Reconfigurable Computing for Analytics Acceleration of Big Bio-Data: The AEGLE Approach

Andreas Raptopoulos[1], Sotirios Xydis[2], and Dimitrios Soudris[2]

[1] EXUS, 73-75 Mesogeion Av and Estias Str 1, 115 26, Athens, Greece
[2] National Technical University of Athens, School of Electrical and Computer Engineering, Athens, Greece

Abstract. This paper presents the main directions of the AEGLE project, that targets to integrate cloud technologies together with heterogeneous reconfigurable computing in large scale healthcare systems for Big Bio-Data analytics. AEGLEs concept brings together the "hot" big-data technologies with the health industry eventually leading to integrated care and creating a win-win situation for both. We provide the addressed Big Data health scenarios and we describe the structural elements of the proposed solution, with emphasis given in the exploitation of high-performance reconfigurable engines for Big Data analytics acceleration integrated to the AEGLE ecosystem, enabling personalized and integrated healthcare services, while also promoting related research activities.

1 Introduction

At the centre of health debates there are open questions on how to manipulate data and how to produce value out of it, share it and secure it [1]. Answers to these questions will create opportunities "to predict long-term health conditions and identify non-traditional intervention points, as well as to design better diagnostics tools, prevent diseases, and increase access to and reduce the costs of healthcare [2]. As discussed in [3], effective use of data in the US health sector could generate USD 300 billion in value per year. The implementation of big data analytics in the healthcare sector has the potential to boost the integration of user-generated data with official medical data, leading to healthcare that is more integrated and personalised [4].

Several European initiatives [5] have already pinpointed the importance and usefulness of healthcare big data, e.g. to predict the outbreak of an epidemic etc. Additionally, business interest is growing like the Open Data initiative, where health big data providers, governmental and research institutes and industry aim to develop a vendor-neutral Big-Data platform [6]. However, the strategic advantage brought by Big-Data in healthcare still materializes at slow paces, as only some large-scale organizations have established few pilot or proof-of-concept projects.

Nowadays, there is an obvious gap in the area of big data analytics for Health Bio-data. Data-driven services are still needed to cater for the data versatility, volume and velocity within the whole data value chain of healthcare analytics. A true opportunity exists to produce value out of big data in healthcare with the goal to revolutionize integrated and personalised healthcare services, which have been recently introduced for the management of complex medical conditions e.g. various chronic disease conditions, chronic malignant and non-malignant disorders. Evidently, the implementation of data value chains for healthcare big data analytics has the potential to lead Europe gain the leading position worldwide, pioneering by leveraging Big Data analytics solutions in a domain that is of critical importance for all European countries and their citizens.

The AEGLE[3] project targets to address the aforementioned open issues by implementing a full data value chain to create new value out of rich, multi-diverse, big health data. The project builds upon the synergy of heterogeneous High Performance Computing (HPC) exploiting reconfigurable architectures, Cloud and Big Data computing technologies, offering tools for a seamless use of different sorts of capacity across HPC, Cloud, Grid and Big Data computing for the benefit of all areas of Horizon 2020. AEGLE will provide a framework for Big Data analytics for healthcare that will overall enable and promote innovation activities that place health at the spotlight.

2 AEGLE's relevance and positioning in respect to other R&D projects

AEGLE aims to generate value from the healthcare data value chain data with the vision to improve translational medicine and facilitate personalized and integrated care services overall improving healthcare at all levels, to promote data-driven research across Europe and to serve as an enabler technology platform enabling business growth in the field of big data analytics for healthcare Currently numerous R&D projects are running, regarding health and ICT technologies. Most of them are targeting to obtain a proof of concept (having limited and controlled validation phases) on the impact of sensing and monitoring devices in the treatment and management of a disease.

Some of the projects have already examining in more depth the concept of integrated care concerning chronic diseases like WELCOME[4] or SWAN-iCare[5]. Additionally health projects have started exploiting cloud capabilities, like e-health GATEway to the Clouds[6] or BIOBANK Cloud[7] as well as perform large scale analysis like MD-Paedigree[8] and OPENPHACTS[9]. Most of these projects however aim to the storage and analysis of mainly biological data (e.g. genomics), and this is the field where commercial products can be found like CLCbio[10] platform aiming to the analysis of DNA, RNA and protein.

