Applied Medical Informatics
Editorial Vol. 42, No. 1 /2020, pp: 9-18

# On the Impact of High Performance Computing in Big Data Analytics for Medicine

**Virginia NICULESCU***

Babeş-Bolyai University, Cluj-Napoca, Mihail Kogalniceanu Str, No. 1, 400084 Cluj-Napoca, Cluj, Romania.
E-mail: vniculescu@cs.ubbcluj.ro

* Author to whom correspondence should be addressed; Tel.: +40 727 792 533.
**Received: December 9, 2019 /Accepted: March 22, 2020 / Published online: March 30, 2020**

**Abstract**
High-Performance Computing (HPC) is a mature domain that proved to be critical in running large-scale modelling and simulation using numerical models. The Big Data Analytics domain (BDA) has been rapidly developed over the last decade to process vast amounts of data now being generated in various fields. Data-intensive applications are needed in various fields of medicine and healthcare ranges from advanced research, as genomics, proteomics, epidemiology, and systems biology, to medical diagnosis and treatments, or to commercial initiatives to develop new drugs. BDA needs the infrastructure and the fundamentals of HPC to face with the required computational challenges. There are important differences in the approaches of these two domains, those that are working in BDA focus on big data specific features such as Volume, Velocity, Variety, Veracity, and Value, while HPC scientists tend to focus on Performance, Scaling, and Power efficiency of computations. In this paper we intend to analyse the need of these two domains in the development of the future medicine that is required to become at least Personalized, Predictive, Preventive, and Participatory.

**Keywords:** Big Data; Data Analytics; High Performance Computing (HPC); Health Informatics; Data Collection

## Introduction

High-Performance Computing (HPC) is a mature domain that proved to be critical in running large-scale modelling and simulation using numerical models. In the last five decades the HPC scientific community focused especially on increasing the ability of scientists to model and simulate natural phenomena [1].

The data analytics domain has been rapidly developed over the last decade to process a huge amount of business, industrial process, and social network data now being generated [2]. But, in general, the Big Data Analytics (BDA) domain was developed much more inside the business oriented community than in the scientific computing community. BDA started to receive public prominence in the first decade of 21st century, being promoting by online enterprises like Google, Amazon, and Facebook. Data-intensive applications are needed in various fields ranges from advanced research - as genomics, proteomics, epidemiology, and systems biology - to commercial initiatives to develop new drugs and medical treatments, agricultural pesticides, and other bio-products. Data processing is still needed in the more traditional domains as physics, climate, and astronomy. Still, even there, the data orientation could bring important advantages, and so HPC needs, on its turn, to adopt data-driven paradigms [3].

Essential differences in these two approaches exist. The BDA focus mainly on the initial 3Vs of big data, namely Volume, Velocity, and Variety, while HPC scientists tend to focus on Performance, Scaling, and the Power efficiency of computations.

As we are heading towards extreme-scale HPC coupled with data-intensive analytics, the integration of Big Data and HPC is currently a hot topic of research [3].

High Performance Data Analytics (HPDA) combines HPC with data analytics [4,5]. This enables HPC's parallel processing software to run powerful analytic software at speeds higher than a teraflop ($10^{12}$ floating-point operations per second). This approach facilitates the examination of large data sets and extracting conclusions about the information these data contain. It is estimated that one of the most socially and economically important HPDA applications will be the transition from today's procedures-based medicine to the personalized, outcomes-based health care system.

The purpose of this paper is to analyse the need and the convergence between BDA and HPC when applied in medicine and healthcare. The next section presents the differences and the commonalities between HPC and BDA, and emphasizes the benefits that could be obtained when the advances in these two domains are using together. The need and the advantages of using HPDA in medicine are emphasized, too. The current challenges in medicine are discussed, together with their requirements regarding HPC and BDA.

