

Confluences among Big Data, Finite Element Analysis and High Performance Computing

¹Lidong Wang, ²Guanghui Wang and ³Cheryl Ann Alexander

¹Department of Engineering Technology, Mississippi Valley State University, USA

²State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, China

³Technology and Healthcare Solutions, Inc., USA

Article history

Received: 20-05-2015

Revised: 18-06-2015

Accepted: 10-07-2015

Corresponding Author:

Lidong Wang

Department of Engineering

Technology, Mississippi Valley

State University, USA

E-mail: lwang22@students.tntech.edu

Abstract: Big Data analyzes correlations from huge raw data and predicts outcomes. It has great impacts on scientific discoveries and value creation. High Performance Computing (HPC) uses parallel processing and advanced programs or software packages to complete complicated jobs quickly. Finite Element Method (FEM) is very powerful in scientific computation and engineering analysis. It has created huge values in almost every area of engineering. In a lot of applications, Finite Element Analysis (FEA) strongly relies on advanced computer technology and HPC. Big Data will play an important role in FEA and HPC. This paper presents confluences among Big Data, FEA and HPC.

Keywords: Big Data, Finite Element Method (FEM), High Performance Computing (HPC), Big Data Analytics, Hadoop, MapReduce, Graphical Processing Unit (GPU)

Introduction

Scientific data is often on a massive scale with complexity and heterogeneity. It is often manipulated through complex and distributed workflows, application-specific (ad hoc) using low-level code libraries. Big Data technology has been expected to perform scalable query processing and scientific workflow management for scientific data (Pacitti and Valduriez, 2012).

Big data is a massive volume of both structured and unstructured data. It is so large that it is difficult to process using traditional database and software techniques (Demchenko *et al.*, 2013). Big data is often heterogeneous. Each organization tends to produce and manage its own data in specific formats and with its own processes. Big data is complicated. Its complexity lies in: Uncertain data (because of data capture), multiscale data (with lots of dimensions) and graph-based data, etc. Continuous data streams are captured (e.g., from sensors or mobile devices), which produces streaming and dynamically changing big data (Pacitti and Valduriez, 2012).

Big Data is likely to be advantageous for comparing differences in competing design options. The combination of Big Data, Artificial Intelligence (AI) and massively parallel computing offered the potential to create a revolutionary way of practicing evidence-based

and personalized medicine (Dilsizian and Siegel, 2014). Big data privacy is a sensitive issue with conceptual, legal and technological implications. Storage and I/O optimization for big-data computing is also an important issue. There is tremendous wealth of information in big data. The information is potentially valuable. High Performance Computing (HPC) can help unlock the wealth contained in big data (Jean-François Lavignon, 2013).

HPC offers immense potential for data-intensive computing. But as data explodes in volume, variety and velocity; it is getting increasingly difficult to scale compute performance (Intel, 2014). Data-intensive HPC, massive storage and file system, I/O Architecture and low-power computing and automatic cloud provisioning for HPC are interesting topics in HPC. Data movement is very expensive. Reducing data movement is critical for HPC. Data locality should be the best solution.

Finite Element Method (FEM) has been widely used in engineering. Traditionally, finite element simulations can be performed on various computers. Advanced numerical methods (e.g., multiscale computation with multiscale material models, or finite element computation with adaptive mesh refinement) can generate data with large volumes and rates. Large-scale simulation workflows can run on large

supercomputers and data are dumped on parallel disk systems (Parashar, 2014).

Data were organized internally within the Finite Element Analysis (FEA) core based on an object-oriented model. Data were represented in three basic data types: Matrix, Vector and ID (integer). For external data representation, eXtensible Markup Language (XML) was used as the standard for representing data in a platform independent manner (Peng *et al.*, 2003). However, ASCII/XML don't adapt well for highly voluminous and complex data such as large-scale finite element analysis data and heterogeneous product data. XML doesn't express entity relationships well either (Folk, 2006).

People have a growing interest in the integration of Big Data and computational mechanics such as FEA. There are algorithmic challenges of big data in huge-scale finite element computation. For example, the data is distributed in space; the requested parts of data are not always available (Nesterov, 2014). The development of finite element meshing and numerical optimization algorithms, parallel scientific computing in the HPC environment and the combination of Big Data and FEA help improve the accuracy and efficiency of FEA greatly.