Figure 1 illustrates the positioning of AEGLE project in respect to other projects on eHealth and Big Data for healthcare. As it can be seen, none of the existing Big-Data projects are completely dedicated to healthcare and the provision of corresponding healthcare services, or the management of diseases. AEGLE combines all elements of

[3] AEGLE: An analytics framework for integrated and personalized healthcare services in Europe
[4] http://www.welcome-project.eu/
[5] http://www.swan-icare.eu
[6] http://www.jisc.ac.uk/whatwedo/programmes/di_research/researchtools/ehealth.aspx
[7] http://www.biobankcloud.com
[8] http://www.md-paedigree.eu
[9] http://www.openphacts.org
[10] http://www.clcbio.com

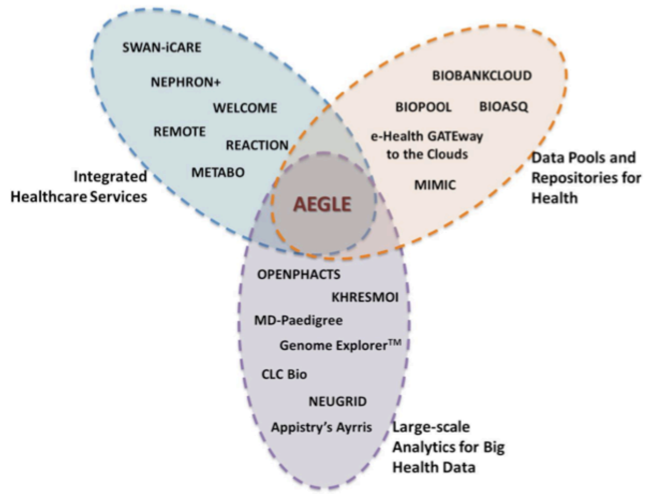

Fig. 1. Positioning in respect to other R&D projects on healthcare.

the full value chain (storage of large volumes of data, big data analytics, cloud computing and provisioning of integrated care services), targeting to cover the whole field of health big data analytics. It will also liaise with other projects (e.g OPENPHACTS etc), for taking advantage from their developments, resulting to a more advanced and extended system.

3 Health scenarios for Big BioData analytics

3.1 Chronic Lymphocytic Leukaemia (CLL) Big BioData Management

CLL is a cancer of the blood system, more precisely a B-cell lymphoma, affecting mainly people of advanced age (median age at diagnosis of 70 years). CLL is a chronic, incurable disease, leading to great distress for patients and their families as well as huge costs for the health care system.The relevant Big Data include: (i) the growing output of multiple -omics technologies, such as proteomics and metabolomics, exome/whole-genome sequencing (genomics), RNA-sequencing and methylation sequencing; (ii) results from functional studies focusing on immune signalling; (iii) results from in vitro drug screening; and (iv) demographic and clinico-biological information from patients. With the advent of high-throughput technologies, a wealth of information has become available about CLL at an unprecedented scale in medical research.

Analysis will be performed both on local repositories provided by AEGLE partners as well as public repositories such as Gene Expression Omnibus (GEO) [8] and Sequence Read Archive (SRA) [9] that archive and freely distribute functional genomics data of multi-terabyte- scales submitted by the scientific community. In the context of CLL data management and analysis services, the AEGLE Big Data platform will operate on integrated sources that will be able to address complex clinical questions and scenarios associating phenotypic data with personal genetic profiles. Ultimately, the AEGLE conceptual and methodological platform will enable to bridge the gap from improved molecular characterization to improved and rational management of CLL/ In addition, it will offer the possibility of proposing and evaluating health interventions towards the goal of integrated care e.g. identifying groups with specific profiles that will be considered as eligible or ineligible for certain treatments and, at the same time, evaluating the cost of this intervention; obtaining accurate information that will help in the design of clinical trials.

3.2 Intensive Care Unit (ICU) Big Bio-signal streams

In an Intensive Care Unit (ICU) context, patient bio-signals are continuously monitored and displayed towards recognizing alerting events. The continuous recordings of clinical, laboratory data and physiologic waveforms could be analysed and displayed in an easy-to-understand manner by clinicians, who cannot interpret more than 3-4 variables at any given time. A basic principle for such analysis is that these signals cannot really be regarded independently, as they express systems and their interactions in stable or deteriorating states. Also in this particular situation, where the human factor steps in with inadequacy, sound critical decisions cannot simply be made without risks. AEGLE scalable data analytics will provide automated analysis of the fast changing multi-dimensional functions of variables for the detection of unusual, unstable or deteriorating states in patients.