## High Performance Computing versus Big Data Analytics

During many years '*big data*' has been synonymous with HPC. This was because HPC tries to solve difficult computations with big volumes of data performantly. At the same time, HPC led to the progress and evolution of hardware systems and networking (the HPC hardware installations). Simulation-based data-intensive computing is a long-standing HPC category with various examples such as weather forecasting/climate modelling, physics-molecular dynamics, engineering - parametric modelling for product design, finance: portfolio optimization, risk analysis, prices analysis; life sciences: genomics, drug discovery and more; security: signal intelligence. All these simulations involve vast amounts of data that should be process and store. HPC is a well-founded domain sustained mainly by the scientific community that proves a strong ability to deliver new scientific discoveries uniquely [1].

BDA is a much newer domain that has been initiated by the industry needs, which become exceptionally well quoted and famous. Big Data represents, beside the proper meaning of the term, a shorthand for advancing trends in technology that are oriented on a new approach to understanding the world and making decisions [2]. Following this, "Data Science" becomes a new scientific domain, and the corresponding professionals and associated jobs started to be highly demanded. The increasing interest in this domain leads to interesting definitions for Data Scientist; such an example is the definition given by J. Wills in [6]: *"Person* [i.e. data scientist] *who is better in statistics than any software engineer and better at software engineering than any statistician*".

The focus in BDA is on a large amount of various data that were initially characterized by using 3Vs – Volume, Variety, and Velocity, which were introduced by D. Laney in 2001 in [7]. Volume refers to the amount of data, the size of the analysed datasets, which are usually ranging from terabytes ($10^{12}$ bytes) to zetabytes ($10^{21}$ bytes). Velocity is related to the motion of data, the fast speed of new data generation, and to the speed of data processing. Variety refers to the various types, and forms (possible unstructured) of the data and its resultant complexity. These 3Vs are the most definitory for Big Data since volume is automatically implied by the term, but if large volumes of data had been used before in HPC, they were usually more structured and with similar types, and without coming in streaming with high velocity.

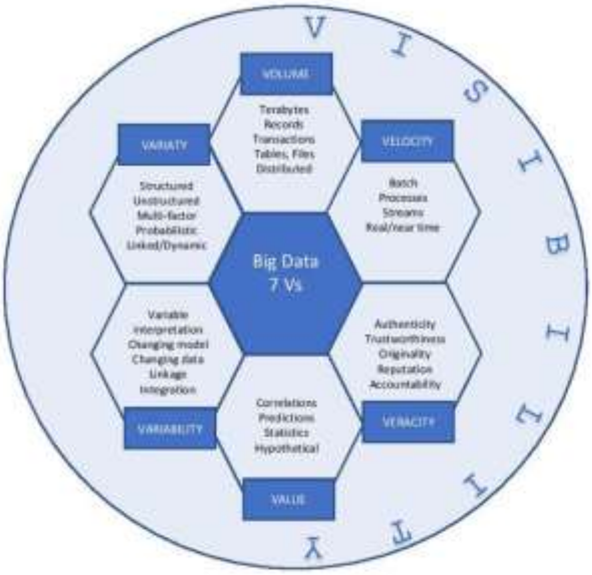


**Figure 1.** The 7Vs of Big Data: Volume, Velocity, Variety, Variability, Veracity, Value, and Visibility

More recently, additional Vs have been proposed to characterize the domain (2013 - [8,9]). These include Variability -- the increase in the range of the values and the continuous data changing, Value, which addresses the need for valuation of enterprise data and to additional information that data can bring to generate knowledge, Veracity that indicates the trustworthiness and inherited ambiguities due to data uncertainty and inconsistency, and Visibility - the state of being visible-accessible. Visibility is important because data from disparate sources need to be analysed, and critical data could exist, but if they are not visible to the processes of a Big Data system, the complete analysis becomes impossible; but in the same time, unauthorised visibility is a risk. Figure 1 emphasizes these Vs and some of their implications. Even more Vs were proposed to be included in this list: Vagueness, Validity, Venue, Viability, Vincularity, Virility, Viscosity, Visible, Vitality, Vocabulary, Volatility.

