**1**

**Towards HPC-ABDS: An Initial High-Performance BigData Stack**

Geoffrey C.Fox, School of Informatics and Computing, Indiana University, Bloomington, USA

Shantenu Jha, RADICAL, Rutgers University, Piscataway, USA

Judy Qiu, School of Informatics and Computing, Indiana University, Bloomington, USA

Andre Luckow, RADICAL, Rutgers University, Piscataway, USA

Many scientific problems depend on the ability to analyze and compute on large amounts of data. This analysis often does not scale well, i.e. its effectiveness is hampered by the increasing volume, variety and rate of change (velocity) of big data. There is a need to integrate features of traditional high-performance computing, such as scientific libraries, communication and resource management middleware, with the rich set of capabilities found in the commercial Big Data ecosystem, resulting in an integrated system generi- cally called high-performance big data system (HPBDS). Our proposed preliminary implementation of the HPBDS – includes many important software systems such as Hadoop available from the Apache open source community and thus referred to as High-Performance Computing-Big Data Stack (HPC–ABDS) – has two fundamental building blocks: (i) Middleware for Data-Intensive Analytics and Science (MIDAS) that will enable scalable applications with the performance of HPC (High Performance Computing) and the rich functionality of the commodity Apache Big Data Stack. (ii) The second building block will design and imple- ment a set of cross-cutting high-performance data-analysis libraries SPIDAL (Scalable Parallel Interopera- ble Data Analytics Library), which will support new programming and execution models for data-intensive analysis in a wide range of science and engineering applications. These libraries will be implemented to be scalable and interoperable across a range of computing systems including clouds, clusters and supercomput- ers. The project libraries will have the same beneficial impact on data analytics that scientific libraries such as PETSc, MPI and SCALAPACK have had for supercomputer simulations. In this paper, we study many Big Data applications from a variety of research and commercial areas and suggest a set of characteristic features and possible kernel benchmarks that stress those features for data analytics. We draw conclusions for the hardware and software architectures that are suggested by this analysis.

General Terms: Big Data, HPC, Apache Hadoop

**ACM Reference Format:**

*ACM* 1, 1, Article 1 (August 2014), 23 pages.

DOI:<http://dx.doi.org/10.1145/0000000.0000000>

1. **INTRODUCTION**

The growing importance of data in many fields including physical and biological sci- ences, and the ability to derive insight and knowledge from increasing volumes of com- plex data, points to the importance of advanced analytics. Analytics needs to be able to utilize the full range of available infrastructure, however, the coupling between tools, analytic engines and infrastructure is often rigid, thus it is often difficult to employ existing solutions for contemporary environments that they were not natively or origi- nally designed for. Further, many tools were developed at a time when parallelism was not essential. In addition, interoperability at multiple levels remains elusive, as well as difficult, and scalable yet general-purpose and broadly applicable solutions in the form of analytic libraries and abstractions are noticeable by their absence.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is per- mitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org.](mailto:permissions@acm.org)

*O*c 2014 ACM 1539-9087/2014/08-ART1 $15.00

DOI:<http://dx.doi.org/10.1145/0000000.0000000>

1:2 G. Fox et al.

The importance of advanced analytics to derive insight and knowledge from increas- ing volumes of complex data will continue to grow. The enterprise community have made impressive gains and seem to have converged around the Apache stack, a dis- tinctive feature is the existence of many implementations of the specific components of the Apache stack, providing sufficient richness in the trade-off between performance and capability. In contrast, within the scientific computing community, progress has been reliant either on long-term foundational advances or short-term hardware fixes. as opposed to integrated approaches that marry the relative technical strengths of the two communities yet deliver these as implementations usable on high performance and distributed computing HPDC infrastructure such as XSEDE, OSG and other domain- specific infrastructure. In both domains, scalable yet general-purpose and broadly ap- plicable solutions in the form of analytic libraries and abstractions are noticeable by their absence.

To remedy this major gap and proffer an integrated solution that brings the best of recent advances to the service of extreme-scale science requirements on current and future science production platforms, we are developing HPC-ABDS – a first imple- mentation of a high-performance Big Data stack (HPBDS) that integrates the best of the Apache developments and HPC capabilities. HPC-ABDS will utilize and expose the integrated relative technical strengths of the two hitherto disjoint approaches and communities, yet it will focus on delivering these as production grade implementa- tions that will bring the best-of-both to shared-infrastructure – such as NSF’s XSEDE, DOE’s leadership machine, OSG and other domain-specific infrastructure, as well as the software developments underway as part of the SI2 software program. HPC-ABDS will translate these applications characteristics, infrastructural requirements and ex- isting capabilities into well-defined and implemented building blocks.

**Application Layer**

**Community and Exemplars**

**Remaining Apache Big Data stack**

**integrated without**

**Message**

**Analytics Libraries**



**HPC-ABDS MapReduce**



**Scalable Parallel Interoperable Data Analytics Libraries (SPIDAL)**



**need for HP**

**enhancements** (SQL-engines, Storm, Impala, Hive, Shark)

**Passing**

**Classic MapReduce**

**Search**

**Iterative MapReduce**

**Iterative Giraph**

*Programming Models*

**Middleware for Data- Intensive Analytics and Science (MIDAS)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Communication**  (MPI, RDMA, Hadoop Shuffle/Reduce, HARP Collectives, Giraph point-to-point) | | |  | **In-memory Data Abstractions**  (HBase, Object Stores, In-Memory, other NoSQL stores, Spatial) | | |
| **Higher-Level Workload Management** (TEZ, LLama) |  | **Workload Management**  (Pilots, Condor) | | |  | **Framework specific Scheduling** (e.g. YARN) |
| **External Data Access**  (Virtual Filesystem, GridFTP, SRM, SSH) | | |  | **Cluster Resource Manager**  (YARN, Mesos, SLURM, Torque, SGE) | | |
|  | | | | | | |
| **Compute, Storage and Data Resources**  (Nodes, Lustre, Cores, HDFS) | | | | | | |



*Resource Fabric*



Fig. 1. Key components of integrated HPBDS stack. Many capabilities unaffected by integration are not shown explicitly

The key components of such an integrated platform are shown in Fig. 1. The aim of HPC-ABDS is to aim for the performance of HPC and the breadth and productivity of ABDS. The resultant integrated architecture is targeted at both production high-end computing platforms (such as leadership machines and XSEDE), as well as (commer- cial) cloud computing. As part of HPC-ABDS, we propose two fundamental building blocks, Middleware for Data-Intensive Analytics and Science (MIDAS) and the Scal- able Parallel Interoperable Data Analytics Library (SPIDAL).

The high-performance community has prospered thanks to libraries like MPI, PETSc and SCALAPACK; SPIDAL brings this concept to data-intensive applications.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:3

SPIDAL Parallel Analytics Libraries will capture system abstractions and expose ap- plication requirements and MIDAS the middleware upon which to build such libraries that are interoperable yet high-performance, SPIDAL will enable interoperable high performance data analytics and is based upon a careful analysis of architectures, tools and application characteristics/requirements. Although built from many existing com- ponents and capabilities, MIDAS is conceptualized and designed from first-principles to ensure our productivity, interoperability and performance goals.

In earlier work [Jha et al. 2014b], we have discussed the need for merging the two “common” stacks. In addition to a qualitative motivation, Ref [Jha et al. 2014b] pro- vided a quantitative analysis of the type of abstractions and support required to enable a successful hybrid stack. In this paper, we will move from a general motivation of the need of a hybrid approach, to a discussion of the design philosophy & objectives of a specific implementation of HPC-ABDS, which serves as a first prototype towards a gen- eral purpose, interoperable high-performance big data stack for to support analytics on high-end clusters, clouds and supercomputers.

1. **SOURCES OF INFORMATION**

In discussing the structure of Big Data applications, let us first examine the inevitably incomplete input that we used to do our analysis. We have gained quite a bit of experi- ence from our research over the years, but 3 explicit sources that we used were a recent use case survey by NIST from Fall 2013 [NIST 2013a]; a survey of data intensive re- search applications by Jha et al. [Jha et al. 2014a; Jha et al. 2013]; and a study of members of data analytics libraries including R [R Project 2012], Mahout [Apache Ma- hout 2012] and MLLib [MLLib 2014]. Following is a summary of the first two sources. The NIST Big Data Public Working Group (NBD-PWG) was launched in June 2013 with a set of working groups covering Big Data Definitions, Taxonomies, Require- ments, Security and Privacy Requirements, Reference Architectures White Paper Sur- vey, Reference Architectures, Security and Privacy Reference Architectures and Big Data Technology Roadmap. The Requirements working group gathered 51 use cases from a public call and then analyzed them in terms of requirements of a reference architecture [NIST 2013b]. Here we will strive to identify common patterns and char- acteristics, which can be used to guide and evaluate Big Data hardware and software. The 51 use cases are organized into nine broad areas with the number of associated use cases in parentheses: Government Operation (4), Commercial (8), Defense (3), Health- care and Life Sciences (10), Deep Learning and Social Media (6), The Ecosystem for Research (4), Astronomy and Physics (5); Earth, Environmental and Polar Science (10)

and Energy (1).

Note that the majority of use cases come from research applications, although com-

mercial, defense and government operations have some coverage. A template was pre-

pared by the Requirements working group, which allowed experts to categorize each

use case by 26 features:

Use case Actors/Stakeholders and their roles and responsibilities; use case goals

and description. Specification of current analysis covering compute system, storage,

networking and software. Characteristics of use case Big Data with Data Source (distributed/centralized), Volume (size), Velocity (e.g. real time), Variety (multiple datasets, mashup), Variability (rate of change). The so-called Big Data Science (collec-

tion, curation, analysis) with Veracity (Robustness Issues, semantics), Visualization, Data Quality (syntax), Data Types and Data Analytics. These detailed specifications were complemented by broad comments including Big Data Specific Challenges (Gaps), Mobility issues, Security and Privacy Requirements and identification of issues for generalizing this use case.

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:4 G. Fox et al.

Table I. What is Parallelism Over for NIST Use Cases

|  |  |
| --- | --- |
| **General Class** | **Examples** |
| People | Users (but see below) or Subjects of application and often both |
| Decision makers | Researchers or doctors (users of application) |
| Items | Experimental observations  Contents of online store  Images or Electronic Information nuggets  EMR: Electronic Medical Records (often similar to people parallelism) Protein or Gene Sequences  Material properties, Manufactured Object specifications, etc., in custom dataset |
| Modeled entities | Vehicles and people |
| Sensors | Internet of Things |
| Events | Detected anomalies in telescope, credit card or atmospheric data |
| Graph Nodes | RDF databases |
| Regular Nodes | Simple nodes as in a learning network |
| Information Units | Tweets, Blogs, Documents, Web Pages, etc. and characters/words in them |
| Files or data | To be backed up, moved or assigned metadata |
| Particles/cells/mesh points | Used in parallel simulations |

The complete set of 51 responses in addition to a summary from the working group of applications, current status and futures as well as extracted requirements can be found in [NIST 2013b]. They are summarized in the Appendix which also gives 20 other use cases coming from the NBD-PWG which do not have the detailed 26 fea- ture template recorded. These 20 cover enterprise data applications and security and privacy.

