databus, Facebook Puma/Ptail/Scribe/ODS, Azure Stream Analytics, Floe 11Grommon, Eduroam OpenStack 14D Streams: Storm, S4, Samza, Granules, Google MillWheel, Amazon Kinesis, LinkedIn Databus, Facebook Puma/Ptail/Scribe/ODS, Azure Stream Analytics, Floe 13) Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbitMQ, NaradaBrokering, QPd, Kafka, Kestrel, JNS, AMQP, Stomp, MQTT, Marionete Collective: Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs SAML OAtut 4) 12) In-memory databases/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Hazelcast, Ehcache, Infinispan 12) Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC 12) Issue Sub, Licuce, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudan, Fivotal Gemfire, HBase, Google BitB) NoSQL: Lucere, Solr, Solandrat, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo			
 Chubby, Zookeeper, Giraffe, Jagave, Atmosphere Agave, Atmosphere Jiligh level Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoenix, Impala, MRQL, SAP HANA, HadoopDB, PolyBase, Pivotal HD/Hawq, Presto, Google Dremel, Google BigQuery, Amazon Redshift, JGroups Steurity & Privacy: InCommon, Eduroam Jenter Programming model and runtime. SPMD, MapReduce: Hadoop, Spark, Twister, MR-MPI, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem Jasic Programming model and runtime. SPMD, MapReduce: Hadoop, Spark, Twister, MR-MPI, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem Jiner process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroNQ, ActiveMQ, RabbitMQ, NaradaBrokering, QPId, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs Admioniting: Ambari, Ganglia, Nagios, Inca Dinermory databases/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Haze(cast, Ehcache, Infinispan Object-relational mapping: Hibernate, OpenIPA, EclipseLink, DataNucleus, ODBC/JDBC Extraction Tools: UIMA, Tika Chuemes, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, PiVA, Sgrt, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titamdb, Jena, Sesame Public Cloud: Azure Table, Amazon Dynamo, Google DataStore May 15 Sh File systems: HDFS, Swith, Haytack, Hd Chuden, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage Netorperability: Libvint, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis DevOps: Dockc			
Zookeper, ISA) High level Programming: Kite, Hive, HCatalog, Tajo, Shark, Phoenix, Impala, MRQL, SAP HANA, Giraffe, Jifkyto Cabinet, Pig, Savzall, Google Cloud DutaFlow, Summingbird 3) Security & Privacy: InCommon Privacy: Biromann Big Streams: Storm, S4, Samza, Granules, Google MilWheel, Amazon Kinesis, LinkedIn Databus, Privacy: Facebook Puma/Ptail/Scribe/ODS, Azure Stream Analytics, Floe InCommon, Edvise Programming model and runtime. SPMD, MapReduce: Hadoop, Spark, Twister, MR-MPI, Startosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem 13 Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, Zork, SAML, OAuth 12 In-memory databases/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key Monitoring: Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs Yule), Hazelcast, Ehcache, Infinispan 10 Object-relational mapping: Hibernate, OpenIPA, EclipseLink, DataNucleus, ODBC/JDBC 12 Isayers Oogle F1, IBM dashDB, NIQL, BlinkDB 11 NosoQL: Ducene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Genrie, HBase, Google Bigtable, LevelDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4			
Giraffe, HadoopDB, PolyBase, Pivotal HD/Hawq, Presto, Google Dremel, Google BigQuery, Amazon Redshift, JGroups Till, Kyoto Cabinet, Pig, Sawzali, Google Cloud DataFlow, Summingbird JS Security & Frivacy: InCommon, Halb Streams: Storm, S4, Samza, Granules, Google MilWheel, Amazon Kinesis, LinkedIn Databus, Frivacy: Frivacy: New Type Frivacy: Groumon, Eduroam OpenStack Keystone, Keystone, Till, Kyoto Cabinet, Pig, Sawzali, Google Other, MagReduee: Hadoop, Spark, Twister, MR-MPI, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem Till Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs SAML OAU 10 Inter process communication Coll cerver, SQLine, MySQL, PostgreSQL, CUBRID, Galera Cluster, Saiba, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cluud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, NIQL, BlinADB 11 Interperesso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Moago DB, Spresso, CouchDB, Couchbase, BM Cloudant, Pivotal Gemfi	-		