The general goal of BDA is to examine large amounts of data to discover hidden patterns, correlations, and other insights. Some benefits that big data analytics brings are speed and efficiency in making decisions, insights identification, which gives the ability to work faster, and that introduces a new competitive edge between organizations [10].

Artificial intelligence (AI) is also a very popular domain nowadays. AI is strongly connected to BDA since many AI algorithms are used in empirical efforts that attempt to generate accurate representations of natural, economical, and social phenomena from *unstructured data* gathered from different sources. The uncontrolled nature of this unstructured data leads to the need of extracting and cleansing meaningful information in order to train AI systems.

Besides the fact the HPC and BDA were initiated and developed by different communities: scientific and industrial, some important differences should be emphasized.

The general HPC focus is on:

- Performance
    - fast computation,
    - accuracy of computation (small errors, high precision, etc.),
- Scaling - big problems on big systems, and
- Power efficiency.

Related to the domain orientation, HPC tends to focus on modelling and simulation (e.g., Monte-Carlo) and to generate accurate representations of what happens in nature**.** In the high energy and astrophysics domains, this could involve examining images using deep learning methods – so to use artificial intelligence methods, too. Difficult operations such as: examining time series, performing signal processing, bifurcation analysis, harmonic analysis, nonlinear system modelling and prediction and/or other numerical analysis methods, are used in many scientific and engineering computations[1,10,11]. Therefore, we may conclude that in general, in HPC models are known, and they should be applied to data, while in BDA the data are analysed to discover new models/patterns.

An elaborate analysis and comparation between Data Analytics and Computational Science software ecosystems has been done by Reed and Dongarra in [11]. The comparison is done on four levels: hardware, system software, middleware& management, and application level. The major differences between the HPC and the BDA ecosystems—software development paradigms and tools, storage models, virtualization and scheduling strategies, resource allocation policies, strategies for redundancy, and fault tolerance—can be explained by the fact that each evolved during a distinctly different phase of the ongoing digital revolution, driven by distinctly different optimization criteria.

*HPC and BDA Integration*

There is a real need to allow these two important domains to collaborate if not even entirely fuse together. The goal is to use the progress already obtained with HPC into BDA, but at the same time to enhance HPC using BDA technologies. Two important directions should be analysed, namely hardware and respectively software.

- **Hardware.** Most of the HPC platforms are compute-centric without efficient storage and I/O (data movement) capabilities. This assures fast computation but not so fast movement of data (data between different computation nodes could be transferred very fast – generally through InfiniBand connections, but the data movements between different system components are not so fast–being based on Ethernet connections). On the other hand, BDA needs to continuously process on large and growing volumes of information, to allow fast and frequent data movement between application servers, network connections, and across the storage.

**Table 1.** Hardware evolution – oriented on HPC and BDA needs.

| Supercomputer (HPC Cluster) | Data center | Cloud infrastructure |
|---|---|---|
| - A system explicitly optimized for a class of tasks that would not be achievable with more conventional computer architectures.<br>- Powerful components arranged ***to achieve the highest possible speeds*** in the execution of the task.<br>- Priority = performance over cost.<br>- The term started to be used since 1960s. | - A facility that contains a large number of computer servers and related equipment.<br>- A collection of general-purpose computers connected with ordinary, commercial class interconnections to ***maximize return on investmen***t.<br>- Priority = cost over performance.<br>- The term started to be used since 1990s | - A facility that contains computers, storage, network, related components, and facilities required for cloud computing and IT-as-a-Service.<br>- The hardware is similar to typical data center infrastructure, but ***it is virtualized and offered as a service to be used via the Internet***.<br>- Priority=elasticity/cost over performance<br>- The term started to be used since 2000s (2006) |

As possible solutions, the HPC community is planning to address BDA challenge related to data movement by reducing data movement at all levels via in-memory processing or by accelerating data movement via more capable fabrics and interconnect networks - this will lead to improved communications. Hardware accelerators as GPU were developed such that to allow simple, efficient, and general-purpose parallel programming. They represent a new trend that leads to many improvements in both domains: HPC and BDA.