* 1. **Properties of the 51 NIST Use Cases**

Tables I to III summarize the characteristics of the 51 use cases, which we will com- bine with other input for the Ogres. Note that Big Data and parallel programming are intrinsically linked as any Big Data analysis is inevitably processed in parallel. Parallel computing is almost always implemented by dividing the data between pro- cessors (data decomposition); the richness here is illustrated in Table I which lists the members of space that is decomposed for different use cases; of course these sources of parallelism are broadly applicable outside the 51 use cases they were extracted from. In Table II, we identify use case features for 15 use cases and map these to Ogre facets. The second column maps to the use case that illustrate this feature; note these are not exclusive so any one use case will illustrate many features.

It’s important to note that machine learning is commonly used, but there is an in- teresting distinction between what is termed Local (LML) or Global machine learning (GML) in Table II. In LML, there is parallelism over items of Table I and machine learning is applied separately to each item; needed machine learning parallelism is limited, typified by use of accelerators (GPU). In GML, the machine learning is ap- plied over the full dataset with MapReduce, MPI or an equivalent. Typically GML comes from maximum likelihood or *χ*2 with a sum over the data items - documents, sequences, items to be sold, images, etc., and often links (point-pairs). Usually GML is a sum of positive numbers. as in least squares, and is illustrated by algorithms like PageRank, clustering/community detection, mixture models, topic determination, Mul- tidimensional scaling, and (Deep) Learning Networks. Somewhat quixotically, GML can be termed Exascale Global Optimization or EGO. \*\*\*AL: maybe some details on what we mean by exascale global optimization would be useful.

The difference between LML and GML is illustrated in Table III, which contrasts 9 of the 51 NIST use cases that involve image-based data. For example, use case 18 with light source data is largely independent machine learning on each image from the

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:5

Table II. Some Features of NIST Use Cases

|  |  |  |
| --- | --- | --- |
| **Abbrev.** | **#** | **Description** |
| PP | 26 | Pleasingly Parallel or Map Only |
| MR | 18 | Classic MapReduce MR (add MRStat below for full count) |
| MRStat | 7 | Simple version of MR where key computations are simple reduction as found in  statistical averages such as histograms and averages |
| MRIter | 23 | Iterative MapReduce or MPI |
| Graph | 9 | Complex graph data structure needed in analysis |
| Fusion | 11 | Integrate diverse data to aid discovery/decision making; could involve sophisti-  cated algorithms or could just be a portal |
| Streaming | 41 | Some data comes in incrementally and is processed this way |
| Classify | 30 | Classification: divide data into categories |
| S/Q | 12 | Index, Search and Query |
| CF | 4 | Collaborative Filtering for recommender engines |
| LML | 36 | Local Machine Learning (Independent for each parallel entity) |
| GML | 23 | Global Machine Learning: Deep Learning, Clustering, LDA, PLSI, MDS, Large  Scale Optimizations as in Variational Bayes, MCMC, Lifted Belief Propagation, Stochastic Gradient Descent, L-BFGS, Levenberg-Marquardt . Can call EGO or Exascale Global Optimization with scalable parallel algorithm |
|  | 51 | Workflow: Universal so no label |
| GIS | 16 | Geotagged data and often displayed in ESRI, Microsoft Virtual Earth, Google  Earth, GeoServer etc. |
| HPC | 5 | Classic large-scale simulation of cosmos, materials, etc. generating (visualization)  data |
| Agent | 2 | Simulations of models of data-defined macroscopic entities represented as agents |

Table III. 9 Image-based NIST Use Cases

|  |  |  |
| --- | --- | --- |
| **Abbrev.** | **#** | **Description** |
| PP, LML,  MR for search | 17 | Moving to terabyte size 3D images, Global Classification |
| PP, LML | 18 | Biology and Materials |
| GML | 26 | Stanford ran 10 million images and 11 billion parameters on a 64 GPU HPC;  vision (drive car), speech, and Natural Language Processing |
| GML | 27 | Fit position and camera direction to assemble 3D photo ensemble |
| PP, LML,  GML | 36 | Processing of individual images for events based on classification of image struc-  ture (GML) |
| PP, LML  moving to GML | 43 | Identify glacier beds and snow layers See GML when one addresses full ice sheet |
| PP | 44 | Find and display slippage from radar images. Includes Data Product Delivery,  and Data Services |
| PP, LML,  GML | 45, 46 | Find paths, classify orbits, classify patterns that signal earthquakes, instabilities,  climate, turbulence |

source, i.e. LML. In contrast deep learning in use case 26 is constructing a learning network integrating all the images.

1. **TOWARDS A HIGH-PERFORMANCE BIG DATA SOFTWARE (HPBDS) ENVIRONMENT**

As alluded to, the HPC-ABDS [Jha et al. 2014b] approach was partially inspired by the NIST big data initiative [NIST 2013a] that generated a collection of 71 use cases as well as a taxonomy, reference architecture, roadmap and study of security and privacy (see collaboration letter of Bob Marcus). Later [Chang 2014] meetings identified use case patterns and mapped them to the NIST reference architecture. Figure 2 summa- rizes the ideas in an HPC-ABDS hourglass. \*\*\*SJ: need to reference hourglass picture

Ref [Geoffrey C. Fox and Luckow 2014] identified a set of Ogres (big data patterns) covering big data analytics with multiple facets including specific algorithms, problem

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:6 G. Fox et al.

HPC ABDS SYSTEM (Middleware)

120 SoIware Projects

**System Abstac2on/Standards**

6 Data Format 6 Storage

6 **HPC Yarn for Resource management**

* **Horizontally scalable parallel programming model**
* **Collec6ve and Point to Point Communica6on**

**! Support for itera6on (in!memory processing)**

**Applica2on Abstrac2ons/Standards**

Graphs, Networks, Images, GeospaDal

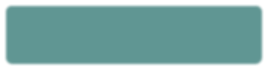
...

**Scalable Parallel Interoperable Data Analy2cs Library (SPIDAL)**

High performance Mahout, R, Matlab …..

**High Perfomance Applica2ons**

Fig. 2. HPC-ABDS Hour Glass \*\*\*SJ: Should we invert the hour glass?



architecture and its features, application class and data source structure. Here we ex- ploit these facets to identify the programming models, data source models and overall end-to-end application models that MIDAS needs to support.

Based upon an analysis of the Ogres, we identify the importance of the 5 parallel programming models. The parallel programming models supported by HPC-ABDS are:

\*\*\*AL: some overlap to table 4

* PM1) Pleasingly Parallel (PP) includes many cases where there are sophisticated

local machine learning applied in parallel – as in parallel image processing without

global optimizations.

* PM2) Search (Srch) includes collaborative filtering (in Mahout), motif (meme) de-

tection in graph (network) algorithms, and spatial relationship based queries for

spatial data, and is implemented using classic MapReduce or non-iterative Giraph.

* PM3) Iterative MapReduce or Map-Collective using Collective Communication are seen in many global machine learning algorithms applied over the complete dis- tributed dataset and are illustrated by clustering and dimensionality reduction us-

ing parallel linear algebra at their core.

* PM4) Iterative Giraph is Map-Communication with point-to-point communication

and includes graph algorithms such as maximum clique, connected component, find-

ing diameter, and community detection. The problems differ in the difficulty of de-

termining the data partitioning and this classic parallel load balancing issue can

need sophisticated runtime techniques.

* PM5) Asynchronous thread-based graph algorithms These are illustrated by short-

est path and betweenness centrality algorithms for shared memory machines and we do not integrate them into HPC-ABDS in this proposal.

In Table IV, we present 5 distinct problem architectures that map into 5 distinct system architectures to cover the Ogres. The central architectures are Tables I and IV which correspond exactly to the four forms of MapReduce that we have presented pre- viously [Ekanayake et al. 2010] summarized in Figure 3. Note this only describes some core features of the facets in [Geoffrey C. Fox and Luckow 2014]. There are many other issues that need to be addressed including support of workflow. In particular the ar-

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:7

Table IV. Distinctive Software/Hardware Architectures for Data Analytics

|  |  |
| --- | --- |
| 1. Pleasingly Paral-  lel (Map Only) | Includes local machine learning (LML) as in parallel decomposition over  items and applying data processing to each item. Hadoop could be used as well as other High Throughput Computing or Many task tools |
| 2. Classic Map-  Reduce | Includes MRStat, search applications and those using collaborative filter-  ing and motif finding implemented using classic MapReduce (Hadoop) |
| 3. Iterative Map-  Collective | Iterative MapReduce using Collective Communication as needed in clus-  tering - Hadoop with Harp, Spark etc. |
| 4. Iterative Map-  Communication | Iterative MapReduce such as Giraph with point-to-point communication,  includes most graph algorithms such as maximum clique, connected com- ponent, finding diameter, community detection). Varys in difficulty of finding partitioning (classic parallel load balancing) |
| 5. Shared (Large)  Memory | Thread-based (event driven) graph algorithms such as shortest path and  Betweenness centrality. Large memory applications |

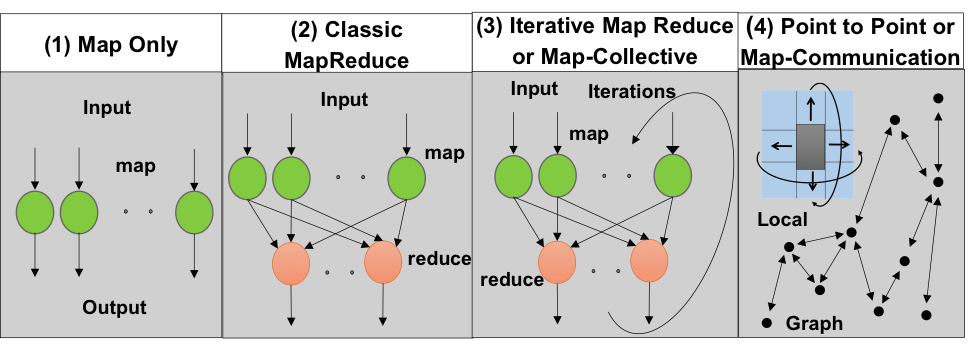


Fig. 3. The Four Forms of MapReduce

chitecture for the rapidly evolving field of streaming (distributed) data needs more work.

Note that we separate Map-Collective [Barrett et al. 1994; der Wijngaart et al. 2012] and Map-(Point to Point) Communication following the Apache projects Hadoop, Spark and Giraph that focus on these cases. These programming models or run times differ in communication style, application abstraction (key-value versus graph) and possible scheduling/load-balancing. HPC with MPI suggests that one could integrate into a single environment. This approach is illustrated by the Harp plug-in [Qiu and Zhang 2014] to Hadoop which supports both models.