Characteristics, Implementation and Technology of Big Data

Big data can be from various sources with low information density and with different structures (structured data; semi-structured document; and unstructured text, graph, image and video). It is often hard to integrate, verify and assess big data (Davenport, 2014).

Big data characteristics can be described by "6Vs". They are: Volume, Variety, Value, Velocity, Veracity and Variability (Bellini *et al.*, 2013; Demchenko *et al.*, 2013; Jean-François Lavignon, 2013; O'Leary, 2013; Jagadish *et al.*, 2014):

- *Volume*: It refers to massive amounts of data. This makes it hard to store and manage
- *Variety*: It refers to heterogeneity of data types, representation and semantic interpretation. This makes it hard to perform data integration
- *Value*: The collected data can bring added-value to the intended process, activity, or predictive analysis
- *Velocity*: Data such as highly streaming data is generated at a rate that exceeds those of traditional systems, which makes it hard to perform online processing
- *Variability*: It refers to data changes during processing and lifecycle. Big data can be constantly changing (dynamic). Thus, analysis often needs to be able to run in real time. There are challenges in dealing with highly varying data

- *Veracity*: It refers to the accuracy, truthfulness and reliability of the data. Veracity makes it hard to perform data analysis. Big data is often noisy (uncertain). Dealing with noisy big data and quantifying data uncertainty have become imperative. These rely on computationally intensive statistical and machine-learning techniques

Capgemini Consulting conducted a global survey of senior Big Data executives in November 2014. The survey covered 226 respondents across Europe, North America and Asia Pacific; spanned a number of industries including retail, manufacturing, pharmaceuticals, financial services and energy and utilities. The survey targeted senior executives across the Analytics, IT and Business functions because they were responsible for overseeing Big Data initiatives in their organization. Only 27% of the executives in the survey described their Big Data initiatives as "successful". Table 1 reveals main challenges that organizations face in Big Data implementation (Colas *et al.*, 2014).

Big Data for development is an issue of turning imperfect, complex, often unstructured data into actionable information. This implies leveraging advanced computational methods (such as machine learning) to reveal correlations and trends within and across large data sets. The intensive mining of socioeconomic data, known as "reality mining," can shed light on processes and interactions. Reality mining can be done in three main ways (Letouzé, 2012):

- Continuous data analysis over streaming data: Using tools to scrape the Web to monitor and analyze high-frequency online data streams including uncertain and inexact data
- Online digestion of semi-structured data and unstructured ones such as news items and product reviews to shed light on hot topics, perceptions, needs and wants
- Real-time correlation of streaming data (fast stream) with slowly accessible historical data repositories

Big data processing is the foundation in applications. In order to improve data processing capability, the Hadoop framework is used to achieve the distribution storage and analysis work of the collected big data (Yan *et al.*, 2014). Hadoop is a Java based framework and heterogeneous open source platform. Hadoop's primary modules are the Hadoop Distributed File System (HDFS) and MapReduce (MR). HDFS provides high throughput access to big data. MR implements a high level and implicit parallel programming model (Davenport, 2014). Hadoop can offer a number of techniques and tools in Table 2 (Eaton *et al.*, 2012; Schneider, 2012; Davenport, 2014; Raghupathi and Raghupathi, 2014):

Table 1. Key challenges for Big Data implementation

Challenges	Percentage (%)
Scattered data lying in silos across various teams	46
Absence of a clear business case for funding and implementation	39
Ineffective coordination of Big Data and analytics teams in the organization	35
Dependency on legacy systems for data processing and management	31
Ineffective governance models for Big Data and analytics	27
Lack of sponsorship from top management	27
Lack of Big Data and analytics skills	25
Lack of clarity on Big Data technology and tools	22
Cost of specific tools and infrastructure for Big Data and analytics	18
Data security and privacy concerns	15
Resistance to change within the organization	12