- **Software** - *Different Systems for Different Jobs*. Depending on the characteristics of the jobs, different systems are needed. We may differentiate partitionable tasks that rely on regular access patterns and non-partitionable which are based on irregular access patterns. In the first category falls most of the big data jobs for which the global memory is not important, and for them, standard clusters systems could be used together with the well-known software infrastructures(middleware) as Hadoop or Cassandra[10]. The second category includes much difficult jobs such as graphing where global memory becomes important and which are orientated not only on model search but on discovery; for these, systems oriented on efficient data movement (e.g., Graph DB) are needed.

In [12] an analysis of the evolution of HPC identifies three epochs:

- Epoch 1: Breaking Supercomputers into Clusters
- Epoch 2: The Multi and Many-core Explosion
- Epoch 3: Melting the Edges of Hardware and Software Through Co-Design

The last epoch emphasizes the need to have a stronger interconnection between software and hardware – the co-design orientation proposes reversing the development direction from software to hardware. The co-design strategy is based on identifying leading edge, high-impact scientific applications and providing concrete optimization targets rather than focusing on speeds and feeds (FLOPs and bandwidth) and percent of peak. This leverages deep understanding of specific application requirements and adapting the hardware development such that to satisfy them [13].

*HPDA - High-Performance Data Analysis*

High-performance data analysis (HPDA) is a term adopted to describe the conversion of the data-intensive HPC analysis into the high-end commercial data analytics (BDA). Classically "big data" tasks are intended to be executed on commodity hardware in a fixed scaled architecture. Still, since there are certain situations where ultra-fast, high-capacity HPC are needed, approaches with better scalability are preferred. This is the domain of HPDA [4,5].

**HPDA and HPC infrastructures.** Due to the various application cases, it cannot be said, which is the best HPC architecture to manage or analyse big data analytics workloads. Heterogeneous systems are needed to obtain productivity, and a solution is to move heterogeneous HPC resources to the cloud, and so the big data analysts may be able to afford access to the latest computation power. For example, important commercial cloud vendors, such as Amazon, are adding various HPC elements to offer high-end services.

**HPDA and Clouds.** Because clouds are much less effective on jobs requiring major inter-processor communications via MPI or other protocols, they are useful primarily for embarrassingly parallel HPC jobs. Highly parallel HPDA problems can be attractive problems for being solved on public clouds. It is estimated that HPDA cloud usage is expanding to include even less partitionable problems such as graph analytics, as long as the problems do not have to be solved in real-time.

**HPDA and AI.** AI is rapidly moving into the mainstream of daily life. Nowadays, AI methods are integrated into an extensive range of enterprise applications, and they also intervene in everyone's daily lives. The following requirements direct the progress in AI:

- Make models bigger.
- Use more data.
- Reduce research cycle time.

So, an important evolution of AI was due to the developments done inside HPC domain. Simple neural networks that were once impossible to train can now be trained in hours and minutes, and the best AI algorithms scale with systems. By using HPC software (tools, frameworks, platforms) and hardware it is possible to accelerate large-scale experiments based on deep learning that solve hard (difficult to solved) problems; training deep neural networks is in fact an HPC problem, and HPC hardware and software approaches could reduce the training time, making it possible to achieve better results.

It should be mentioned that at the same time, this has also elevated HPC, together with machine learning and data analytics, as an essential asset for modern businesses.