SPIDAL will capture these common characteristics and requirements by identifying key abstractions; it will utilize capabilities of the underlying middleware that will be exposed via well-designed and engineered libraries. The MIDAS middleware imple- ments these with high performance in an ABDS context. MIDAS is based on abstrac- tions in the areas: a) Software defined System, b) Storage layer including a spatial access abstraction, c) Scheduling layer using advances in multi-level and application- level scheduling, d) Collective layer that permits Map Collective generalization of Iter- ative MapReduce e) Parallelism or Programming model which generalizes the popular Giraph and MPI SPMD models. \*\*\*SJ: what else should we write here?

1. **SCALABLE PARALLEL INTEROPERABLE DATA-ANALYTICS LIBRARY (SPIDAL)**

This section summarizes the core algorithms proposed initially for SPIDAL.

* 1. **Graph and Network Algorithms**
     1. *Architectures for Graph Algorithms.* Distributed-memory vs. shared-memory. For graph problems, researchers have developed both distributed-memory algorithms

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:8 G. Fox et al.

**Govt. Operations Defense**

**Commerical**

**Earth, Env. & Polar Science**

**HealthCare & Life Science**

Deep learning and social media

**Ecosystem for Research**

**Astronomy & Physics**

**Community and Exemplars**

**Remaining Apache Big Data stack**

**integrated without**

**Message**

**Analytics Libraries**

**HPC-ABDS MapReduce**

**Scalable Parallel Interoperable Data Analytics Libraries (SPIDAL)**

**need for HP**

**enhancements** (SQL-engines, Storm, Impala, Hive, Shark)

**Passing**

**Classic MapReduce**

**Search**

**Iterative MapReduce**

**Iterative Giraph**

*Programming Models*

**MI**ddleware for **D**ata-Intensive **A**nalytics and **S**cience (MIDAS) API



**Middleware for Data- Intensive Analytics and Science (MIDAS)**



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Communication**  (MPI, RDMA, Hadoop Shuffle/Reduce, HARP Collectives, Giraph point-to-point) | | |  | **In-memory Data Abstractions**  (HBase, Object Stores, In-Memory, other NoSQL stores, Spatial) | | |
| **Higher-Level Workload Management** (TEZ, LLama) |  | **Workload Management**  (Pilots, Condor) | | |  | **Framework specific Scheduling** (e.g. YARN) |
| **External Data Access**  (Virtual Filesystem, GridFTP, SRM, SSH) | | |  | **Cluster Resource Manager**  (YARN, Mesos, SLURM, Torque, SGE) | | |

**Compute, Storage and Data Resources**



*Resource Fabric*

(Nodes, Lustre, Cores, HDFS)

Fig. 4. SPIDAL: Library and Algorithms \*\*\*AL: programming models inconsistent with Figure 4 and Table 4.

[N. Edmonds 2010; Alam et al. 2013; Arifuzzaman et al. 2013; Zhao et al. 2010; Zhao et al. 2012][38-42] and shared-memory algorithms [Prountzos and Pingali 2013; Edi- ger et al. 2012; Bader and Cong 2005; 2004; Madduri et al. 2007; Madduri et al. 2009][43-48]. In a distributed memory system, each processor has its own local mem- ory, and data is partitioned so that each processor contains one partition in its memory. Since processors may need to communicate and exchange data with one another, poor locality is a major challenge for distributed-memory systems, causing communication overhead that can lead to decreased performance. A distributed memory system is good for graph problems with high locality. In a shared-memory system, data is stored in a common shared memory accessed by all processors and locality is not critical, although efficient thread parallelism may still be hard.

Message Passing Interface (MPI) vs. Giraph. MPI is a general-purpose distributed memory system for parallel programming, with efficient communication primitives. Efficient MPI implementations have been developed for a number of graph problems, which scale to very large networks, using problem-specific knowledge of the computa- tion and communication patterns; this requires significant HPC expertise. In contrast, Giraph is easy to use, but does not allow easy access to partitioning and load balancing. Two main challenges in parallelization are: obtaining good estimates of the computa- tion cost for each partition, and load balancing, both of which require problem specific insights [Nguyen et al. 2013]. This makes the problem of finding good partitions that minimize communication cost very challenging [Alam et al. 2013; Arifuzzaman et al. 2013].

MapReduce and Giraph for graph algorithms. We briefly summarize some of the main theoretical advances in understanding problems where these frameworks are effective. Karloff et al. [Karloff et al. 2010] show that any PRAM algorithm using sub- quadratic space and processors can be implemented in MapReduce in the same time. This implies that, in theory, many shared memory algorithms can be implemented in MapReduce, e.g., for approximating the diameter [Kang et al. 2011a; 2008], clustering

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:9

[Das et al. 2007] and counting triangles [Suri and Vassilvitskii 2011]. However, the theoretical analyses do not consider the complexity of the shuffle operation [Goel and Munagala 2012]. Klauck et al. [Klauck et al. 2013] study the Giraph framework from a theoretical perspective. They consider a number of graph problems, e.g., Spanning Tree and PageRank, and derive strong lower and upper bounds for the time complexity. For instance, they show the Minimum Spanning Tree can be computed in this model within O(n polylog(n)/K) time, while there is a lower bound of (n/K polylog(n)) on the MST problem, for a graph with n nodes on a machine with K processors. Furthermore, their results do not require any specific graph partitioning scheme. This suggests that Giraph is a promising approach for many graph problems.

* + 1. *Graph Algorithms in SPIDAL.* To deal with the challenges posed by massive net- works, we will develop new algorithmic techniques based on MapReduce and Giraph. In iterative MapReduce, we will explore techniques to decompose the problem and en- able efficient information exchange through the reduce operation. The problem of find- ing network motifs/subgraphs is well-studied [Zhao et al. 2010; Zhao et al. 2012; Alon et al. 2008; Aravind and Raman 2002; Gonen and Shavitt 2009; Milo et al. 2002; Qin and Gao 2012] with sequential algorithms, although [Ribeiro et al. 2012] addresses a shared-memory architecture. We developed distributed-memory parallel algorithms [Zhao et al. 2010; Zhao et al. 2012]; however, these algorithms are for a few special classes of motifs, e.g., trees. We propose to develop Giraph-based distributed-memory parallel algorithms for a more general class of motifs. We will also build on the tech- niques of Karloff et al. [Karloff et al. 2010] to explore efficient MapReduce implemen- tations for other PRAM algorithms with super-quadratic key-value space complexity, and identify problem classes which do not scale in MapReduce. For such problems, we will use Giraph. We will build on the general conversion theorem [Klauck et al. 2013] to develop implementations of algorithms for problems such as connectivity, subgraph enumeration and random graph generation in Giraph. Some of the results in [Klauck et al. 2013] yield constant factor or logarithmic approximation algorithms for problems such as shortest paths in near optimal time. We will explore the complexity of optimal algorithms for these problems.

Community detection problems (including clustering approaches) have only received serious attention recently. Current methods [Satuluri et al. 2011; Meyerhenke et al. 2011; Satuluri and Parthasarathy 2009; Fortunato 2010; Sarkar and Dong 2011; Zhang et al. 2009] do not scale well for large networks, including parallel algorithms [Meyerhenke et al. 2011; Satuluri and Parthasarathy 2009; Fortunato 2010; Sarkar and Dong 2011; Zhang et al. 2009] for shared-memory architecture that cannot sup- port massive-scale networks. We propose to develop scalable and distributed-memory parallel algorithms using Giraph. David Baders group (Georgia Tech) developed par- allel algorithms for some other network problems, such as shortest path and between- ness centrality [Madduri et al. 2007; Bader and J. JJ 1996; Bader and Madduri 2008; Madduri and Bader 2009; Bader 2010; Green et al. 2012; Jiang et al. 2009]. Edmonds, Hoefler and Lumsdaine [N. Edmonds 2010] recently developed a space-efficient par- allel algorithm for computing betweenness centrality in distributed-memory systems, while efficient sequential algorithms have been developed for selected network ana- lytics problems such as diameter, pagerank, and counting triangles, at CMU (Chris- tos Faloutsos) [Faloutsos ; Kang et al. 2009; Kang et al. 2011b; Tsourakakis et al. 2009] and Sandia National Lab [Zhang et al. 2009; Seshadhri et al. 2012a; Seshadhri et al. 2012b]. There also are sequential libraries for network analytics such as Pajek [Batagelj and Mrvar 1998], Pegasus [Faloutsos ] (Christos Faloutsos), SNAP [Leskovec

] (Jure Leskovec at Stanford), NetworkX [Hagberg et al. 2008] (at Los Alamos Na-

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:10 G. Fox et al.

tional Lab) and statnet [Handcock et al. 2003]. Similar libraries for parallel graph algorithms are needed to work with emerging massive networks.

* 1. **Spatial Queries and Spatial Analytics Algorithms Related Work**

Spatial Data Management Systems (SDBMS) have major limitations on managing and querying large scale spatial data. SDBMSs [pro 2013d; 2013g; ; 2013f] rely on parallel DBMS architectures such as a shared nothing architecture [Mehta and DeWitt 1995; Patel et al. 1997; pro 2013h; 2013b; 2013e] to scale out. Parallel SDBMSs through par- titioning are not optimized for computationally intensive operations such as geometric computations [Wang et al. 2012], and lack spatial partitioning to balance data and task loads [pro 2007]. Data loading overhead is another major bottleneck [Pavlo et al. 2009; Aji et al. 2013; Wang et al. 2013]. GIS systems [Chang 2003; pro 2013a] often use SDBMS as the backend spatial engine. Work in [Cary et al. 2009; Zhang et al. a; Zhang et al. b; Wang et al. ; Liu et al. 2010; Pozdnoukhov and Kaiser 2011; Li et al. 2012; Guo et al. 2010; Ballesteros et al. 2011; Ma et al. 2009; Akdogan et al. 2010] tries to tackle specific spatial algorithms using MapReduce, and ongoing MapReduce spatial query- ing systems include [Eldawy and Mokbel 2013; Daszak 2000]. Our Hadoop-GIS [Aji et al. 2013; pro 2013c] provides a general framework for spatial queries and analytics with MapReduce. Hadoop-GIS is integrated into Apache Hive to support declarative spatial queries.