IGroups Drill, Kyoto Cabinet, Pig. Sawzall, Google Cloud DataFlow, Summingbird 3) Security & Table Streams: Storm, S4, Samza, Granules, Google MillWheel, Amazon Kinesis, LinkedIn Databus, Facebook Puma/Paul/Scribe/ODS, Azure Stream Analytics, Floe Incommon, Eduroam Hab) Streams: Storm, S4, Samza, Granules, Google MillWheel, Amazon Kinesis, LinkedIn Databus, Facebook Puma/Paul/Scribe/ODS, Azure Stream Analytics, Floe Incommon, Eduroam IAA) Basic Programming model and runtime, SPMD, MapReduce: Hadoop, Spark, Twister, MR-MPI, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem Yonno, Eduroam Yaroby Reef (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem Yaroby Charley, Coll (Cloud) Active MQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs SAML OAuth 12) In-memory databases/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Hazelcast, Ehcache, Infinispan Magins, Inca 12) Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC Z1 layers Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC Ganglia, Nagios, Inca 110, SQL (NewSQL): Oracle, DB2, SQL Server, SQL ite, MySQL, PostgreSQL, CUBRID, Galera Cluster, SciDB, Rasdaman, Pache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, N1QL, BlinkDB<			
 3) Security & Privacy: InCommon, Eduroam 3) Bstreams: Storn, 54. Samza, Granules, Google MillWheel, Amazon Kinesis, LinkedIn Databus, Facebook Puma/Ptail/Scribe/ODS, Azure Stream Analytics, Floe 14.N Basic Programming model and runtime, SPMD, MapReduce: Hadoop, Spark, Twister, MR-MPI, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem 13. Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs SAML OAuth 12. In-memory database/scaches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Hazelcast, Ehcache, Infinispan 12. Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC Ambari, Ganglia, Magios, Inca 12. Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC 11B) NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrtl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, 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: BifTorrent, HTTP, FTP, SSH, Globus Tools, Pilot Jobs Si File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage DeVOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, Salfstack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Forem			
Privacy: InCommon, Eduroam OpenStack Keystone, LDAP, Settry, Sqr1, OpenID, ActiveKQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kestrel, JMS, AMOP, Stomp, MQTT, MarGraph, Totem13 Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroMQ, ActiveKQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kestrel, JMS, AMOP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs SAML OAuth12 Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC 12) Extractional mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC 13) Ruse and Analytice, Solit, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrtl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titandb, 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: Bitforrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD GPFDIST 9) Cluster Resource Management: Mesos, Yarn, Helix, Llama, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs8) File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cl			
InCommon, Eduroarn OpenStack Keystone, LDAP, Sentry, Santi Libert, Rest, Libert, Rest, Santia, Ganglia, Nagios, Inca14A) Basic Programming model and runtime. SPMD, MapReduce: Hadoop, Spark, Twister, MR-MPI, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem13) Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kesterl, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs21) In-memory databases/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Hazelcast, Ehcache, Infinispan21) Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC21) Rayers Over 350 Software PackagesMay 15 20152015May 15 20152015May 15 201520159. Cluster Resource Management: iRODS, NetCDF, CDF, HDF, OPeNDAP, FITS, RCFile, ORC, Parquet 110) Data Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD/GPFDIST9. Cluster Resource Management: Mesos, Yarn, Helix, Llama, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurn, Torque, Globus Tools, Pilot Jobs9. File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage 9. Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisc			
Eduroam DenStack Keystone, Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem Stratosphere (Apache Flink), Reef, Hama, Giraph, Pregel, Pegasus, Ligra, GraphChi, Galois, Medusa-GPU, MapGraph, Totem 13 Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs SAML OAuth 12 In-memory databases/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Hazelcast, Encache, Infinispan 12 Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC 13 Extraction Tools: UIMA, Tika 14 Ci SQL(NewSQL): Oracle, DB2, SQL Server, SQLite, MySQL, PostgreSQL, CUBRID, Galera Cluster, SciDB, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, N1QL, BlinkDB 11B NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongOB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, 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, Pivotal GPLOAD/GPFDIST 9 OLuster 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, Gluter, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage 7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr			