Table 2. Hadoop techniques, tools and their functions

Techniques and tools	Functions
HDFS	A highly fault tolerant distributed file system; responsible for storing data.
MapReduce	A powerful parallel programming technique for distributed processing.
Pig	A scripting language for describing operations like reading, filtering, transforming, joining and writing data.
Hive	A scripting language similar to Pig, but more batch oriented; being able to transform data into the relational format for structured query language (SQL) queries.
HBase	A scalable and distributed database for random read/write access.
Oozie	A workflow for dependent Hadoop jobs.
ZooKeeper	A centralized service (coordination) for providing distributed synchronization and group services.
Sqoop	A project (data exchange) for transferring data between relational databases and Hadoop.
Flume	Log collector; performing reliable and distributed streaming log collection.
Avro	A system of data serialization.
Chukwa	A Hadoop subproject as data accumulation system for monitoring distributed systems.

Big Data can extend applications by the following: (1) Robust and highly distributed file systems capable of managing Big Data applications; (2) flexible methods for handling large quantities of data in highly parallel computing environments such as MapReduce, a parallel merge-sort algorithm; (3) Structured Query Language (SQL) and Not only SQL (NoSQL) capabilities for retrieving and storing data from dynamically growing databases or data streams (Douglas, 2014).

Features, Functions and Challenges of High Performance Computing

High Performance Computing (HPC) uses parallel processing for running advanced application programs quickly and reliably. High performance computer clusters can complete different kinds of jobs. For example, optimizing a parallel genetic algorithm was implemented on a HPC architecture using clusters of computers with several Graphical Processing Units (GPUs). Clusters can be high performance computing clusters, load balancing clusters and fail-over clusters. Linux and Microsoft Azure are common operating systems used for HPC. Lustre is a POSIX-compliant and parallel system used in HPC. It is currently an open-source system that is often used in conjunction with Hadoop to provide a distributed, scalable and high

performance file system in place of HDFS. It enables HPC engines to maintain performance by aggregating multiple I/O paths to multiple servers in the compute cluster (Guillén *et al.*, 2014; Slack, 2014).

There are the following challenges (Parashar, 2014) faced by traditional HPC data pipelines:

- Scalable data analytics challenge
- I/O and storage challenge
- Increasing performance gap
Disks are outpaced by computing speed
- Data movement challenge
There is a lot of data movement between simulation and analysis as well as between coupled multi-physics simulation components, which leads to long latencies.
- Energy challenge
Much power is consumed due to the memory and data movement.

According to IBM data-centric design principles, data motion should be minimized and workflow parallelism should be increased to leverage low-power cores. Computation and analytics should be moved closer to the data. It is the best for work to be done where the data resides. It is also important to integrate in-situ analytics and in-transit analytics. Primary resources execute the

main simulation and in-situ computations; secondary resources provide a staging area whose cores act as containers for in-transit computations (Parashar, 2014).

Confluence between Big Data and Finite Element Analysis

Traditionally, Finite Element Analysis (FEA) is often conducted using only one geometric configuration, one set of boundary conditions and one set of material properties or parameters. Results only for every *n*-th time step (*n* is normally greater than 2) of analysis are stored. The computed results are often only for a set of points in the geometric design. Only part of variables (such as minimal and maximal stresses/strains) are quantitatively analyzed. In a lot of applications, there are changing geometric shapes with service time due to loads; changing material parameters and the stress-strain relationship of materials under various temperatures; and changing boundary conditions. The FEA issue should be very complicated if the combination of the above changing aspects occurs.

Statistical approaches or stochastic approaches to handling changing geometric configurations, changing boundary conditions and changing material properties are expected to get better solutions to the numerical analysis of a true physics system behavior and performance. There are analysis results with the level of Terabytes (TB) in some applications of large FEA. Big Data (with statistical methods and stochastic methods) can help FEA handle large size results and deal with highly varying conditions and properties. It is also powerful in resolving problems in large data storage, analysis and visualization. The confluence between Big Data and FEA can provide more comprehension analysis, predict defects and failures and optimize design in time.

A generalized workflow was outlined for the integration of multi-modal measurements and multi-physics models at multiple hierarchical length scales to accelerate materials development. Protocols for direct and efficient linking of materials models/databases into process/performance simulation codes (e.g., crystal plasticity finite element method) were studied. A Big Data based workflow was given for integrating

mesoscale heterogeneities in material structure with process simulation. The efforts based on Big Data were for improving the accuracy of predictive modeling (Salem *et al.*, 2014).