Spatial queries and analysis are the crux of GIS and other spatial applications [Lon- gley et al. 2010]. We focus on developing the following representative methods. Spa- tial relationship based queries include spatial joins of multiple datasets (two-way or multi-way spatial joins), point-in-polygon queries, and window queries. Distance based queries identify correction of spatial objects based on proximity and common queries include nearest neighbor search. Density based spatial patterns include spatial clus- ters, hotspots, and anomalies. We will focus on density and statistics based spatial clustering. Spatial relationship modeling involves constructing spatial weights matri- ces or modeling spatial relationships using regression analysis, e.g. Geographically Weighted Regression (GWR) [Daniel 2004]. The library will draw on our earlier work motivated by pathology imaging spatial analytics [Aji et al. 2012] and geospatial spa- tial queries [Aji et al. 2013; pro 2013c]

Spatial data has multiple dimensions and there is no concept of a key in a spatial space for data partitioning or task partitioning. Our approach partitions space into tiles, and breaks down time-consuming spatial operations into smaller tasks on tiles, running them in parallel in HPC-ABDS MapReduce. Thus a tile serves as a key, and the objects of a tile are the value. To avoid data skew, density aware spatial parti- tioning methods are provided. Spatial query processing then becomes a problem of designing query methods that can run on tiles independently while preserving correct semantics. A generic framework for distributed spatial data processing has the follow- ing steps. Initialization includes data partitioning, uploading and indexing. For each spatial query, the index is used to identify regions of interest for the query. After that, spatial query processing is executed independently for each tile, in parallel. During each tile based query, on-demand indexing can be provided to accelerate queries. Then a normalization step will be performed to fix results due to partitioning. Most spatial processing methods can be mapped into above patterns, and spatial clustering follows global machine learning patterns.

* 1. **Core Image Processing Algorithms Related Work on Pathology informatics**

The availability of advanced scanners has fueled efforts in whole slide image analysis. Algorithms have been developed for detection and extraction of features from whole slide images [121-149]. Data models and databases for representation and manage-

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:11

ment of microscopy image data have also been developed [Goldberg et al. 2005; Schiff- mann et al. 2006; Martone et al. 2008; Martone et al. 2003; Hayden et al. 2007; Wang et al. 2010][150-155]. ITK and VTK are two well-known libraries mainly designed to support Radiology images [156-158] but sequential processing can take hours for an image. Several recent research efforts, have investigated the use of GPUs, distributed memory parallel machines, and Grid computing to analyze large images and datasets [159-179]. SPIDAL will build on and complements this and will enable much larger scale investigations. The abstractions will enable interoperability of image analysis with other methods such as clustering, motif detection, complexity reduction, and net- work abstraction, leading to broader deployment and novel analyses. Related work in computer vision.

While computer vision has a 50 year history, only recently have researchers con- sidered large-scale datasets driven by the dramatic rise of online social media. For instance, papers in 3D reconstruction [Crandall et al. 2011; Agarwal et al. 2009; Cran- dall et al. 2011; Crandall and Snavely 2012], visual geolocation [Li et al. 2009; Hays and Efros 2008], denoising [Hays and Efros 2007], scene classification [Xiao et al. ], and object recognition [University ] use millions of images from Flickr. While there are established libraries for single machine computer vision, like OpenCV, Matlab, CImg, VLFeat, and ImageJ, there are not widely adopted libraries for large-scale computer vision, leading researchers to roll their own ad-hoc libraries (although some nascent parallel libraries are under development, like CloudCV (Virginia Tech) and Hadoop Image Processing Interface (Virginia)).

In SPIDAL, we will provide a generic high-performance framework that frees vi- sion researchers from considering systems-level issues. Core image processing library. Our core image processing library encapsulates common operations employed by com- puter vision, pathology informatics, spatial informatics, and radar informatics. These routines include low-level image preprocessing like convolution, edge detection, color correction; core primitive object detection (e.g. Hough transform) and segmentation (MinCuts); low-level feature extracting [Lowe 2004; Dalal and Triggs 2005; Torralba et al. ]); registering images with 3D models; object matching; and 3D feature extrac- tion.

The library also includes derived routines for specific domains. For example, in pathology and spatial informatics, whole slide images are extremely large (10 billion pixels per image) and cannot fit in memory requiring tiling as described above. This is not the case in computer vision, where images are typically from consumer cameras and the challenge is number of images and not their individual size. We will provide a generic parallel image processing pipeline that supports different image analysis al- gorithms with HPC-ABDS. For large images, we provide tile-based partitioning with buffered tiles at boundaries. Tiled images can be processed independently in paral- lel for image pre-processing, segmentation, and feature extraction. The results from tiles are converted into vector-based geometric shapes and structural features with correction for boundary object normalization. For 3D images, one performs 3D object grouping from 2D segmentation results, and 3D feature extraction, which again can be run in parallel.

Instead of re-implementing this code from scratch, we will port high-quality imple- mentations from open-source libraries like VLFeat, CImg and OpenCV, as well as our own existing code [160-162, 189] and for 3D [Mosaliganti et al. 2008; Pan et al. 2006; Sharp et al. ; Wenzel et al. 2007; Pan 2004; Mosaliganti et al. 2005]. These generic routines benefit multiple application domains.

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:12 G. Fox et al.

* 1. **Existing General Libraries and Machine Learning**

Mahout [Apache Mahout 2012] is an open source project that builds machine learning libraries based on Hadoop MapReduce. It supports supervised and unsupervised learn- ing algorithms such as Collaborative Filtering, Kmeans Clustering, Latent Dirichlet Allocation, and Naive Bayes Classifiers. While it offers fast implementations for some algorithms like Na”ive Bayes, for problems with iteration (like clustering) it sacri- fices performance for scalability by writing intermediate data out to disk after each iteration. For several algorithms, like Hidden Markov Models, Multilayer Perceptron, and Logistic Regression, it does not provide any parallel implementation at all (just a single-machine version).

We have studied Mahout [Apache Mahout 2012] carefully, and have chosen key algo- rithms to include in SPIDAL, including SVD, Random Forests, SVMs, Kmeans, HMMs, and LDA. For some of these, either IU or Mahout already has parallelized code, which we will simply port to HPC-ABDS to make them available in our framework. Other Mahout algorithms are either not parallelized or require iteration (which in Hadoop requires costly disk I/O); for these, HPC-ABDS will offer much higher-performance implementations.

The Spark MLLib (MLbase) library [MLLib 2014] is another promising effort, offer- ing a few algorithms (SVM, Kmeans, gradient descent) supported by a small commu- nity. We will monitor developments in that library as it matures and port from there if possible. We will also port clustering routines from R [pro 2012] to test our ability to move algorithms from this important library to HPC-ABDS.

Parallelized implementations from IU include clustering with a sophisticated de- terministic annealing (DA) method which was originally developed with Rose [Rose 1998; Rose et al. 1990] but with recent parallel implementations by Fox [Fox 2013; Fox et al. 2009a]. These codes cover the cases of both metric and non-metric spaces and have been used extensively in bioinformatics [Qiu et al. 2011; Fox et al. 2009b; Ruan et al. 2012a; Stanberry et al. 2012; Ruan et al. 2012b; Qiu et al. 2010; Qiu et al. 2010; Hughes et al. 2012; Ruan and Fox 2013; Ruan et al. 2014; Fruhwirth et al. 2011]. These will build upon [Choi 2012], wherein developed DA improvements to mixture model approaches to find hidden factors and we will also implement this in HPC-ABDS. This was successfully applied to find the context of HPC jobs and improve pre-fetching [Choi et al. 2012].

The non-metric case uses Multi-Dimensional Scaling (MDS) using either SMACOF improved with DA [Ruan and Fox 2013; Bae 2012], or a method using nonlinear 2 minimization [Kearsley et al. 1995] which is typically slower but allows a more flex- ible weighted objective function. There are several methods for dimension reduction for metric spaces and we will deploy in SPIDAL, e.g., a parallel DA enhanced GTM (Generative Topographic Mapping) from [Choi 2012; Choi et al. 2011].

1. **MIDDLEWARE FOR DATA-INTENSIVE ANALYTICS AND SCIENCE (MIDAS)**

\*\*\*why Midas?

Midas provides the underlying resource management middleware and heteroge-

neous infrastructure access layer which will support SPIDAL libraries to work effi- ciently across a range of resources. The aim of MIDAS is to provide a scalable runtime system and appropriate resource-management abstractions enabling Spidal and high-

performance, interoperable Big Data applications. For this purpose Midas introduces HPC extension, e. g. for scalable interprocess communication, to the Apache stack. At the same time Midas will ensure that Midas abstractions can be interoperable used across heterogeneous infrastructure, e. g. both in HPC and Apache Hadoop environ- ments.

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:13

**Execution Processing Models**

Execution- Processing Model Pleasingly Parallel

Execution- Processing Model Classical MR

Execution- Processing Model Iterative MR

Execution- Processing Model GiRAPH

**MI**ddleware for **D**ata-Intensive **A**nalytics and **S**cience (MIDAS) API

**Communication**

(MPI, RDMA, Hadoop Shuffle/Reduce, HARP Collectives, Giraph point-to-point)

Data and Communication

**In-memory Data Abstractions**

(HBase, Object Stores, In-Memory, other NoSQL stores, Spatial)

Higher- Level Scheduling Task Execution

schedule compute units to pilot compute

**Higher-Level Workload Management** (TEZ, LLama)

**Workload Management**

(e.g. Condor)

**Framework specific Scheduling** (e.g. YARN)

Pilot (Data and Compute) Layer Application-level Scheduiling

Pilot-based Resource Management

XSEDE

Future Grid

Azure

Amazon

OSG

Fig. 5. MIDAS: Layered Architecture View



MIDAS is based on abstractions in the areas: a) Software defined System, b) Stor- age layer including a spatial access abstraction, c) Scheduling layer using advances in multi-level and application-level scheduling, d) Collective layer that permits Map Col- lective generalization of Iterative MapReduce e) Parallelism or Programming model which generalizes the popular Giraph and MPI SPMD models.

Figure 5 shows the architecture of Midas. There are two primary design objectives of MIDAS: (i) Provide high-level abstractions (e.g. scalable data processing, inter-process communication and storage supporting both query and analysis), so as to hide de- tails of different lower level implementations (e.g. for accessing data or resources via HPC schedulers such as SLURM, Big Data schedulers, such as YARN [Hortonworks 2014] or Cloud backends like Amazon and Google), (ii) provide a flexible middleware to support four key programming models identified in Figure 3: (PM1) Pleasingly par- allel, (PM2) Search using Classic MapReduce, (PM3) Iterative MapReduce with Col- lectives and (PM4) Iterative Graph Processing. These programming models require different runtime systems to extract performance on different platforms varying from clouds to HPC. Using constructs identified in this section, MIDAS will provide the scal- able runtime system to support these programming models via appropriate execution- processing capabilities on different platforms.

To support these programming models, we propose two abstractions: an inter- process communication layer and a layer handling computation, offering, for exam- ple, different data abstractions for Hadoop (key-value, dataflow) and Giraph (graph). Communication abstractions enable the coordination and exchange of data between tasks. In particular iterative MapReduce tasks need collective operations while graph processing largely needs point-to-point communication. We have already shown that using classic MPI techniques can provide a collective layer that outperforms existing (iterative) MapReduce approaches on both cloud and HPC environments [Jha et al. 2014b; Qiu and Zhang 2013; 2014]. We will expand this concept in MIDAS covering all 4 programming models and defining the HPC-ABDS MapReduce model shown in sys- tem architecture. The communication layer is designed in a pluggable, infrastructure agnostic way and can be used within Hadoop applications and HPC application, based

e. g. on the Pilot abstraction.