## New Challenges in Medicine

"*Medicine is going to become an information science. In 10 years or so, we may have billions of data points on each individual, and the real challenge will be to develop information technology that can reduce that to real hypotheses about that individual.*" -Leroy Hood [14]

This is an interesting point of view based on an approach that leads to some new developments in medicine. Future medicine is considered to become a much-evolved medicine [13-15]. The first step in defining the characteristics of this advanced medicine was through P3 medicine that is Personalized, Predictive, Preventive, but very soon was updated to P4 by adding the Participatory characteristics [5]. P4 medicine is also based on three converging megatrends driving the transformation of healthcare**.** These megatrends are [14]:

- ***Systems biology & systems medicine*** - these increase the ability to decipher the biological complexity of disease;
- ***The digital revolution*** - this radically enhanced capability for collecting, integrating, storing, analysing and communicating data and information, including conventional medical histories, clinical tests and the results of the tools of systems medicine;
- ***Consumer-driven healthcare and social networks*** – having access to a lot of information, the consumer interest in managing their health increases more and more. At the same time, the social networks offer the needed information for extracting statistics and prediction models related to healthcare.

As happened with the Vs of Big Data, more (P) characteristics were introduced leading to P6 Medicine [15,16], which is Personalized, Predictive, Preventive; Participatory, Psycho-cognitive and Public. P6 medicine was preceded by "P4 + $C^n$ Hippocratic revolution" defined in [17] where $C^n$ stands for "community, collaboration, self-caring, co-creation, co-production, and co-development using technologies delivered via the internet". The public aspect is in P4 implicitly considered through Participatory feature – in order to allow people to participate, the system should be public. Still, emphasizing this feature in P6 definition enforces the orientation on "Health Web Science", that combines e-health, e-medicine, and telemedicine, where computers and innovative online health communities play a very important role [15]. Since World Health Organization defines health as a state of complete physical, mental, and social well-being, psychology was considered more and more important for the future modern medicine [16]. On the other hand, we cannot consider that P4 excludes this, since naturally, all the symptoms and characteristics of the patients should be considered.

Following the strategy of adding new (P) features (encountered for the Vs of Big Data, too) we may also add Predispositions–genetic and environmental. In this way, we would arrive to P7 medicine represented in Figure 2.

Facilitated by rapid advances in data science and technology, driven by the request to utilize up-to-date information, a big data revolution in medicine is to have happened. This new wave has the potential to positively transform medicine. It is estimated that a transition from today's procedures-based medicine to personalized, outcomes-based health care will take place. This new approach will allow the identification of highly effective treatments in near-real-time by comparing: an individual's genetic makeup, health history, and symptomology against millions of archived patient records. Electronic health records (EHRs) could be used as a base for health care providers in finding the desired data more efficiently and precisely [18-20].

This transition could be achieved only through a global decision-support tool of utility for the general health care community. Such a tool is not impossible to be designed and developed but represents a tremendous challenge that implies high-end technology, especially related to big data and high-performance computing.

These tremendous technology requests are emphasized in many different medical studies. In [21] we found an interesting comparison of the evolution of the volume and complexity of data in neuro-imagining and genomics and the computational power – Table 2. The data complexity measures in this case the data heterogeneity (e.g., new imaging data acquisition modalities or sequence coverage depth; complexity of 5 indicates a 5-fold increase of the data diversity over 1985).