OpenStack Keystone, LDAP, Sentry, Monitoring: Amagination of the sentral sentr			
 Keystone, LDAP, Senty. 13) Inter process communication Collectives, point-to-point, publish-subscribe: MPI, Harp, Netty, ZeroMQ, ActiveMQ, RabbitMQ, NaradaBrokering, QPid, Kafka, Kestrel, JMS, AMQP, Stomp, MQTT, Marionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event Hubs 14) In-memory databases/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Hazelcast, Ehcache, Infinispan 12) Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC 13) Inter process (Job Coll, Cloud: Amazon SDS, Lambda, Google Cloud SQL, Azure SQL, CUBRID, Galera Cluster, SciDB, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google FI, IBM dashDB, N1QL, BlinkDB 11B, NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, 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, Pivotal GPLOAD/GPEDIST 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 (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor			
LDAP, Sentry, Sqrt, OpenD, SAML OAuthJonus Process Communication Concent Concent System Concent Concent C			
Sqr1, OpenID, SAML OAuthMarionette Collective, Public Cloud: Amazon SNS, Lambda, Google Pub Sub, Azure Queues, Event HubsMonitoring: Ambari, Ganglia, Nagios, Inca12) In-memory databases/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Hazelcast, Ehcache, Infinispan12) Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC12) Nosicet, Fatana Mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC12) Extraction Tools: UIMA, Tika11C: SQL(NewSQL): Oracle, DB2, SQL Server, SQLite, MySQL, PostgreSQL, CUBRID, Galera Cluster, SciDB, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, NIQL, BlinkDB11B) NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqr1, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame Public Cloud: Azure Table, Amazon Dynamo, Google DataStoreMay 15 20159) Cluster Resource Management: Mesos, Yarn, Helix, Llama, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs8) File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S, Azure Blob, Google Cloud Storage7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuint MaaS, Facebook Tupervare, A			
 SAML OAuth 12) In-memory databases/caches: Gora (general object from NoSQL), Memcached, Redis, LMDB (key value), Hazelcast, Ehcache, Infinispan 12) Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC Distraction Tools: UIMA, Tika 11C) SQL(NewSQL): Oracle, DB2, SQL Server, SQLite, MySQL, PostgreSQL, CUBRID, Galera Cluster, SciDB, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, N1QL, BlinkDB 11B) NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrtl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame May 15 Otata Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD/GPFDIST Ol Data Transport: BitTorrent, Methy, Torque, Globus Tools, Pilot Jobs File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api Iaas Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds <th></th><th></th>			
 4) value), Hazelcast, Ehcache, Infinispan 12) Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC 13) Object-relational mapping: Hibernate, OpenJPA, EclipseLink, DataNucleus, ODBC/JDBC 12) Extraction Tools: UIMA, Tika 11) C) SQL(NewSQL): Oracle, DB2, SQL Server, SQLite, MySQL, PostgreSQL, CUBRID, Galera Cluster, SciBB, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, N1QL, BlinkDB 11B) NoSQL: Lucene, Soir, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, 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, Pivotal GPLOAD/GPFDIST 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: Libivit, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack, Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) Iaas Managem	SAML OAuth		
Monitoring: Ambari, Ganglia, Nagios, Inca 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, CUBRID, Galera Cluster, SciDB, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, N1QL, BlinkDB 21 layers Google F1, IBM dashDB, N1QL, BlinkDB 11B) NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrfl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame Packages 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, Pivotal GPLOAD/GPFDIST 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: Libivirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent	4)		
Ambari, Ganglia, Nagios, Inca12) Extraction Tools: UIMA, Tika11. SQL (NewSQL): Oracle, DB2, SQL Server, SQLite, MySQL, PostgreSQL, CUBRID, Galera Cluster, SciDB, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, N1QL, BlinkDB21 layers Over 350118) NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame Public Cloud: Azure Table, Amazon Dynamo, Google DataStoreMay 15 201510) Data Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD/GPFDIST9) Cluster Resource Management: Mesos, Yarn, Helix, Llama, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs8) File systems: HDFS, Swift, Haystack, f4, Cinder, Cep, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api5) Iaas Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux- Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud,			