There will be more applications of FEA enhanced by Big Data. The following can be the challenges or driving forces: (1) FEA model sizes are increasing continuously; (2) analysis types such as nonlinear and dynamics are becoming more common; (3) software efficiency and hardware performance are increasing; (4) one of the limits to realistic model size is the ability to store and retrieve the vast amount of data generated in FEA (Abbey, 2014); (5) there are problems with sparse data in large-scale FEA and sometimes the data is changing with time; (6) hierarchical meshes through adaptive mesh refinement are often needed in huge-scale FEA for fatigue crack incubation/crack growth and fracture issues in Aeronautical and Astronautical engineering; (7) multiscale materials models (macro, micro and nano) and multiscale computation (macro FEA, crystal plasticity FEA and molecular dynamics) can provide more accurate results in some applications.

Confluence between Big Data and High Performance Computing

Traditionally, HPC was chiefly compute-intensive and not data-intensive. There have been new HPC usage models that deal with HPC workloads in cloud computing, increasing heterogeneity of data and explosion of data volumes (“big data”). Big Data Analytics (BDA) techniques and tools can handle large and diverse semi-structured and unstructured datasets, which increases the overhead of data access and must be handled in parallelized methods. Big data is by nature distributed. Therefore, distributed algorithms are key; data migration is necessary and important. The computational complexity of analyzing big data and its sheer size indicate that data-intensive computing is very important (Jean-François Lavignon, 2013).

The comparison between Big Data analytics (BDA) and High Performance Computing (HPC) is shown in Table 3 (Valduriez, 2015). The combination of HPC and BDA is sometimes called Data-intensive HPC.

Table 3. Big Data analytics (BDA) Vs. High Performance Computing (HPC)

	BDA	HPC
Computing model	Data-centric: move tasks to data and reduce	Compute-centric: move data to tasks and accelerate
Data storage	Uniform storage (sharding) on disks	Hierarchical storage (disks, tapes, etc.)
Parallel file management	Designed for few big files, e.g., Hadoop Distributed File System (HDFS)	Designed for many small files, e.g., Lustre
Programming model	Algebraic operators, e.g., Map Reduce, Spark	Message Passing Interface (MPI) versus Open Multi Processing (OpenMP)
Languages	Java, Python, C++	C, C++

As Big Data frameworks find their way into production environments, users are facing challenges of integrating scale performance, scaling and stabilizing the performance of the clusters. The growing ubiquity of Solid State Disks (SSDs), high performance CPUs and advanced network clustering enables users to build efficient clusters, scale performance and provide a best-in-class environment for Hadoop and other Big Data applications (Gutkind, 2013).

One of challenges of Big Data is to access data in an efficient way, applying massive parallelism not only for the computation, but also for the storage. Although HPC clusters are typically equipped with high-level interconnections, the real problems arise when it is necessary to transfer data between geographically distributed sites because the Internet connection might not be suitable to transfer big data. Therefore, reducing data movement or data locality is critical. In cluster computing, the data parallel approach subdivides the data to analyze among almost independent processes. It can be a suitable solution for Big Data analysis. In GPU computing, GPUs deliver extremely high floating-point performance and massive parallelism at a very low cost (Merelli *et al.*, 2014).

High Performance Data Analytics (HPDA) represents the confluence of HPC and Big Data analytics that is deployed onto HPC-style configurations. For implementing a high-performance, scalable and agile information foundation to support near real-time analytics and HPDA capabilities, users can use Hadoop with HPC infrastructure to reduce the processing time of the growing volumes of data that are common in today's distributed computing environments. There is a pressing need for a Hadoop platform that enables Big Data analytics applications to process data stored on HPC clusters. Ideally, users should be able to run Hadoop workloads like running any other HPC workload with the same manageability and performance. This can only happen when Hadoop is tightly integrated with the file systems and schedulers that have long served HPC environments (Intel, 2014).