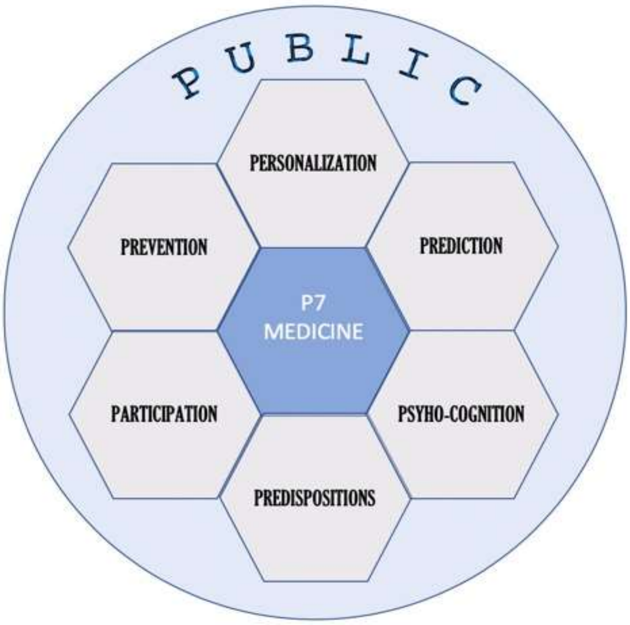


**Figure 2.** The P7 Medicine: Personalization, Prediction, Prevention, Participation, Psycho-cognition, Predispositions, and Public.

**Table 2.** The increase of data volume and complexity rel. to computational power[1] (adapted from [21]).

| Neuroimaging (annually) | | Genomics (BP/Yr) | | Moore's Law (transistor counts) | Bandwidth (Edholm's Law[2]) | Years |
|---|---|---|---|---|---|---|
| Size | Complexity | Size | Complexity | | | |
| 200 GB | 1 | 10 MB | 1 | $1\times10^{5}$ | $10^{5}$ | 1985-1989 |
| 1 TB | 2 | 100 MB | 2 | $1\times10^{6}$ | $10^{6}$ | 1990-1994 |
| 50 TB | 5 | 10 GB | 3 | $5\times10^{6}$ | $10^{8}$ | 1995-1999 |
| 250 TB | 6 | 1TB | 4 | $1\times10^{7}$ | $10^{9}$ | 2000-2004 |
| 1 PB | 7 | 30TB | 5 | $8\times10^{6}$ | $10^{10}$ | 2005-2009 |
| 5 PB | 8 | 1 PB | 7 | $1\times10^{9}$ | $10^{11}$ | 2010-2014 |
| 10+ PB | 9 | 20+ PB | 8 | $1\times10^{11}$ | $10^{13}$ | 2015-2019 |

The computational power is measured there only based on transistors counts (Moore's Law[3]), which is not exactly right, considering the latest computational development. A better and more reliable analysis should consider the computational power achieved by using high performance supercomputing that could go now until 148,600.0 TFlop/s (from Top500 Supercomputing site [22]). There are many other medical fields such as omics' disciplines, cardiology, diabetology, rheumatology, which emphasized an increasing requirement in size of the data needed to be analysed [18-20].

*Big Data Solutions*

**Medical data characteristics.** Many current studies [19-24], also emphasize other important characteristics of the healthcare data that should be analysed:

[1] FLOPS, flops or flop/s - floating point operations per second - is a measure of computer performance, useful in fields of computations that require floating-point calculations; when it is a more accurate measure than measuring instructions per second.
[2] Edholm's law predicts that the bandwidth and data rates double every 18 months.
[3] Moore's law is the observation that the number of transistors in a dense integrated circuit doubles about every 2 years.

- they have various types (high level of heterogeneity)
- they have complex correlations and patterns
- they are collected from different sources, and
- they could be incomplete.

Besides these, two important things characterize the big healthcare data, their energy, and life-span [21]. It is considered that the data energy encapsulates the holistic information included in the data. Because of the huge data size, the data energy may often represent a significant portion of the joint distribution of the underlying healthcare process. The life-span of big healthcare data characterizes the value of the data over time; this also follows an exponential model. But in this case, the lifespan and value of healthcare data rapidly decay, and this represents information devaluation.

Through big healthcare data analytics, it is intended to solve some of the holistic information-processing challenges and rapidly provide effective estimation and prediction based on dynamic and heterogeneous data. Still, big healthcare data analytics do not come without cost and challenges, and to achieved their potential, significant investments are needed. Due to the characteristics of the medical and healthcare data, these developments are crucially influenced by the domain of HPC.