Gangua, Nagios, Inca11C) SQL(NewSQL): Oracle, DB2, SQL Server, SQLite, MySQL, PostgreSQL, CUBRID, Galera Cluster, SciDB, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, N1QL, BlinkDB21 layers Over 350I1B) NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame Public Cloud: Azure Table, Amazon Dynamo, Google DataStoreMay 15 2015OLuster Resource Management: MOS, NetCDF, CDF, HDF, OPeNDAP, FITS, RCFile, ORC, Parquet 10) Data Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD/GPFDISTMay 15 20159) 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: Libvint, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaliStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) LaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux- Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere<			
Nagios, indaSciDB, Rasdaman, Apache Derby, Pivotal Greenplum, Google Cloud SQL, Azure SQL, Amazon RDS, Google F1, IBM dashDB, N1QL, BlinkDB21 layersGoogle F1, IBM dashDB, N1QL, BlinkDBOver 350I1B) NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LeveIDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrtl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame Public Cloud: Azure Table, Amazon Dynamo, Google DataStoreMay 15I1A) File management: iRODS, NetCDF, CDF, HDF, OPeNDAP, FITS, RCFile, ORC, Parquet10) Data Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD/GPFDIST9) Cluster Resource Management: Mesos, Yarn, Helix, Llama, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs8) File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux- Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, A	-		
21 layersGoogle F1, IBM dashDB, N1QL, BlinkDBOver 35011B) NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LevelDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame Public Cloud: Azure Table, Amazon Dynamo, Google DataStoreMay 15 201510) Data Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD/GPFDIST9) Cluster Resource Management: Mesos, Yarn, Helix, Llama, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs8) File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux- Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds	Nagios, Inca		
 11By NoSQL: Lucene, Solr, Solandra, Voldemort, Riak, Berkeley DB, Kyoto/Tokyo Cabinet, Tycoon, Tyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LevelDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, 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, Pivotal GPLOAD/GPFDIST 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 (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) JaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux- Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 	21		
Over 350 Software PackagesTyrant, MongoDB, Espresso, CouchDB, Couchbase, IBM Cloudant, Pivotal Gemfire, HBase, Google Bigtable, LevelDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame Public Cloud: Azure Table, Amazon Dynamo, Google DataStoreMay 15 2015Public Cloud: Azure Table, Amazon Dynamo, Google DataStoreMay 15 20159 Cluster Resource Management: Mesos, Yarn, Helix, Llama, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs8) File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux- Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds	ZI layers		
Software PackagesBigtable, LevelDB, Megastore and Spanner, Accumulo, Cassandra, RYA, Sqrrl, Neo4J, Yarcdata, AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame Public Cloud: Azure Table, Amazon Dynamo, Google DataStoreMay 15 2015IIA) File management: iRODS, NetCDF, CDF, HDF, OPeNDAP, FITS, RCFile, ORC, ParquetMay 15 2015Olusta Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD/GPFDIST9) Cluster Resource Management: Mesos, Yarn, Helix, Llama, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs8) File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux- Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds	Over 350		
 Packages 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, Pivotal GPLOAD/GPFDIST 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 (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 			
 11A) File management: iRODS, NetCDF, CDF, HDF, OPeNDAP, FITS, RCFile, ORC, Parquet 10) Data Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD/GPFDIST 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 (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 	Software	AllegroGraph, Blazegraph, Facebook Tao, Titan:db, Jena, Sesame	