Confluence between Finite Element Analysis and High Performance Computing

In some applications, a simulation using FEA can produce over 50 Terabytes (TB) of data. Not all the data and results produced by the simulation are stored; results of several time steps are finally deleted from storage device. An architecture and platform was designed to fit with specific requirements of engineering simulations such as FEA. These requirements include (Lange and Nguyen, 2014):

- The architecture should be scalable, extensible and supported by a large community of users
- The data produced by the simulation engines can be computed by several nodes
- The visualization queries can be simple or complex
- The visualization queries can be performed in batch or in real-time
- The platform should be compatible with various computing infrastructure: HPC, grids and clouds, etc

Four steps were used in large finite element analysis and simulations. The first two steps in the analysis workflow have costly data-intensive computation. Preferably these take place in a parallel server infrastructure where the bulky data is stored, therefore avoiding transfer of the big data and exploiting the parallel HPC resources. The four steps are as follows (Teran and Garcke, 2014):

- Extraction: The raw data for the numerical analysis were obtained directly from the simulation, these variables can be scalars, vectors, or tensors defined on nodes or elements of a finite element mesh
- Preprocessing: This step was usually necessary to handle the huge data size (millions of nodes and elements) and the use of only subsets of the datasets in areas of interest
- Dimensionality reduction: A low dimensional representation was obtained from the dataset in this step
- Exploration: The simulated variables were organized in the low dimensional embedding space. Each numerical simulation was represented as a point in the obtained low dimensional space. The datasets could be explored efficiently due to the reduced dimensionality.

A parallel finite element analysis system was developed to simulate a large-scale analysis model on different supercomputers. In the system, the Hierarchical Domain Decomposition Method (HDDM), a very effective method for large-scale analysis, was developed to achieve the purpose of high performance computing and simulation (Murotani *et al.*, 2014).

Confluence of Big Data, Finite Element Analysis and High Performance Computing

IBM offers a full range of solutions that are able to create dynamic and flexible clusters, HPC cloud environments and Big Data analytics infrastructure that address compute and data intensive challenges specific to industries, especially in the Aerospace and Defense industry. IBM Application Ready Solutions for MSC Software provides a high performance simulation

environment for MSC Nastran, MSC Patran and SimManager simulation data manager. Platform Computing grid, cluster and HPC cloud solutions are ideal for application in every aspect of the Aerospace and Defense industry, including, Finite Element Analysis (FEA), Computational Fluid Dynamics (CFD) and Computer Aided Design (CAD), etc. Some main solutions are (IBM, 2014):

- Performance analytics and visualization (for such as FEA results).
- Hadoop/MapReduce support and file system for dealing with large data sets including big data
- Heterogeneous application, operating system and hardware support including IBM Technical computing systems

New analytic and finite element models have been used to build surgical simulations. Simulations have been developed and applied directly to medical device design processes. Computational methods such as FEA typically generate very large mesh and simulation datasets. Computational methods and Big Data can play an important role in medical device design. It is possible for us to develop new big data design tools that combine the power of computational engines and for the design team to steer the optimization process in real time. This can lead to a greatly improved medical devices solution platform. Designers will interact directly with an ensemble of thousands of simulation runs integrated into a virtual design environment. The design environment supports large-scale multidimensional data analysis and design optimization, leading to an improved and holistic understanding of design spaces supported by big data. Major advances are required in real-time and scalable visualization. FEA and other analyses are displayed in a single registered space. Novel interfaces and visualizations are used for real engineering work directly in an environment that includes a real-time connection to High Performance Computing (HPC) and provides the ability to visualize results in real time (Erdman *et al.*, 2014).

Discussion

Big Data has become an important method in recent years. It has been used in many areas at different levels. Some areas include government administration, supply chain and business management, Homeland Security, cyber warfare and cyber defense, financial services, manufacturing and medical applications and health care, etc. However there are some challenges in Big Data applications. But it has great potential and significant impacts in the above application areas. HPC is not a new discipline though there is a lot of new progress in this