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:14 G. Fox et al.

In-Memory abstractions can be used to implicitly facilitate the data exchange be- tween multiple tasks, or to efficiently retain data between generations of tasks. This is essential to efficiently support iterative processing or graph processing (patterns (MP3) and (MP4)). In particular, Harp [Qiu and Zhang 2014] building on Iterative MapReduce Twister [J.Ekanayake et al. 2010] is an open source collective commu- nication library which can be plugged into Hadoop. With this plug-in, Map-Reduce jobs can be transformed into Map-Collective jobs and users can invoke efficient in- memory message passing operations such as collective communication (e.g. broadcast and group-by) and point-to-point communication directly in Map tasks. For the first time, Map-Collective brings high performance to ABDS in a clear communication ab- straction, which did not exist before in the Hadoop ecosystem. We expect Harp to equal MPI performance with straightforward optimizations. The above is illustrative of a critical design challenge that faces MIDAS: balancing performance with flexibility.

In earlier Hadoop versions, it was necessary to retrofit non-MapReduce applications, as in for pleasingly parallel processing and machine learning, into the rigid MapRe- duce programming model. This is not necessary anymore for Hadoop-2 (YARN) [Hor- tonworks 2014], as YARN can co-locate and run any application MPI, MapReduce, iterative MR, graph, or based on any other library/framework on a Hadoop cluster. Ensuring support for YARN on XSEDE platforms will thus be an important objective. However, while YARN removes the programming model limitation, various challenges remain: (i) the current usability of YARN on HPC environments is limited (as it was de- signed as a system-level framework), (ii) resource containers cannot easily be re-used across multiple tasks, and (iii) interoperability with HPC environments (Blue Waters, to a variety of XSEDE resources). To address these limitations, we propose to build upon and extend the Pilot-abstraction, which is a proven approach in supporting data- parallel tasks on top of heterogeneous infrastructures [Luckow et al. 2012; Luckow et al. 2014]. Using Pilot-Jobs, we allow processing frameworks to interoperably exploit YARN, common HPC and cloud resource managers. Further, we will extend the Pilot- Data abstraction to support in-memory processing required for iterative MapReduce in an infrastructure and data source agnostic way.

Looking at data source aspects of the Ogres (big data patterns), most applications use a collection of files but we expect a growing interest in object stores and will sup- port both with the Pilot-Data abstraction. Several search problems need customizable dynamic indices. Also, relevant is in-memory data storage, data format, data partition- ing and the access model. Thus, we will extend canonical abstraction of Pilot-Jobs to Pilot-Data to support interoperable access to data stored in Hadoop, databases, and NOSQL as well as other types of HPC storage (Lustre, iRODS, etc.) as needed.

In addition to data access, MIDAS will support the common data processing pat- terns (partition, process, merge, etc.) on top of datasets managed by Pilot-Data. As demonstrated by the evolution to YARN, there is an increasing need to provide appli- cation specific scheduling and resource management control; the Pilot abstraction has demonstrated such capabilities. Application-level scheduling as provided by the Pilot- layer will be an essential tool to integrate library resource usage modes (see processing patterns) with resource allocation/usage. This will be required for the network science community, where flexible integration with a logical resource partitioner is important. For example, we will have to examine the trade-offs of partitioning at the MPI level or at the Giraph programming model depending upon data size and architecture. In gen- eral, MIDAS is intended to be agnostic to workflow models and will support gateways, scripting and specific workflow tools. Auto-tuning (akin to ATLAS for linear algebra software) and load balancing for each execution processing model to optimize data lo- cality and communication is a critical capability to provide to analytical application. Last but not least, the infrastructure access layer will use the existing capabilities

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:15

of SAGA, which provides most resource access functionality using a standards-based layer. This will ensure that the new software footprint of MIDAS will be small, whilst providing reuse of established and existing building blocks.

1. **DISCUSSION AND CONCLUSION**

As alluded to in the title, HPC-ABDS is the first attempt led by the Indiana-Rutgers collaboration at creating a software system to meet Big data analytics, i.e., a high- performance Big Data Stack (HPBDS). There are several incipient and emerging ef- forts at addressing the requirements for big data on high-end systems. HPC-ABDS is distinguished by its efforts to address the requirements of a well defined set of ap- plication patterns and their primary characteristics (referred to as Ogres), that are commonly required of data-analytics applications, via the construction of libraries that work across a range of high-end computing platforms, including clouds. We believe the design of such SPIDAL libraries – based upon an analysis of Big Data Ogres, as op- posed to a limit set of applications – along with a careful co-design of the resource management MIDAS layer makes HPC-ABDS distinctive.

In addition to the NIST use cases, HPC-ABDS will enable multiple many other com- munities to use a core set of libraries SPIDAL – graphs, imaging, spatial, remote sensing, HPC simulation and modeling – on top of an abstractions-based (Job, data, communication, in-memory operations, parallelism) high performance middleware MI- DAS. By integrating the two fundamental building blocks – MIDAS and SPIDAL, HPC-ABDS will provide new levels of scalability, application performance along mul- tiple dimensions by combining application expertise, the broad Apache Big Data stack and best practice HPC. Specifically, HPC-ABDS will, (i) implement resource manage- ment capabilities of MIDAS using job and data abstractions, as well as scalable and reusable fine-grained building blocks for HPC/high-end resources that realize com- munication and in-memory abstractions, (ii) define library interfaces (SPIDAL, with appropriate language bindings) to these fine-grained building blocks, (iii) provide scal- able and parallel analytic libraries by integrating the aforementioned interfaces, the fine-grained building blocks with the resource management middleware capabilities of MIDAS.

SPIDAL libraries will be provided with a set of simple benchmark kernels. The ini- tial focus will be on data analytics libraries and needed abstractions built on a skeletal MIDAS middleware whose key features have already been demonstrated.

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:16 G. Fox et al.

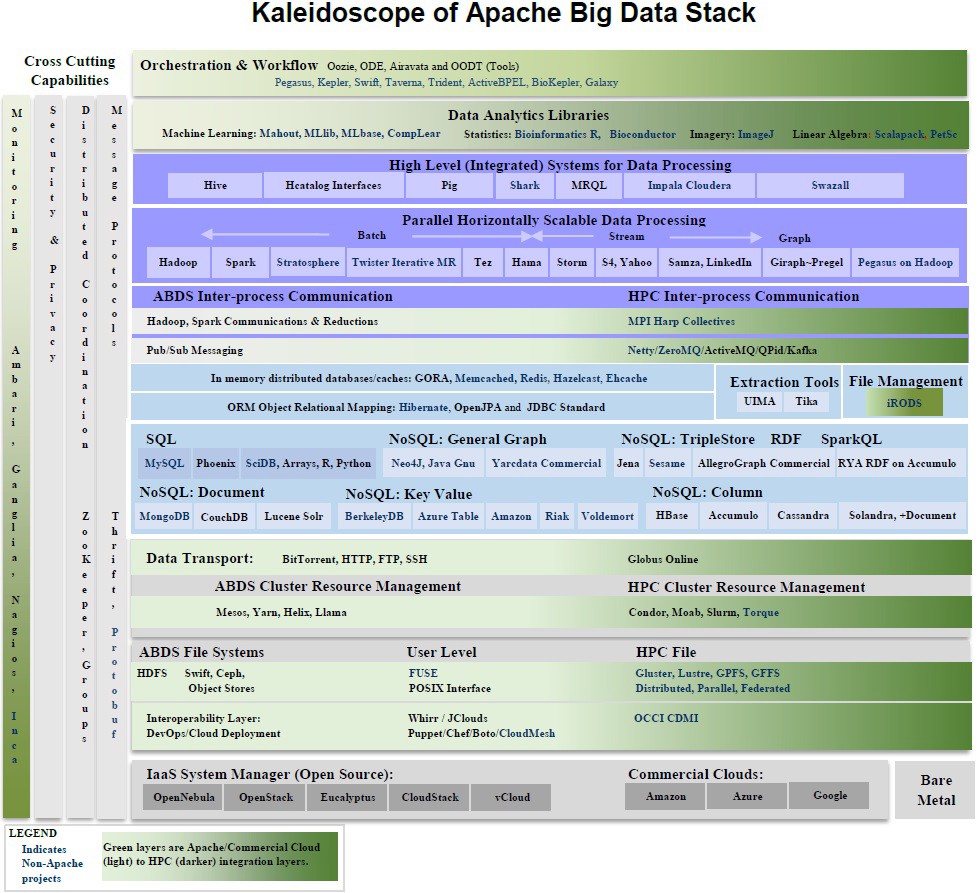


Fig. 6. For updated figure see: <http://hpc-abds.org/kaleidoscope/>

1. **APPENDIX REFERENCES**

Delivering Location Intelligence with Spatial Data. (????). <http://www.microsoft.com/sql/techinfo/> whitepapers/spatialdata.mspx

2007. Database Partitioning,Table Partitioning, and MDC for DB2 9. (2007). <http://www.redbooks.ibm.com/> redbooks/pdfs/sg247467.pdf

2012. CRAN Task View: Cluster Analysis and Finite Mixture Models. (July 21 2012). [http://cran.cnr.berkeley.](http://cran.cnr.berkeley/) edu/web/views/Cluster.html

2013a. ArcGIS. (2013). <http://www.esri.com/software/arcgis>

2013b. Greenplum. (2013). <http://www.greenplum.com/products/greenplum-database> 2013c. Hadoop-GIS Wiki. (2013). https://web.cci.emory.edu/confluence/display/HadoopGIS 2013d. IBM DB2 Spatial. (2013). <http://www-01.ibm.com/software/data/spatial/>

2013e. IBM Netezza. (2013). <http://www-01.ibm.com/software/data/netezza/>

2013f. Oracle Spatial and Oracle Locator. (2013). <http://www.oracle.com/us/products/database/options/> spatial/overview/index.html

2013g. PostGIS. (2013). <http://postgis.refractions.net/>

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:17

2013h. Teradata. (2013). <http://www.teradata.com/>

Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz, and Richard Szeliski. 2009. Building Rome in a day. (Sept. 29-Oct. 2 2009). DOI:<http://dx.doi.org/10.1109/ICCV.2009.5459148>

Ablimit Aji, Fusheng Wang, and Joel Saltz. 2012. Towards Building a High Performance Spatial Query System for Large Scale Medical Imaging Data. (2012).

Ablimit Aji, Fusheng Wang, Hoang Vo, Rubao Lee, Qiaoling Liu, Xiaodong Zhang, and Joel Saltz. 2013.

Hadoop-GIS: A High Performance Spatial Data Warehousing System Over MapReduce. *Proc. VLDB Endow.* 6, 11 (2013), 1009–1020.