 11A) File management: iRODS, NetCDF, CDF, HDF, OPeNDAP, FITS, RCFile, ORC, Parquet 10) Data Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal GPLOAD/GPFDIST 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 (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 	Packages		
May 15 2015GPLOAD/GPFDIST9) Cluster Resource Management: Mesos, Yarn, Helix, Llama, Google Omega, Facebook Corona, Celery, HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs8) File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux- Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds	U	11A) File management: iRODS, NetCDF, CDF, HDF, OPeNDAP, FITS, RCFile, ORC, Parquet	
 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 (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 		10) Data Transport: BitTorrent, HTTP, FTP, SSH, Globus Online (GridFTP), Flume, Sqoop, Pivotal	
 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 (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux- Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 	May 15	GPLOAD/GPFDIST	
 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 (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 	2015		
 Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage 7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 	2013	HTCondor, SGE, OpenPBS, Moab, Slurm, Torque, Globus Tools, Pilot Jobs	
 7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 		8) File systems: HDFS, Swift, Haystack, f4, Cinder, Ceph, FUSE, Gluster, Lustre, GPFS, GFFS	
 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 		Public Cloud: Amazon S3, Azure Blob, Google Cloud Storage	
 6) DevOps: Docker (Machine, Swarm), Puppet, Chef, Ansible, SaltStack, Boto, Cobbler, Xcat, Razor, CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 		7) Interoperability: Libvirt, Libcloud, JClouds, TOSCA, OCCI, CDMI, Whirr, Saga, Genesis	
 CloudMesh, Juju, Foreman, OpenStack Heat, Sahara, Rocks, Cisco Intelligent Automation for Cloud, Ubuntu MaaS, Facebook Tupperware, AWS OpsWorks, OpenStack Ironic, Google Kubernetes, Buildstep, Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 			
 Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 			
 Gitreceive, OpenTOSCA, Winery, CloudML, Blueprints, Terraform, DevOpSlang, Any2Api 5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds 			
Vserver, OpenStack, OpenNebula, Eucalyptus, Nimbus, CloudStack, CoreOS, rkt, VMware ESXi, vSphere and vCloud, Amazon, Azure, Google and other public Clouds			
and vCloud, Amazon, Azure, Google and other public Clouds		5) IaaS Management from HPC to hypervisors: Xen, KVM, Hyper-V, VirtualBox, OpenVZ, LXC, Linux-	
Networking: Google Cloud DNS, Amazon Route 53			
		Networking: Google Cloud DNS, Amazon Route 53	

One also sees systems like Apache Pig offering data parallel interfaces. At the high layer we see both programming models and Platform as a Service toolkits where the Google App Engine is well known but there are many entries including the recent BlueMix from IBM. The implementation of the analytics layer depends on details of orchestration and especially programming layers but probably most important are quality parallel algorithms. As many machine learning algorithms involve linear algebra, HPC expertise is directly applicable as is fundamental mathematics needed to develop O(NlogN) algorithms for analytics that are naively $O(N^2)$ [13].

3. Implications for Workflow

The orchestration or workflow layer has seen an explosion of recent activity in the commodity space although with systems like Pegasus, Taverna, Kepler, and Swift, HPC has substantial experience stemming from the Grid and service oriented architecture communities. There are ABDS orchestration dataflow systems like Tez but more interestingly projects like Apache Crunch with a data parallel emphasis based on ideas from Google FlumeJava. A modern version of the latter presumably underlies Google's recently announced Cloud Dataflow that unifies support of multiple batch and streaming components; a capability that we expect to become common. Cascading which is open source but not Apache is a popular ABDS workflow toolkit. NiFi is a recent Apache workflow environment from NSA, the US National Security Agency. NSA has a strong ABDS commitment having previously donated the popular NoSQL store Accumulo and spun off Sqrrl at the interface of big data and security. It seems important to evaluate the ABDS workflow systems compared to HPC approaches like Pegasus, Kepler, Taverna, and Swift; this evaluation should include ease of integration of other levels of HPC-ABDS including batch and streaming processing as well as SQL and NoSQL data systems. One area of HPC and ABDS commonality is the use of Python (IPython) to script orchestration environments.