Advanced data analytics shall be employed under the AEGLE for the critical detection and prevention of patient states to deterioration to eventually save lives and reduce the rates of mortalities in hospital ICUs. AEGLE analytics infrastructure for sepsis diagnosis and prediction in ICU shall promote the development of large data bases for variability and complexity analysis of physiological data (such as heart rate, blood pressure and temperature) captured from monitors in real time and alerting clinicians when the patients clinical status may deteriorate, particularly in the early stages of life- threatening illnesses. In this respect, early and personalized treatment will be feasible using AEGLE technology for higher survival in ICUs around European Hospitals.

3.3 Multi-parametric Management for Type II Diabetes

The risk of developing T2D can be increased by various factors; usually a mixture of modifiable and non-modifiable elements: i) age, ii) weight, iii) genetics and iv) ethnicity [10], [11]. Based on the risk factors and screening measurements outlined, and their variety from biochemical data, ultrasound, to general demographic data it is necessary to develop a scalable infrastructure such as the AEGLE platform that will process streams of distributed and heterogeneous biomedical, imaging and demographic Big Data, while offering multi-parametric data analytics services. AEGLE infrastructure will cope with important computational issues generated by: a) the complex and heterogeneous, huge volumes of data and, b) the inherent complexity of the multi-parametric data analyses to provide meaningful data to multi-level of healthcare professionals, researchers to inform diabetes early diagnoses, screening and counselling. It must be noted that high- throughput data are usually provided in semi-structured or unstructured format and hence the AEGLE platform will need to provide a multi-step processing modification and analysis of data in order to be operational for the end users. The processing of all the above prognostic and complication indicators in their variety should enable the early detection, treatment and modifications of complications developing in type II diabetes, leading to a decrease in morbidity, mortality and excessive health care costs.

The Diabetic Databases already available can be utilised for early AEGLE development. The process will take place mainly in AEGLE cloud. More specifically, the AEGLE system will analyse the inter dependences of the variables that are known to have a detrimental effect in type 2 diabetes to give a prediction on the potential deterioration - this would enable intervention to enable reduction of mortality, complications and hospitalization which would all lead to reduction in overall health costs.

4 The AEGLE Infrastructure

Figure 2 depicts the main building blocks of AEGLEs big data analytics framework. Reflecting the requirements of different stakeholders involved in the full data value chain for healthcare analytics, the AEGLE framework consists of big data analytics services at two levels:

- **Local level:** The local level implements big data analytics services for real-time processing of large volumes, fast generated and multiple-formatted raw data originating from patient monitoring services deployed within a healthcare unit, complemented with dedicated medical databases. An example is the real-time analytics service that AEGLE will implement over an HPC substrate for the scenario of Intensive Care Unit (ICU). The goal of the analytics service is to detect unusual, unstable or deteriorating states of patients given the fast changing multi- dimensional variables conveyed within the bio-signals generated by ICU dedicated equipment. The stakeholders of the local level analytics are healthcare units/systems and of course the patients a ultimate beneficiaries that will benefit from the advanced treatment modalities enabled by adopting the analytics services. For example, in the ICU scenario, a prompt reaction to detected instabilities or abnormal behaviour of the patients status could significantly help to save the lives of patients being treated within the ICU.
- **Cloud level:** The cloud level analytics services will offer an experimental big data research platform to data scientists, workers and data professionals across Europe. The platform consists of a large pool of semantically-annotated and anonymized healthcare data, a set of libraries implementing state-of-the-art big data analytics methods including the local level big data analytics AEGLE services and APIs for federating with public and private data sets. Advanced visualisation tools will be implemented by AEGLE as an instrument for gaining new knowledge and expertise, advancing the European know-how in healthcare big data analytics, by allowing data scientists to steer the cloud level analytics mechanisms with their own insights. SMEs across Europe will be given the ability to use of the AEGLE platform (according to the business model that AEGLE will define) in order to deploy and