On the other hand, there are two other big problems of Big Data approach in medicine, namely data security and data integrity. As in other domains, as financial, military, or high-end technology, the security of personal medical information is essential. Data security is a general problem of the world today, and there are huge investments in developing solutions for this, from both academic and industry environments. Related to data integrity, the healthcare domain comes with specific problems. Based on an information-theoretic interpretation of Gödel's incompleteness principle [25] it is stated in [21] that health data cannot be consistent and complete at the same time. In essence, Gödel's incompleteness theorems essentially say that in general a mathematical formal system beyond a certain complexity is either inconsistent or incomplete. On the other hand, in algorithmic information theory, an output that can be generated by a computer program with binary size much lesser than the output itself is called "compressible" or "reducible". Given that a mathematical formal system and its theorems are comparable to an algorithm and its outputs, the incompleteness principle in math (improvable theorems do exist in a formal system) leads to its equivalence in incompressibility (algorithmically incompressible outputs do exist). This is confirmed also by the fact that data items in some empirical data sets appear to show correlations with one another, but this is in many interesting cases arbitrary. This would mean that any computational inference, or decision making, based on big healthcare data would be expected to either be reliable within a restricted domain (e.g., time, space) or be more broadly applicable (e.g., population studies) but less consistent, but not both. This is also supported by scientific experiences where statistical inference on small or large sample sizes depends on corresponding large or small variances of the parameter estimations, respectively. But, because of the expected violations of the parametric assumptions and the lack of control, when there are huge number of samples, the accuracy is inversely proportional to the sample-size of the data. Still, the application of this principle for the Big Data approach in medicine should be more carefully analysed.

## Conclusions

The new developments lead to new approaches and developments in both of the two discussed domains – HPC and BDA. On one side, HPC community realizes that there are more things in data storage than just files, and the 'self-service' mechanisms could bring important advantages. On the other side, BDA realizes that HPC hardware and solutions can really speedup analytics.

It could be noticed now that all major public cloud services have an HPC offerings, and many academic HPC centers offer cloud infrastructures and BDA tools. At the same time, the advances in HPC research could expand analytics. HPDA is a very new domain that intend to use the HPC technologies in data analysis.

Using HPC for data-intensive challenges enlarges the HPC's contributions to science, commerce, and society. Also, the classical HPC application areas – modelling and simulation – are moving to new advances by using more modern high-performance analytics.

Integrated HPC platforms and software stacks will be able to provide capabilities for the new demanding computational needs across different domains, offering more opportunities for progress and so facilitating innovative products and services. One of the most important application arias is represented by the newly proposed health systems that could support a more advanced medicine. Identifying highly effective treatments in near-real-time by comparing an individual's genetic makeup, health history, and symptomology against other millions of patient records induces enormous computational challenges. HPDA technologies, and in general, integration of HPC with BDA, will offer the framework that allows the development of complex decision-support tools of unprecedented utility for the global health care community that will facilitate the future medicine, which has to support at least personalization, prediction, prevention, and public participation.

**List of abbreviations**

HPC – High-Performance Computing
BDA – Big Data Analytics
HPDA – High-Performance Data Analytics
AI – Artificial Intelligence
GPU – Graphics Processing Unit