We suggest that there we can find another important opportunity by looking at levels 5 (Infrastructure), 6 (DevOps), 15B (PaaS hosting environments) and 17 (Workflow and orchestration) which are naturally linked. At the IaaS level 5, there has been lots of recent progress in both hypervisor and Linux containerbased virtual environments with OpenStack and Docker being key technologies. This progress has been accompanied by the rapid emergence of DevOps (level 6) to build software defined or automation systems. At level 6 one finds in particular, tools like Chef, Ansible, OpenStack Heat and Kubernetes. These DevOps tools define computer systems in a variety of languages and use this definition to automate system deployment. This approach was aimed at easing the systems administration task but has other interesting features. Firstly it makes interoperability at the infrastructure level easier as the DevOps tools can take the scripted system and deploy it on different hypervisors, containers or just bare metal without any user or administrative actions needed. Here standards at level 7 - especially libcloud or TOSCA and OCCI are important. One can use the same script on Docker, OpenStack or Amazon for example. Now let's understand importance of these ideas for workflow. DevOps defines a system for deployment whilst workflow defines systems for execution; these are different but closely related goals. This close relationship can be seen by comparing the two OASIS standards: BPEL for workflow and TOSCA for DevOps. They are very similar and have key people in common including the group at Stuttgart led by Frank Leymann. This group has developed the OpenTOSCA software automation environment and tools such as Winery that span orchestration and deployment. We don't recommend this particular approach - partly because BPEL was not so successful but do consider the linking of workflow with DevOps as an important area that should be included in any workflow initiative. We can quote the mission of OpenStack Heat "The mission of the OpenStack Orchestration program is to create a human- and machine-accessible service for managing the entire lifecycle of infrastructure and applications within OpenStack clouds". There are several other DevOps tools in the lifecycle and management area which could be very valuable in developing a robust HPC workflow stack that can run on a broad IaaS and not just OpenStack.

We can illustrate another idea with OpenStack Sahara that is a project to automatically build Hadoop systems on OpenStack. This combines DevOps with an idea often called PaaS or Platform as a Service, listed as level 15B in the Figure. PaaS identifies a set of popular tools and offers a complete managed environment supporting them. It can be considered as integrating DevOps and workflow for a particular application stack as Sahara does for Hadoop. Although some PaaS are focused, others like Heroku and Stackato support general application stacks. The integration with DevOps is illustrated by Dokku that

implements Heroku PaaS functionality on Docker where Heroku has buildpacks that are its DevOps style system definitions. The field is pretty incoherent but the trend is clear. Application stacks are defined as scripts which are used through the full life cycle from deployment through execution. These form well supported Platforms as a Service that can be deployed on a variety of IaaS. We suggest that HPC should investigate this model and build HPC systems in the DevOps supported PaaS style.

4. Conclusions

We have found it very fruitful to consider HPC-ABDS, which merges High Performance Computing with the Commodity Big Data Stack ABDS. In this white paper we have focused on the special implications of this approach for workflow and orchestration. Our analysis suggests an important broadening of HPC workflow to consider both ABDS workflow systems and the integration of workflow with DevOps and PaaS ideas. We stress that our ideas are synergistic with the concept of Software as a Service argued for at the workshop.

References

- 1. Judy Qiu, Shantenu Jha, Andre Luckow, and Geoffrey C.Fox, *Towards HPC-ABDS: An Initial High-Performance Big Data Stack*, in *Building Robust Big Data Ecosystem ISO/IEC JTC 1 Study Group on Big Data*. March 18-21, 2014. San Diego Supercomputer Center, San Diego. <u>http://grids.ucs.indiana.edu/ptliupages/publications/nist-hpc-abds.pdf</u>.