Afsin Akdogan, Ugur Demiryurek, Farnoush Banaei-Kashani, and Cyrus Shahabi. 2010. Voronoi-Based Geospatial Query Processing with MapReduce. (2010).

Maksudul Alam, Maleq Khan, and Madhav V. Marathe. 2013. Distributed-memory parallel algo- rithms for generating massive scale-free networks using preferential attachment model. (2013). DOI:<http://dx.doi.org/10.1145/2503210.2503291>

N. Alon, P. Dao, I. Hajirasouliha, F. Hormozdiari, and S. Sahinalp. 2008. Biomolecular network motif count- ing and discovery by color coding. *Bioinformatics* 24, 13 (2008), i241.

Apache Mahout. 2012. Apache Mahout Scalable machine learning and data mining. [http://mahout.apache.](http://mahout.apache/) org/. (2012).

V. Aravind and V. Raman. 2002. Approximate counting of small subgraphs of bounded treewidth and related problems. (2002).

Shaikh Arifuzzaman, Maleq Khan, and Madhav Marathe. 2013. PATRIC: a parallel algorithm for counting triangles in massive networks. (2013). DOI:<http://dx.doi.org/10.1145/2505515.2505545>

D.A. Bader. 2010. Analyzing Massive Social Networks using Multicore and Multithreaded Architectures.

*Facing the Multicore-Challenge: Aspects of New Paradigms and Technologies in Parallel Computing, Lecture Notes in Computer Science* 6310, 1 (2010).

D.A. Bader and G. Cong. 2004. Fast Shared-Memory Algorithms for Computing the Minimum Spanning Forest of Sparse Graphs. (April 26-30 2004).

D.A. Bader and K. Madduri. 2008. A Graph-Theoretic Analysis of the Human Protein-Interaction Network Using Multi-core Parallel Algorithms. *Parallel Comput.* 34, 11 (2008), 627–639.

David A. Bader and Guojing Cong. 2005. A fast, parallel spanning tree algorithm for sym- metric multiprocessors (SMPs). *J. Parallel Distrib. Comput.* 65, 9 (2005), 994–1006. DOI:<http://dx.doi.org/10.1016/j.jpdc.2005.03.011>

D. A. Bader and . J. JJ. 1996. Parallel Algorithms for Image Histogramming and Connected Components with an Experimental Study. *J. Parallel and Distrib. Comput.* 35, 2 (1996), 173–190.

Seung-Hee Bae. 2012. *SCALABLE HIGH PERFORMANCE MULTIDIMENSIONAL SCALING*. Thesis.

<http://grids.ucs.indiana.edu/ptliupages/publications/SeungheeBae>Dissertation.pdf

Jaime Ballesteros, Ariel Cary, and Naphtali Rishe. 2011. SpSJoin: Parallel Spatial Similarity Joins. (2011).

DOI:<http://dx.doi.org/10.1145/2093973.2094054>

R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and

H. Van der Vorst. 1994. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition*. SIAM, Philadelphia, PA.

V. Batagelj and A. Mrvar. 1998. Pajek - Program for Large Network Analysis. *Connections* 21, 2 (1998), 47–57. <http://vlado.fmf.uni-lj.si/pub/networks/doc/pajek.pdf>

Ariel Cary, Zhengguo Sun, Vagelis Hristidis, and Naphtali Rishe. 2009. Experiences on Processing Spatial Data with MapReduce. (2009). DOI:<http://dx.doi.org/10.1007/978-3-642-02279-1>24

Kang-Tsung Chang. 2003. *Introduction to Geographic Information Systems*. Prentice Hall PTR. 384 pages. Wo Chang. 2014. ISO/IEC JTC 1 Study Group on Big Data. In *1st Big Data Interoperability Framework*

*Workshop: Building Robust Big Data Ecosystem*, Vol. 2014. NIST.

Jong Youl Choi. 2012. *Unsupervised Learning Of Finite Mixture Models With Deterministic Annealing For Large-scale Data Analysis*. Thesis. [http://grids.ucs.indiana.edu/ptliupages/publications/damix.final.v1.](http://grids.ucs.indiana.edu/ptliupages/publications/damix.final.v1) pdf

Jong Youl Choi, Mohammad H. Abbasi, David Pugmire, Scott Klasky, Judy Qiu, and Geoffrey Fox. 2012.

Mining Hidden Mixture Context With ADIOS-P To Improve Predictive Pre-fetcher Accuracy. (October 8-12 2012). <http://grids.ucs.indiana.edu/ptliupages/publications/hcming(1).pdf>

Jong Youl Choi, Seung-Hee Bae, Judy Qiu, Bin Chen, and David Wild. 2011. Browsing Large Scale Cheminformatics Data with Dimension Reduction. *Concurr. Comput. : Pract. Exper.* Special Issue on ECMLS2010 (2011). <http://grids.ucs.indiana.edu/ptliupages/publications/plotviz.v6.pdf>

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:18 G. Fox et al.

David Crandall, Andrew Owens, Noah Snavely, and Daniel P. Huttenlocher. 2011. Discrete-continuous op- timization for large-scale structure from motion. (2011).

David Crandall and Noah Snavely. 2012. Modeling people and places with internet photo collections. *Com- mun. ACM* 55, 6 (2012), 52–60. DOI:<http://dx.doi.org/10.1145/2184319.2184336>

Navneet Dalal and Bill Triggs. 2005. Histograms of Oriented Gradients for Human Detection. (2005).

DOI:<http://dx.doi.org/10.1109/cvpr.2005.177>

P. McMillen Daniel. 2004. Geographically Weighted Regression: The Analysis of Spatially Varying Relation- ships. *American Journal of Agricultural Economics* 86, 2 (2004), 554–556. <http://ideas.repec.org/a/oup/> ajagec/v86y2004i2p554-556.html

Abhinandan S. Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: scalable online collaborative filtering. (2007). DOI:<http://dx.doi.org/10.1145/1242572.1242610>

P. Daszak. 2000. Emerging Infectious Diseases of Wildlife– Threats to Biodiversity and Human Health.

*Science* 287, 5452 (2000), 443–449. DOI:<http://dx.doi.org/10.1126/science.287.5452.443>

Rob F. Van der Wijngaart, Srinivas Sridharan, and Victor W. Lee. 2012. Extending the BT NAS parallel benchmark to exascale computing. (2012).

D. Ediger, K. Jiang, E.J. Riedy, and D.A. Bader. 2012. GraphCT: Multithreaded Algorithms for Massive Graph Analysis. *’IEEE Transactions on Parallel & Distributed Systems* (2012).

Jaliya Ekanayake, Thilina Gunarathne, Judy Qiu, Geoffrey Fox, Scott Beason, Jong Youl Choi, Yang Ruan, Seung-Hee Bae, and Hui Li. 2010. *Applicability of DryadLINQ to Scientific Applications*. Report. Com- munity Grids Laboratory, Indiana University.

Ahmed Eldawy and Mohamed Mokbel. 2013. A Demonstration of SpatialHadoop: An Efficient MapReduce Framework for Spatial Data. (2013).

Christos Faloutsos. Project Pegasus Peta-Scale graph mining. (????). <http://www.cs.cmu.edu/>*∼*pegasus/

S. Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3-5 (2010), 75–174.

Geoffrey Fox. 2013. Robust Scalable Visualized Clustering in Vector and non Vector Semimetric Spaces.

*Parallel Processing Letters* 23, 2 (2013). DOI:<http://dx.doi.org/doi/abs/10.1142/S0129626413400069>

Geoffrey Fox, Seung-Hee Bae, Jaliya Ekanayake, Xiaohong Qiu, and Huapeng Yuan. 2009a. *Parallel Data Mining from Multicore to Cloudy Grids*. IOS Press, Amsterdam. <http://grids.ucs.indiana.edu/ptliupages/> publications/CetraroWriteupJune11-09.pdf

Geoffrey Fox, Xiaohong Qiu, Scott Beason, Jong Youl Choi, Mina Rho, Haixu Tang, Neil Devadasan, and Gilbert Liu. 2009b. Biomedical Case Studies in Data Intensive Computing. (December 1-4 2009). http:

//grids.ucs.indiana.edu/ptliupages/publications/SALSACloudCompaperOct10-09.pdf

Rudolf Fruhwirth, D. R. Mani, and Saumyadipta Pyne. 2011. Clustering with position-specific constraints on variance: Applying redescending M-estimators to label-free LC-MS data analysis. *BMC Bioinformatics* 12, 1 (2011), 358. <http://www.biomedcentral.com/1471-2105/12/358>

Judy Fox Geoffrey C. Fox, Shantenu Jha and Andre Luckow. 2014. Towards an Understanding of Facets and Exemplars of Big Data Applications. (October 13-14 2014).

Ashish Goel and Kamesh Munagala. 2012. *Complexity Measures for Map-Reduce, and Comparison to Paral- lel Computing*. Report. <http://www.stanford.edu/>*∼*ashishg/papers/mapreducecomplexity.pdf

I. G. Goldberg, C. Allan, J. M. Burel, D. Creager, A. Falconi, H. Hochheiser, J. Johnston, J. Mellen, P. K. Sorger, and J. R. Swedlow. 2005. The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biol* 6, 5 (2005), R47. DOI:[http://dx.doi.org/gb-2005-6-5-r47[pii];10.1186/gb-](http://dx.doi.org/gb-2005-6-5-r47)2005-6-5-r47

M. Gonen and Y. Shavitt. 2009. Approximating the number of network motifs. (2009).

O. Green, R. McColl, and D.A. Bader. 2012. A Fast Algorithm for Incremental Betweenness Centrality. (September 3-5 2012).

Danhuai Guo, Kaichao Wu, Jianhui Li, and Yuwei Wang. 2010. Spatial scene similarity assessment on Hadoop. (2010). DOI:<http://dx.doi.org/10.1145/1869692.1869700>

A. A. Hagberg, D. A. Schult, and P. J. Swart. 2008. Exploring network structure, dynamics, and function using NetworkX. (2008). <http://conference.scipy.org/proceedings/scipy2008/paper>2/

M. Handcock, D. Hunter, C. Butts, S. Goodreau, and M. Morris. 2003. statnet: Software tools for the Statis- tical Modeling of Network Data, Version 2.0. (2003). [http://statnetproject.org](http://statnetproject.org/)

Linda Hayden, Geoffrey Fox, and Prasad Gogineni. 2007. CYBERINFRASTRUCTURE FOR REMOTE SENSING OF ICE SHEETS. (June 4-8 2007). <http://grids.ucs.indiana.edu/ptliupages/publications/> TeraGrid07 paper.pdf

James Hays and Alexei A. Efros. 2007. Scene completion using millions of photographs. *ACM Trans. Graph.*

26, 3 (2007), 4. DOI:<http://dx.doi.org/10.1145/1276377.1276382>

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:19

James Hays and Alexei A. Efros. 2008. IM2GPS: Estimating Geographic Information from a Single Image. (2008).