field every year. FEA has almost been theoretically mature. A lot of commercial software packages such as Adina, Abaqus, MSC Nastran have been developed and have been applied well in most areas of engineering. Much research has been conducted in the confluence between FEA and HPC. Some research has been started in the confluence between Big Data and HPC. High Performance Data Analytics (HPDA) is actually the combination of HPC and Big Data. Little research has been done in following two aspects: (1) The confluence between Big Data and FEA; (2) the confluence of Big Data, FEA and HPC. Actually, FEA has met challenges due to highly varying conditions and properties in some situations such as real-time simulations in meteorological sciences and severe weather, real-time surgical simulations and numerical analysis in Aeronautical and Astronautical engineering. Big Data (or with HPC) helps provide solutions to these challenges. Besides the Finite Element Method (FEM), there are other methods such as the finite difference method (FDM), Finite Volume Method (FVM) and Boundary Element Method (BEM), etc. in scientific computation and numerical analysis. The confluence of Big Data, HPC and any of these methods (FDM, FVM and BEM) will be an interesting research topic and will bring great benefits to some engineering applications like numerical analysis in Aeronautical and Astronautical engineering.

Conclusion

Big Data can unlock significant value by making information transparent and perform predictive modeling based on Big Data analytics. HPC can offer immense potential for data-intensive computing and help unlock the wealth contained in big data.

The confluence between Big Data (with statistical methods and stochastic methods) and FEA can handle large size data and deal with highly varying conditions and properties. It helps resolve problems in large data storage and visualization. The confluence between Big Data and FEA can also provide more comprehensive analysis, predict defects and failures and optimize design in time.

The confluence between Big Data and HPC using Hadoop with HPC infrastructure helps reduce the processing time of large volumes of data that are in distributed computing environments. The confluence between HPC and FEA improves the efficiency and capability of FEA in complicated engineering and scientific jobs.

The confluence of Big Data, FEA and HPC enhances the capability of dealing with data intensive challenges, especially in the Aerospace and Defense engineering. These challenges include large and

heterogeneous data, analytics and real-time visualization, etc. The confluence also improves the integration of surgical simulations, multidimensional data analysis and medical device design and optimization in a real-time and scalable visualization environment. There will be more applications of the confluence of Big Data, FEA and HPC.

Acknowledgment

This study was supported in part by Technology and Healthcare Solutions, Inc. in Mississippi, USA. No conflict of interest to disclose.

Author's Contributions

This study was conducted by all the authors as group efforts at all stages. They equally contributed in this study.

Ethics

This article has not been published elsewhere. The corresponding author confirms that all the authors have read and approved the manuscript and no ethical issues involved.