- 2. Geoffrey Fox, Judy Qiu, and Shantenu Jha, *High Performance High Functionality Big Data Software Stack*, in *Big Data and Extreme-scale Computing (BDEC)*. 2014. Fukuoka, Japan. http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/whitepapers/fox.pdf.
- 3. Shantenu Jha, Judy Qiu, Andre Luckow, Pradeep Mantha, and Geoffrey C. Fox, A Tale of Two Data-Intensive Approaches: Applications, Architectures and Infrastructure, in 3rd International IEEE Congress on Big Data Application and Experience Track. June 27- July 2, 2014. Anchorage, Alaska. <u>http://arxiv.org/abs/1403.1528</u>.
- 4. Geoffrey Fox, Judy Qiu, Shantenu Jha, Supun Kamburugamuve, and Andre Luckow, HPC-ABDS High Performance Computing Enhanced Apache Big Data Stack, in Invited talk at 2nd International Workshop on Scalable Computing For Real-Time Big Data Applications (SCRAMBL'15) at.CCGrid2015, the 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. 2015, IEEE. Shenzhen, Guangdong, China. http://dsc.soic.indiana.edu/publications/HPC-#ABDSDescribedv2.pdf.
- 5. *HPC-ABDS Kaleidoscope of over 300 Apache Big Data Stack and HPC Tecnologies*. [accessed 2014 April 8]; Available from: <u>http://hpc-abds.org/kaleidoscope/</u>.
- Geoffrey C. Fox, Shantenu Jha, Judy Qiu, and Andre Luckow, *Towards an Understanding of Facets and Exemplars of Big Data Applications*, in 20 Years of Beowulf: Workshop to Honor Thomas Sterling's 65th Birthday October 14, 2014. Annapolis <u>http://grids.ucs.indiana.edu/ptliupages/publications/OgrePaperv9.pdf</u>.
- 7. Geoffrey Fox and Wo Chang, *Big Data Use Cases and Requirements*, in *1st Big Data Interoperability Framework Workshop: Building Robust Big Data Ecosystem ISO/IEC JTC 1 Study Group on Big Data* March 18 - 21, 2014. San Diego Supercomputer Center, San Diego. <u>http://grids.ucs.indiana.edu/ptliupages/publications/NISTUseCase.pdf</u>.
- 8. *NIST Big Data Use Case & Requirements*. 2013 [accessed 2015 March 1]; Available from: <u>http://bigdatawg.nist.gov/V1 output docs.php</u>.
- 9. Geoffrey C.Fox, Shantenu Jha, Judy Qiu, and Andre Luckow, *Ogres: A Systematic Approach to Big Data Benchmarks*, in *Big Data and Extreme-scale Computing (BDEC)* January 29-30, 2015. Barcelona. http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/whitepapers/OgreFacets.pdf.
- 10. Geoffrey C. FOX, Shantenu JHA, Judy QIU, Saliya EKANAYAKE, and Andre LUCKOW, *Towards a Comprehensive Set of Big Data Benchmarks*. February 15, 2015. <u>http://grids.ucs.indiana.edu/ptliupages/publications/OgreFacetsv9.pdf</u>.
- Geoffrey Fox. Data Science Curriculum: Indiana University Online Class: Big Data Open Source Software and Projects.
 2015 [accessed 2015 March 31]; Available from: <u>http://bigdataopensourceprojects.soic.indiana.edu/</u>.
- 12. Bingjing Zhang, Yang Ruan, and Judy Qiu, in *IEEE International Conference on Cloud Engineering (IC2E)*. March 9-12, 2015. Tempe AZ. <u>http://grids.ucs.indiana.edu/ptliupages/publications/HarpQiuZhang.pdf</u>.
- 13. Committee on the Analysis of Massive Data; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Their Applications; Division on Engineering and Physical Sciences; National Research Council, *Frontiers in Massive Data Analysis*. 2013: National Academies Press. <u>http://www.nap.edu/catalog.php?record_id=18374</u>