**References**


1. Dongarra J. Trends in High Performance Computing. The Computer Journal 2004;47(4):399-403.
2. Lohr S. The age of big data. New York Times 11. 2012
3. Asch M, Moodre T, Badia R, Beck M, Beckman P, Bidot T, et al. Big data and extreme-scale computing: Pathways to Convergence-Toward a shaping strategy for a future software and data ecosystem for scientific inquiry. The International Journal of High Performance Computing Applications 2018;32(4):435-79.
4. Pacific Northwest National Laboratory(PNNL) staff. High Performance Data Analytics. [Online] [cited 2019 November 20] Available from: https://www.pnnl.gov/computing/HPDA/Images/HPDA_2017.pdf
5. Damkroger PA. The Intersection of AI, HPC and HPDA: How Next-Generation Workflows Will Drive Tomorrow's Breakthroughs. [Online] [Cited 2020 January 20] Available from: https://www.top500.org/features/content/the-intersection-of-ai-hpc-and-hpda-how-next-generation-workflows-will-drive-tomorrows-breakthroughs/
6. Wills J. Data Scientist definition. [Online] [Cited 2019 November 20] Available from: https://twitter.com/josh_wills/status/198093512149958656.
7. Laney D. 3D data management: Controlling data volume, variety and velocity. META Group Research Note 6 (2001) [Online] 2012 [Cited 2019 November 20] Available from https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
8. Rijmenam M. Why The 3V's Are Not Sufficient to Describe Big Data. [Online] [Cited 2019 December 12]. Available from: https://datafloq.com/read/3vs-sufficient-describe-big-data/166
9. Livingstone R. The 7 vs of big data. [Online] 2013 [Cited 2019 November 20]. Available from: http://rob-livingstone.com/2013/06/big-data-or-black-hole/
10. Demchenko Y, Grosso P, de Laat C, Membrey P. Addressing big data issues in Scientific Data Infrastructure. 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, 2013, pp. 48-55. doi: 10.1109/CTS.2013.6567203
11. Reed DA, Dongarra J. Exascale computing and big data. Commun. ACM 2015;8(7):56-68.
12. InsideHPC staff. The Evolution of HPC. [Online] [Cited 2019 November 20]. Available from: https://insidehpc.com/2016/08/the-evolution-of-hpc/

13. Barrett RF, Borkar S, Dosanjh SS, Hammond SD, Heroux MA, et al. On the Role of Co- Design in High Performance Computing. Transition of HPC Towards Exascale Computing. E.H. D'Hollander et al. (Eds.) IOS Press, 2013, pp. 141-54. doi:10.3233/978-1-61499-324-7-14
14. Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. Science 2004;306(5696):640-3.
15. Bragazzi NL. From P0 to P6 medicine, a model of highly participatory, narrative, interactive, and "augmented" medicine: some considerations on Salvatore Iaconesi's clinical story. Patient Preference and Adherence 2013;7:353-9.
16. Bragazzi NL, Del Puente G. Why P6 medicine needs clinical psychology and a trans-cultural approach. Health Psychology Research 2013;1(1):e5. doi: 10.4081/hpr.2013.e5
17. Cumming G, Fowlie A, Brooks E, McClusky C, French T, Eckl C, et al. H = P4 + C and Health Web Science: "A Hippocratic Revolution in Medicine". [Online] [Cited 2019 November 20]. Proceedings of the ACM WebSci '11 3rd International Conference on Web Science; June 14–17, 2010. Koblenz, Germany. Available from: http://www.websci11.org/fileadmin/websci/Papers/Health_WS_Workshop_A_Hippocratic_Revolution.pdf.
18. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. Personalized Medicine. 2013;10(6):565-76.
19. Hood L, Balling R, Auffray C. Revolutionizing medicine in the 21st century through systems approaches. Biotechnology Journal. 2012;7(8):992-1001.
20. Weston AD, Hood L. Systems biology, proteomics, and the future of healthcare: toward predictive, preventative, and personalized medicine. Journal of Proteome Research. 2004;3(2):179-96.
21. Dinov ID. Volume and Value of Big Healthcare Data. Journal of Medical Statistics and Informatics 2016;4:3. doi:10.7243/2053-7662-4-3
22. TOP500 List - November 2019 [Online] 2019 [Cited 2019 November 20] Available from: https://www.top500.org/lists/2019/11/
23. Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. Journal of Integrative Bioinformatics. 2018;15(3):20170030. doi: 10.1515/jib-2017-0030
24. Austin CO, Kusumoto FM. The application of Big Data in medicine: current implications and future directions. Journal of Interventional Cardiac Electrophysiology 2016;47(1):51-9.
25. Gödel K. Some basic theorems on the foundations of mathematics and their implications. In: Feferman S. Ed., Kurt Gödel Collected Works, Vol. III, Oxford University Press, 1995, pp. 304-323.