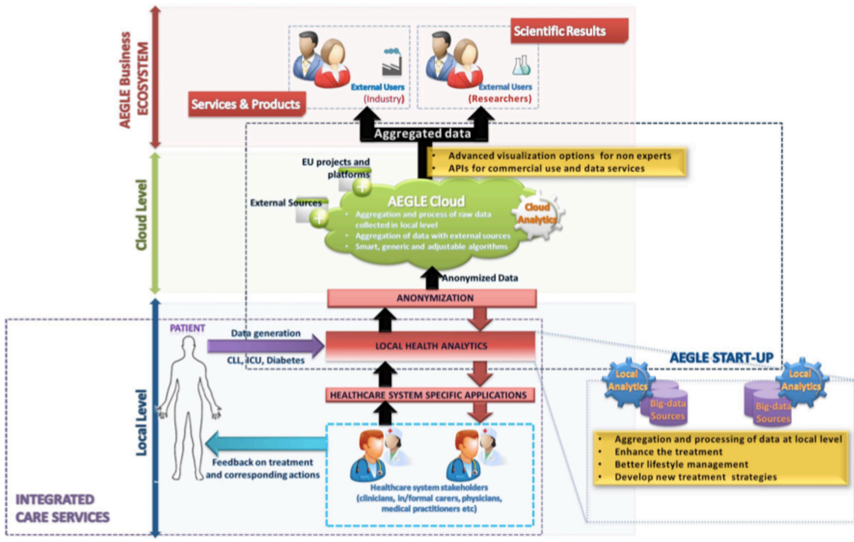

Fig. 2. AEGLE infrastructure.

assess the validity of their innovative data analytics solutions which aim at creating new value in the field of healthcare.

4.1 Exploiting Heterogeneous High Performance Reconfigurable Architectures

Over the past few years, high performance computing environments have increasingly embraced hardware heterogeneity as a means to offering improved performance and particularly, performance/Watt. Rather than a homogeneous collection of a single-type of processing element (i.e. single core CPUs), heterogeneous systems utilize multiple types of resources allowing workloads to be partitioned across those resources depending on the best match of the type of processing to the capabilities of the processing unit. A typical heterogeneous system will contain conventional multi-core processors providing a general purpose computing capability and one or more type of co-processors such as GPUs, DSPs or FPGA-based Dataflow Engines (DFEs) which are optimized for parallel workloads.

Most hardware systems consist of rackmount servers, where each server contains some number of CPUs and PCI Express attached co-processor cards such as GPUs or DFEs. This type of system is simple to deploy but can be hard to optimize since either CPUs or co-processors are often under-utilized due to the fixed ratio of resources within the chassis. An alternative deployment methodology is represented by the MAXELER MPC-X Series [12], which decouple DFEs from CPUs by having separate CPU servers and DFE systems connected via a high speed network. This method achieves somewhat lower bandwidth between the CPUs and DFEs but at the benefit of significantly increased flexibility of task assignment which can lead to much improved performance and efficiency.

MAXELER high-performance dataflow solutions are designed to integrate into production server environments, supporting standard operating systems and management tools. Mature software tools are provided by MAXELER for dataflow oriented application customization, i.e. MaxIDE and MaxCompiler programming tool suite [13]. In addition, MAXELERs dataflow computing platforms form promising infrastructures for big data problems, with large on-board memories backed by high performance computation capabilities. For example, the MPC-X series dataflow nodes provide up to 384GB of DRAM and ultra-high speed connectivity to other nodes, allowing problems that might normally use disk or multiple nodes to run in memory on a single node. In addition, DFEs can incorporate lossless or lossy data compression into application data flows, directly multiplying memory capacity and bandwidth and allowing terabytes of data to be held in memory within a single node.

Despite growing adoption in narrow contexts, heterogeneous infrastructures are not widely used for the kind of Big Data analytics applications that will be tackled in AEGLE. This is due to a number of factors, including that not all analytics algorithms map well to co-processors (and particularly to GPUs, which are optimized specifically for floating point arithmetic), the increased programming difficulty in a heterogeneous environment, complexities of managing multiple resource types and the prevalence of a simplified view of the Map Reduce programming model which maps well to homogeneous processors. AEGLE will address the final steps necessary to enable heterogeneous HPC infrastructure to have a real impact on healthcare. In particular, we will utilize MAXELER DFEs which are based on reconfigurable hardware and can be optimized for specific algorithms and the project brings together both algorithm experts and computational experts to ensure that the analysis applications are efficiently implemented on the target technology. The project will develop optimized implementations of the core algorithms on the MAXELER platform, as well as developing the technology necessary to manage a large-scale infrastructure of heterogeneous systems.