Hortonworks. 2014. Apache Hadoop YARN is a sub-project of Hadoop introduced in Hadoop 2.0. (2014). <http://hortonworks.com/hadoop/yarn/>

Adam Hughes, Yang Ruan, Saliya Ekanayake, Seung-Hee Bae, Qunfeng Dong, Mina Rho, Judy Qiu, and Geoffrey Fox. 2012. Interpolative Multidimensional Scaling Techniques for the Identification of Clusters in Very Large Sequence Sets. *BMC Bioinformatics* 13(Suppl 2):S9, Special Issue of for Proceedings of GLBIO Great Lakes Bioinformatics Conference Ohio University Athens Ohio May 2-4 2011 (2012). DOI:<http://dx.doi.org/10.1186/1471-2105-13-S2-S9>

J.Ekanayake, H.Li, B.Zhang, T.Gunarathne, S.Bae, J.Qiu, and G.Fox. 2010. Twister: A Run- time for iterative MapReduce. (2010). <http://grids.ucs.indiana.edu/ptliupages/publications/> hpdc-camera-ready-submission.pdf

S. Jha, M. Cole, D. Katz, O. Rana, M. Parashar, and J. Weissman. 2013. Distributed Computing Practice for Large-Scale Science & Engineering Applications. *Concurrency and Computation: Practice and Experi- ence* 25, 11 (2013), 1559–1585. DOI:<http://dx.doi.org/10.1002/cpe.2897>

Shantenu Jha, Neil Chue Hong, Simon Dobson, Daniel S. Katz, Andre Luckow, Omer Rana, and Yogesh Simmhan. 2014a. Introducing Distributed Dynamic Data-intensive (D3) Science: Understanding Appli- cations and Infrastructure. (2014).

Shantenu Jha, Judy Qiu, Andre Luckow, Pradeep Mantha, and Geoffrey C. Fox. 2014b. A Tale of Two Data- Intensive Approaches: Applications, Architectures and Infrastructure. [http://arxiv.org/abs/1403.1528.](http://arxiv.org/abs/1403.1528) (June 27- July 2 2014).

K. Jiang, D. Ediger, and D.A. Bader. 2009. Generalizing k-Betweenness Centrality Using Short Paths and a Parallel Multithreaded Implementation. (September 22-25 2009).

U. Kang, C. Tsourakakis, A. Appel, C. Faloutsos, and J. Leskovec. 2008. *HADI: Fast diameter estimation and mining in massive graphs with hadoop*. Report. <http://ra.adm.cs.cmu.edu/anon/anon/home/ftp/ml2008/> CMU-ML-08-117.pdf

U Kang, Charalampos Tsourakakis, and Christos Faloutsos. 2009. PEGASUS: A Peta-Scale Graph Mining System - Implementation and Observations. (December 2009).

U Kang, Charalampos E. Tsourakakis, Ana Paula Appel, Christos Faloutsos, and Jure Leskovec. 2011a.

HADI: Mining radii of large graphs. (2011).

1. Kang, Charalampos E. Tsourakakis, Ana Paula Appel, Christos Faloutsos, and Jure Leskovec.

2011b. HADI: Mining Radii of Large Graphs. *ACM Trans. Knowl. Discov. Data* 5, 2 (2011), 1–24.

DOI:<http://dx.doi.org/10.1145/1921632.1921634>

Howard Karloff, Siddharth Suri, and Sergei Vassilvitskii. 2010. A model of computation for MapReduce. (2010).

Anthony J. Kearsley, Richard A. Tapia, and Michael W. Trosset. 1995. *The Solution of the Metric STRESS and SSTRESS Problems in Multidimensional Scaling Using Newtons Method*. Report. Rice University.

Hartmut Klauck, Danupon Nanongkai, Gopal Pandurangan, and Peter Robinson. 2013. *The Distributed Complexity of Large-scale Graph Processing*. Report. <http://arxiv.org/abs/1311.6209>

J. Leskovec. Stanford Network Analysis Project. (????). <http://snap.stanford.edu/>

Wengen Li, Weili Wang, and Ting Jin. 2012. *Evaluating Spatial Keyword Queries under the MapReduce Framework Database Systems for Advanced Applications*. Lecture Notes in Computer Science, Vol. 7240. Springer Berlin / Heidelberg, 251–261. DOI:<http://dx.doi.org/10.1007/978-3-642-29023-7>26

Yunpeng Li, David Crandall, and Daniel P. Huttenlocher. 2009. Landmark Classification in Large-scale Image Collections. (2009).

Yan Liu, Kaichao Wu, Shaowen Wang, Yanli Zhao, and Qian Huang. 2010. A MapReduce Approach to Gi\*(d) Spatial Statistic. (2010). DOI:<http://dx.doi.org/10.1145/1869692.1869695>

Paul A. Longley, Mike Goodchild, David J. Maguire, and David W. Rhind. 2010. *Geographic Information Systems and Science* (third ed.). Wiley.

David G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision* 60, 2 (2004), 91–110. DOI:http://dx.doi.org/10.1023/b:visi.0000029664.99615.94

Andre Luckow, Mark Santcroos, and Shantenu Jha. 2014. Pilot-Data: An Abstraction for Distributed Data.

*J. Parallel and Distrib. Comput.* In press (2014). <http://arXiv.org/abs/1301.6228>

Andre Luckow, Mark Santcroos, Ole Weidner, Andre Merzky, Pradeep Mantha, and Shantenu Jha. 2012.

P\*: A Model of Pilot-Abstractions. (2012). DOI:<http://dx.doi.org/10.1109/eScience.2012.6404423>

Qiang Ma, Bin Yang, Weining Qian, and Aoying Zhou. 2009. Query Processing of Massive Trajectory Data Based on Mapreduce. (2009). DOI:<http://dx.doi.org/10.1145/1651263.1651266>

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:20 G. Fox et al.

K. Madduri and D.A. Bader. 2009. Compact Graph Representations and Parallel Connectivity Algorithms for Massive Dynamic Network Analysis. (May 25-29 2009).

K. Madduri, D.A. Bader, J.W. Berry, and J.R. Crobak. 2007. An Experimental Study of A Parallel Shortest Path Algorithm for Solving Large-Scale Graph Instances. (January 6 2007).

Kamesh Madduri, David Ediger, Karl Jiang, David A. Bader, and Daniel Chavarria-Miranda. 2009. A faster parallel algorithm and efficient multithreaded implementations for evaluating betweenness centrality on massive datasets. (2009). DOI:<http://dx.doi.org/10.1109/ipdps.2009.5161100>

M. E. Martone, J. Tran, W. W. Wong, J. Sargis, L. Fong, S. Larson, S. P. Lamont, A. Gupta, and M. H. Ellisman. 2008. The cell centered database project: an update on building commu- nity resources for managing and sharing 3D imaging data. *J Struct Biol* 161, 3 (2008), 220–31. DOI:[http://dx.doi.org/S1047-8477(07)00254-7[pii];10.1016/j.jsb](http://dx.doi.org/S1047-8477(07)00254-7).2007.10.003

M. E. Martone, S. Zhang, A. Gupta, X. Qian, H. He, D. L. Price, M. Wong, S. Santini, and M. H. Ellisman. 2003. The cell-centered database: a database for multiscale struc- tural and protein localization data from light and electron microscopy. 1, 4 (2003), 379–95. DOI:http://dx.doi.org/NI:1:4:379[pii];10.1385/NI:1:4:379

Manish Mehta and David J. DeWitt. 1995. Managing Intra-operator Parallelism in Parallel Database Sys- tems. (1995).

Henning Meyerhenke, David Ediger, and David A. Bader. 2011. Parallel Community Detection for Massive Graphs. (September 2011).

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. Network motifs: simple building blocks of complex networks. *Science* 298, 5594 (2002), 824.

MLLib. 2014. Machine Learning Library (MLlib). [http://spark.apache.org/docs/0.9.0/mllib-guide.html.](http://spark.apache.org/docs/0.9.0/mllib-guide.html) (2014).

Kishore Mosaliganti, Tony Pan, Raghu Machiraju, Kun Huang, and Joel Saltz. 2005. ITK-based Registration of Large Images from Light Microscopy: A Biomedical Application. *The Insight Journal - 2005 MICCAI Open-Source Workshop* (2005).

K. Mosaliganti, T. Pan, R. Ridgway, R. Sharp, L. Cooper, A. Gulacy, A. Sharma, O. Irfanoglu, R. Machiraju,

T. Kurc, A. de Bruin, P. Wenzel, G. Leone, J. Saltz, and K. Huang. 2008. An imaging workflow for characterizing phenotypical change in large histological mouse model datasets. *J Biomed Inform* 41 (2008), 863–73.

A. Lumsdaine N. Edmonds, T. Hoefler. 2010. A space-efficient parallel algorithm for computing betweenness centrality in distributed memory. (2010).

Donald Nguyen, Andrew Lenharth, and Keshav Pingali. 2013. A lightweight infrastructure for graph ana- lytics. (2013). DOI:<http://dx.doi.org/10.1145/2517349.2522739>

NIST. 2013a. Big Data Initiative Reports from V1. <http://bigdatawg.nist.gov/V1>output docs.php. (2013). NIST. 2013b. NIST Big Data Public Working Group (NBD-PWG) Use Cases and Requirements. http:

//bigdatawg.nist.gov/usecases.php, (2013).

T Pan, S Jewel, U Catalyurek, P Wenzel, G Leone, S Hastings, S Oster, S Langella, T Kurc, and J Saltz.

2006. *Virtual Microscopy: Distributed Image Storage, Retrieval, Analysis, and Visualization*. John Wiley

& Sons, Honoken, NJ, 737–763.

Tony C Pan. 2004. *Three Dimensional Reconstruction of Microscopy Images*. Report. BMI Technical Report: Osubmi tr 2004 n09, The Ohio State University, Department of Biomedical Informatics.

Jignesh Patel, JieBing Yu, Navin Kabra, Kristin Tufte, Biswadeep Nag, Josef Burger, Nancy Hall, Karthikeyan Ramasamy, Roger Lueder, Curt Ellmann, Jim Kupsch, Shelly Guo, Jo- han Larson, David De Witt, and Jeffrey Naughton. 1997. Building A Scaleable Geo-Spatial DBMS: Technology, Implementation and Evaluation. *SIGMOD Rec.* 26, 2 (1997), 336–347. DOI:<http://dx.doi.org/10.1145/253262.253342>

A. Pavlo, E. Paulson, A. Rasin, D. J. Abadi, D. J. DeWitt, S. Madden, and M. Stonebraker. 2009. A Compari- son of Approaches to Large-Scale Data Analysis. (2009).

Alexei Pozdnoukhov and Christian Kaiser. 2011. Scalable Local Regression for Spatial Analytics. (2011).

DOI:<http://dx.doi.org/10.1145/2093973.2094023>

Dimitrios Prountzos and Keshav Pingali. 2013. Betweenness centrality: algorithms and implementations.