Reference

- Abbey, T., 2014. Femap 11.0 new features. White Paper, Siemens PLM Software, Texas, USA.
- Bellini, P., M.D. Claudio, P. Nesi and N. Rauch, 2013. Taxonomy and Review of Big Data Solutions Navigation. In: Big Data Computing, Akerkar, R. (Ed.), Chapman and Hall/CRC, ISBN-10: 978-1-4665-7837-1, pp: 57-101.
- Colas, M., I. Finck, J. Buvat, R. Nambiar and R.R. Singh, 2014. Cracking the data conundrum: How successful companies make big data Operational. Technical Report, Capgemini Consulting. f
- Davenport, T., 2014. Big Data at Work: Dispelling the Myths, Uncovering the Opportunities. 1st Edn., Harvard Business Review Press, Boston, ISBN-10: 1422168166, pp: 240.
- Demchenko, Y., P. Grosso, C. de Laat and P. Membrey, 2013. Addressing big data issues in scientific data infrastructure. Proceedings of the International Conference on Collaboration Technologies and Systems, May 20-24, IEEE Xplore Press, San Diego, pp: 48-55. DOI: 10.1109/CTS.2013.6567203
- Dilsizian, S.E. and E.L. Siegel, 2014. Artificial intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Curr. Cardiol. Rep.*, 16: 441-441. DOI: 10.1007/s11886-013-0441-8
- Douglas, C.C., 2014. An open framework for Dynamic Big-Data-Driven Application Systems (DBDDAS) development. *Proc. Comput. Sci.*, 29: 1246-1255. DOI: 10.1016/j.procs.2014.05.112
- Eaton, C., D. Deroos, T. Deutsch, G. Lapis and P.C. Zikopoulos, 2012. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. 1st Edn., McGraw-Hill Companies, ISBN-10: 978-0-07-179053-6, pp: 141.
- Erdman, A.G., D.F. Keefe and R. Schiestl, 2013. Grand challenge: Applying regulatory science and big data to improve medical device innovation. *IEEE Trans. Biomed. Eng.*, 60: 700-706. DOI: 10.1109/TBME.2013.2244600
- Folk, M., 2006. NCSA-NARA investigations of HDF5 for long term retention of engineering data. Technical report, PDES, Inc.
- Guillén, A., M.I.G. Arenas, M. van Heeswijk, D. Sovilj and A. Lendasse *et al.*, 2014. Fast feature selection in a gpu cluster using the delta test. *Entropy*, 16: 854-869. DOI: 10.3390/e16020854
- Gutkind, E., 2013. The high performance computing approach to hadoop deployment. *Sci. Comput.*
- IBM, 2014. IBM platform computing. IBM. l
- Intel, 2014. Big Data meets high performance computing. White Paper, Intel.
- Jagadish, H.V., J. Gehrke, A. Labrinidis, Y. Papakonstantinou and J.M. Patel *et al.*, 2014. Big Data and its technical challenges. *Commun. ACM*, 57: 86-94. DOI: 10.1145/2611567
- Jean-François Lavignon, B., 2013. European technology platform for high performance computing: ETP4HPC strategic research agenda achieving HPC leadership in Europe. Technical Report, ETP4HPC.
- Lange, B. and T. Nguyen, 2014. Big data architecture for large-scale scientific computing. World Academy of Science - ABDA 2014, Las Vegas, United States.
- Letouzé, E., 2012. Big data for development: Challenges and opportunities. White Paper, Global Pulse, New York, USA.
- Merelli, I., H. Pérez-Sánchez, S. Gesing and D. D'Agostino, 2014. Managing, analysing and integrating big data in medical bioinformatics: Open problems and future perspectives. *BioMed Res. Int.*, 2014: 1-13. DOI: 10.1155/2014/134023
- Murotani, K., R. Shioya, S. Koshizuka, M. Ogino and H. Kawai, 2014. Development of a numerical library based on hierarchical domain decomposition for post peta-scale simulation. Proceedings of the ATIP Workshop: Japanese Research Toward Next-Generation Extreme Computing, (GEC' 14), Post-PETA CREST Talk #7.
- O'Leary, D.E., 2013. 'Big Data', the 'internet of things' and the 'internet of signs'. *Intell. Syst. Account. Finance Manage.*, 20: 53-65. DOI: 10.1002/isaf.1336

- Pacitti, E. and P. Valduriez, 2012. Zenith: A hybrid P2P/cloud for big scientific data. ERCIM News.
- Parashar, M., 2014. Big data challenges in simulation-based science. Proceedings of the 6th international workshop on Data intensive distributed computing, Jun. 23-27, ACM, New York, USA, pp: 1-2.
DOI: 10.1145/2608020.2612731
- Peng, J., D. Liu and K.H. Law, 2003. An online data access system for a finite element program. Technical Report, Stanford University.
- Salem, A.A., J.B. Shaffer, D.P. Satko, S.L. Semiatin and S.R. Kalidindi, 2014. Workflow for integrating mesoscale heterogeneities in materials structure with process simulation of titanium alloys. *Integrat. Mater. Manuf. Innovat.*, 3: 1-24.
DOI: 10.1186/s40192-014-0024-6
- Schneider, R.D., 2012. Hadoop for Dummies. 1st Edn., John Wiley and Sons, Canada, ISBN-10: 1118250516, pp: 64.
- Slack, E., 2014. The evolution of high-performance computing storage architectures in commercial environments. White Paper, Storage Switzerland, LLC.
- Teran, R.I. and J. Garcke, 2014. Data analytics for simulation repositories in industry. *Lecture Notes in Informatics*.
- Nesterov, Y., 2014. Algorithmic challenges of big data. CORE/INMA (UCL), Max Planck Institute.
- Valduriez, P., 2015. Data-intensive HPC: Opportunities and challenges. Proceedings of the Workshop on Big Data and Extreme-scale Computing, Jan. 28-20, Barcelona.
- Yan, J., S. Xin, Q. Liu, W. Xu and L. Yang *et al.*, 2014. Intelligent supply chain integration and management based on cloud of things. *Int. J. Distributed Sensor Netw.*, 2014: 1-15. DOI: 10.1155/2014/624839