5 Big Data Analytics Acceleration

5.1 State-of-Art Technological Approaches for Big Data

BigData healthcare analytics operate on collections of large and complex data sets which are difficult to process using common database management tools. Improving the performance of Big Data analytics has currently gained a lot of attention in industry and academia. Especially in healthcare systems, analytics acceleration is expected to form the main enabler for making decisions faster, analyse bigger data sizes and processing live data streams under real-time constraints.

Several techniques and tools have been emerged for Big Data acceleration to address the increased complexity of applications requirements. Widely used for BigData processing, MapReduce is a distributed data processing framework for large clusters built of commodity servers. MapReduce frameworks are exploiting the inherent parallelism found in data driven applications, by splitting computation to set of parallel jobs (mappers) that generate intermediate results and propagate them to the reduction tasks to produce the final outcome. Google originally proposed and developed the MapReduce framework to process their proprietary Big Data [14]. Apache Hadoop [15] is open source MapReduce software that has emerged as the de facto standard of MapReduce for big data processing. Typical MapReduce frameworks target cluster based organizations of computing working nodes, thus assuming that each working nodes has its own address space. These computing fabric organizations usually impose a communication bottleneck between the maps and reduce phase. The Phoenix framework [16] proposes a runtime environment for supporting MapReduce on shared-memory multicore processors, taking advantage of the physical proximity of the computing and memory resources.

At the data management level, in-memory databases have gained a lot of attention in the field of Big Data analytics. Modern commodity servers equipped with 1 TByte main memory are currently available for a very reasonable price. An in-memory database [17], [18] relies on main memory storage mechanisms rather than disk storage, in order to eliminate seek time when querying the data, thus providing faster and more predictable performance. In- memory databases are faster than disk-optimized databases since the internal optimization algorithms are simpler and execute fewer CPU instructions. Recently, IBM has launched DB2 with BLU [19] acceleration database infrastructure to speed up BigData analytics using dynamic in-memory columnar technologies. Unnecessary processing is eliminated by data skipping that automatically detects and skips large sections of data that do not qualify for a query. In-memory columnar technologies provide an efficient way to scan and find relevant data. BLU acceleration utilizes instruction level parallelism through SIMD vector processing techniques to pack instructions and execute them in a single time slot, thus improving processing efficiency. In general, there is a growing interest in accelerating data management and event stream processing with devices exhibiting high capabilities on parallel resources. Within the context of exploiting parallelism for query acceleration, database management solutions exploiting GPU [20] and FPGA [21], [22] implementations of the basic query structures have been also emerged.

5.2 Dataflow Reconfigurable Computing for Big Data Analytics

Within AEGLE, we target to go beyond state-of-art on Big Data services by introducing an integrated infrastructure that exploits FPGA-based dataflow acceleration across three different software levels, i.e.:

- the algorithmic level,
- the MapReduce runtime level and
- the storage and data management level.

More specifically, at the algorithmic level, customized DataFlow Engines (DFEs) will be developed in order to accelerate the computation intensive kernels found in the targeted Big Data analytics procedures. At this level, acceleration is achieved through the well-known hardware-software co-design paradigm [23], in which the software kernels that are utilized frequently and exhibit high computation demands will be designed as hardware DFEs and mapped to MAXELERs device. Advanced compiler- and datapath-level optimization techniques [30], [31] will be adapted for spatial computing with DFEs. Along with per accelerator datapath optimization, maximization of accelerator's scalability issues given the FPGA's computational and memory organization constraints [28] will be a considered.

At the runtime level, specialized DFEs will be designed targeting to the acceleration of the underlying MapReduce programming model, i.e. the map, combine and reduce functions. MapReduce allocates several resources from the software processors, reducing the overall performance of the Big Data application. In this case, acceleration targets to the efficient implementation of internal procedures found in MapReduce runtime. Early experimental analysis considering the implementation of a MapReduce accelerated framework for FPGAs, showed speedup gains up to 32x in respect to a purely software solution [24].

In addition, customized memory management schemes tailored to the memory hierarchy and organization of MAXELERs devices will be incorporated to efficiently handle the large number of key-value pairs usually generated by MapReduce semantics, as well as platform specific task schedulers for balancing the load across the software processors and the DFEs. Previous work on customized dynamic memory management for multi-threaded workloads showed that significant performance and memory footprint gains can be achieved, when careful design of customized [25], platform dependent [26] and adaptable [27] memory management mechanisms are performed.