*SIGPLAN Not.* 48, 8 (2013), 35–46. DOI:<http://dx.doi.org/10.1145/2517327.2442521>

G. Qin and L. Gao. 2012. An algorithm for network motif discovery in biological networks. *International Journal of Data Mining and Bioinformatics* 6, 1 (2012), 1–16.

Judy Qiu, Jaliya Ekanayake, Thilina Gunarathne, Jong Youl Choi, Seung-Hee Bae, Hui Li, Bingjing Zhang, Tak-Lon Wu, Yang Ryan, Saliya Ekanayake, Adam Hughes, and Geoffrey Fox. 2010. Hybrid cloud and cluster computing paradigms for life science applications. *BMC Bioin-*

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

Towards HPC-ABDS: An Initial High-Performance BigData Stack 1:21

*formatics* Proceedings of BOSC 2010 (2010). <http://grids.ucs.indiana.edu/ptliupages/publications/> HybridCloudandClusterComputingParadigmsforLifeScienceApplications Pub.pdf

Judy Qiu, Jaliya Ekanayake, Thilina Gunarathne, Jong Youl Choi, Seung-Hee Bae, Yang Ruan, Saliya Ekanayake, Stephen Wu, Scott Beason, Geoffrey Fox, Mina Rho, and Haixu Tang. 2011. *Data Inten- sive Computing for Bioinformatics*. IGI Publishers. DOI:<http://dx.doi.org/10.4018/978-1-6152971-2>

Judy Qiu, Thilina Gunarathne, Jaliya Ekanayake, Jong Youl Choi, Seung-Hee Bae, Hui Li, Bingjing Zhang, Yang Ryan, Saliya Ekanayake, Tak-Lon Wu, Scott Beason, Adam Hughes, and Geoffrey Fox. 2010. Hybrid Cloud and Cluster Computing Paradigms for Life Sci- ence Applications. (July 9-10 2010). <http://grids.ucs.indiana.edu/ptliupages/publications/> HybridCloudandClusterComputingParadigmsforLifeScienceApplications.pdf

Judy Qiu and Bingjing Zhang. 2013. *Clustering Social Images with MapReduce and High Per- formance Collective Communication*. IOS Press. <http://grids.ucs.indiana.edu/ptliupages/publications/> MammothDataintheCloudClusteringSocialImage.pdf

Judy Qiu and Bingjing Zhang. 2014. Harp: a runtime for efficient in-memory communication. (2014). http:

//salsaproj.indiana.edu/harp/

R Project. 2012. R open source statistical library. [http://www.r-project.org/.](http://www.r-project.org/) (2012).

P. Ribeiro, F. Silva, and L. Lopes. 2012. Parallel discovery of network motifs. *J. Parallel and Distrib. Comput.*

72, 2 (2012), 144–154.

Ken Rose. 1998. Deterministic Annealing for Clustering, Compression, Classification, Regression, and Re- lated Optimization Problems. *Proc. IEEE* 86 (1998), 2210–2239.

Ken Rose, Eitan Gurewitz, and Geoffrey Fox. 1990. A deterministic annealing approach to clustering. *Pattern Recogn. Lett.* 11 (1990), 589–594.

Yang Ruan, Saliya Ekanayake, Mina Rho, Haixu Tang, Seung-Hee Bae, Judy Qiu, and Geoffrey Fox.

2012a. DACIDR: Deterministic Annealed Clustering with Interpolative Dimension Reduction using Large Collection of 16S rRNA Sequences. (October 7-10 2012). <http://grids.ucs.indiana.edu/ptliupages/> publications/DACIDR camera ready v0.3.pdf

Yang Ruan and Geoffrey Fox. 2013. A Robust and Scalable Solution for Interpolative Multidimensional Scaling with Weighting. (October 22-25 2013). DOI:<http://dx.doi.org/10.1109/eScience.2013.30>

Yang Ruan, Zhenhua Guo, Yuduo Zhou, Judy Qiu, and Geoffrey Fox. 2012b. HyMR: a Hybrid MapRe- duce Workflow System. (June 18 2012). <http://grids.ucs.indiana.edu/ptliupages/publications/HyMR> submission HPDC workshop final.pdf

Yang Ruan, Geoffrey L. House, Saliya Ekanayake, Ursel Schtte, James D. Bever, Haixu Tang, and Geoffrey Fox. 2014. Integration of Clustering and Multidimensional Scaling to Determine Phylogenetic Trees as Spherical Phylograms Visualized in 3 Dimensions. (May 26-29 2014). <http://grids.ucs.indiana.edu/> ptliupages/publications/PhylogeneticTreeDisplayWithClustering.pdf

S. Sarkar and A. Dong. 2011. Community detection in graphs using singular value decomposition. *Physical Review E* 83, 4 (2011), 04611.

1. Satuluri and S. Parthasarathy. 2009. Scalable graph clustering using stochastic ows: applications to com- munity discovery. (2009).

Venu Satuluri, Srinivasan Parthasarathy, and Yiye Ruan. 2011. Local graph sparsification for scalable clus- tering. (2011). DOI:<http://dx.doi.org/10.1145/1989323.1989399>

D. A. Schiffmann, D. Dikovskaya, P. L. Appleton, I. P. Newton, D. A. Creager, C. Allan, I. S. Nathke, and I. G. Goldberg. 2006. Open microscopy environment and findspots: integrating image infor- matics with quantitative multidimensional image analysis. *Biotechniques* 41, 2 (2006), 199–208. DOI:[http://dx.doi.org/000112224[pii]](http://dx.doi.org/000112224)

C. Seshadhri, T. Kolda, and A. Pinar. 2012a. Community structure and scale-free collections of Erdos–Renyi graphs. *Physical Review E* (2012). <http://arxiv.org/abs/1112.3644>

C. Seshadhri, A. Pinar, and T. Kolda. 2012b. Fast Triangle Counting through Wedge Sampling. (2012). http:

//arxiv.org/abs/1202.5230

Richard Sharp, Randall Ridgway, Kishore Mosaliganti, Okan Irfanoglu, Pamela Wenzel, Raghu Machiraju, Alain de Bruin, Gustavo Leone, Tony Pan, and Kun Huang. Examining Phenotype Differences in Mouse Placenta with Volume Rendering and Segmentation. In *Life Science Systems and Applications Work- shop, 2006. IEEE/NLM*. IEEE, 1–2.

Larissa Stanberry, Roger Higdon, Winston Haynes, Natali Kolker, William Broomall, Saliya Ekanayake, Yang Ruan, Judy Qiu, Eugene Kolker, Geoffrey Fox, and Adam Hughes. 2012. Visualizing the Protein Sequence Universe. (June 18 2012). <http://grids.ucs.indiana.edu/ptliupages/publications/paperDelft> final.pdf

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.

1:22 G. Fox et al.

Siddharth Suri and Sergei Vassilvitskii. 2011. Counting triangles and the curse of the last reducer. (2011).

DOI:<http://dx.doi.org/10.1145/1963405.1963491>

Antonio Torralba, Kevin P Murphy, William T Freeman, and Mark A Rubin. Context-based vision system for place and object recognition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Con- ference on*. IEEE, 273–280.

Charalampos E. Tsourakakis, U Kang, Gary L. Miller, and Christos Faloutsos. 2009. DOULION: Counting triangles in massive graphs with coin. (2009).

Stanford University. ImageNet image database organized according to the WordNet hierarchy. (????). http:

[//www.image-net.org/](http://www.image-net.org/)

Fusheng Wang, Jun Kong, Jingjing Gao, Lee A.D. Cooper, Tahsin Kurc, Zhengwen Zhou, David Adler, Cristo- bal Vergara-Niedermayr, Bryan Katigbak, Daniel J Brat, and Joel H Saltz. 2013. A high-performance spatial database based approach for pathology imaging algorithm evaluation. *J Pathol Inform.* 4, 5 (2013).

F Wang, T Kurc, P Widener, T Pan, J Kong, L Cooper, D Gutman, A Sharma, S Cholleti, V Kumar, and J Saltz. 2010. High-performance Systems for In Silico Microscopy Imaging Studies. *The 7th International Conference on Data Integration in the Life Sciences, Gothenburg, Sweden* (2010).

Kai Wang, Jizhong Han, Bibo Tu, Jiao Dai, Wei Zhou, and Xuan Song. Accelerating Spatial Data Processing with MapReduce. In *IEEE International Conference on Parallel and Distributed Systems (ICPADS)*. 229–236. DOI:<http://dx.doi.org/10.1109/icpads.2010.76>

Kaibo Wang, Yin Huai, Rubao Lee, Fusheng Wang, Xiaodong Zhang, and Joel Saltz. 2012. Accelerating Pathology Image Data Cross Comparison on CPU-GPU Hybrid Systems. *Proc. VLDB Endow.* 5, 11 (2012), 1543–1554.

P. L. Wenzel, L. Wu, A. de Bruin, J. L. Chong, W. Y. Chen, G. Dureska, E. Sites, T. Pan, A. Sharma, K. Huang,

R. Ridgway, K. Mosaliganti, R. Sharp, R. Machiraju, J. Saltz, H. Yamamoto, J. C. Cross, M. L. Robinson, and G. Leone. 2007. Rb is critical in a mammalian tissue stem cell population. *Genes Dev* 21, 1 (2007), 85–97. DOI:<http://dx.doi.org/10.1101/gad.1485307>

Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba. SUN database: Large- scale scene recognition from abbey to zoo. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*. IEEE, 3485–3492.

Shubin Zhang, Jizhong Han, Zhiyong Liu, Kai Wang, and Shengzhong Feng. Spatial Queries Evaluation with MapReduce. In *Eighth International Conference on Grid and Cooperative Computing*. 287–292. DOI:<http://dx.doi.org/10.1109/gcc.2009.16>

Shubin Zhang, Jizhong Han, Zhiyong Liu, Kai Wang, and Zhiyong Xu. SJMR: Parallelizing Spatial Join with MapReduce on Clusters. In *IEEE International Conference on Cluster Computing*. 1–8. DOI:<http://dx.doi.org/10.1109/clustr.2009.5289178>

1. Zhang, Z. Wang, Y. Wang, and L. Zhou. 2009. Parallel community detection on large networks with propin- quity dynamics. (2009).

Zhao Zhao, Maleq Khan, V. S. Anil Kumar, and Madhav V. Marathe. 2010. Subgraph Enumer- ation in Large Social Contact Networks Using Parallel Color Coding and Streaming. (2010). DOI:<http://dx.doi.org/10.1109/icpp.2010.67>

1. Zhao, G. Wang, A. Butt, M. Khan, V. S. Anil Kumar, and M. Marathe. 2012. SAHAD: Subgraph Analysis in Massive Networks Using Hadoop. (May 2012).

ACM, Vol. 1, No. 1, Article 1, Publication date: August 2014.