Finally, regarding the storage and data management level, the database management system (DBMS) would be extended to support both adaptive data layout optimizations, e.g. columnar versus row-wise storage model according to the type of queries, and query-specific hardware pipelines dataflow-based acceleration. There is a lot of calculation intensive operations that are executed by the DBMS to maintain the stored data e.g. merging the update buffer into the main storage of an in-memory column store, that can be efficiently accelerated though dataflow-based FPGA acceleration. Regarding the DBMS acceleration for data analytics, MAXELERs in-memory capabilities is expected to fulfil the needs for high demanding and fast data retrieval. Several DFEs organizations and design options will be investigated in order to tailor their hardware architecture of DFEs to the set of most demanding queries.

6 Expected scientific and societal benefits

Business growth and technology advancements have resulted in growing amounts of enterprise data. To gain valuable insights and competitive advantages, there is a growing demand for performing fast and in some cases, e.g. responsive health-care, real-time analytics on such data. AEGLE project invests on effective acceleration of such workloads, by incorporating the latest advances in hardware, i.e. HPC computing nodes to support parallelization and heterogeneous data-flow acceleration on reconfigurable devices to perform several tasks in just one instruction. Data- flow acceleration has been already proven extremely beneficial for several domains, e.g. financial analysis, oil and gas discovery analytics [32]. Within AEGLE it is foreseen to extent the scope and the benefits of acceleration also in domains related with big bio and health data and database management. It is expected that in this era of big data, the methodologies and tools developed within AEGLE project will transform to an extremely helpful tool-suite not only for health practitioners and researchers but also for business and IT leaders across all industries that are looking for ways to easily and cost-effectively unlock the value of enterprise data and are trying to quickly implement new solutions to gain additional insight from this data to improve outcomes across all areas of the business and science.

Additionally the outcomes of this project that would comprise the AEGLE platform shall impact the biomedical informatics domain in providing new and fast tools to the end users for performing high throughput analysis, as well as provide to the end users a solid and efficient framework for developing clinical studies (observational and/or interventional) for reliable medical decision support, and for the production of evidence based medicine incorporating information from the social media. It would impact the issue of accessing data from various inputs and the evolution of an open health data virtual space comprising of all the personalized and translational medicine related data. Thus, the impact to the ICT world is expected to be significant since health is one of the most important social and financial areas for R&D and innovation, and further, the S/W products are expected to be useable to a large extend to other big data that are related with the economy, environment, energy, behavior and psychology.

Regarding the healthcare domain AEGLE will provide more advanced mechanisms for analysis of data, while offer predictive modelling especially in critical situations. This could lead in short term evaluation of patients condition estimation of evolution for any type of illness and at the same time facilitate doctors role by providing significant assistance via pattern recognition of the severity and the countermeasures for patients condition. The social benefits in terms of integrated care and the use of big-data are the following:

- Improved interaction between patients, their relatives and carers, facilitating more active participation of patients and relatives in care processes
- Improved cooperation between the providers of health, social and informal care
- Reinforced medical knowledge with respect to efficient management of comorbidities
- Increased confidence in decision support systems for disease/patient management
- Increased level of education and acceptance by patients and care givers of ICT solutions for personalised care
- Reduced admissions and days spent in care institutions, improved disease management and treatment at the point of need, actual improvements in the daily activities of patients through the effective use of ICT and the better coordination of care processes

7 Conclusions

In this paper, we presented the AEGLE approach for enabling high performance Big Bio-Data analytics. AEGLE aims to efficiently integrate cloud computing together with heterogeneous high performance computing technologies to enable both a publicly available global medical repository for wide adoption within the healthcare research , as well as to support fast analysis for aiding medical decisions at the local level of intervention. Here, we focused our analysis on the advanced reconfigurable technologies to be exploited throughout the AEGLE project. After a brief introduction of the considered Big Data health scenarios, we described the overall cloud- and local-level AEGLE infrastructure and the high performance reconfigurable architectures employed. We reviewed the state-of-art approaches supporting high performance Big Data analytics, and we then discussed the planned activities that will enable accelerated Big Data services through effective use of reconfigurable dataflow engines. We concluded our presentation by discussion the expected scientific and societal impact of the AEGLE approach.

References

1. Transforming health care through big data, Institute for Health Technology Information. Available online at http://www.ihealthtran.com/big$_{d}ata_{i}n_{h}ealthcare.html$
2. Alex Pentland, Todd G. Reid and Tracy Heidbeck, BIG Data and Health, Revolutionizing Medicine and Public Health, Report of the Big Data and Health Working Group 2013
3. http://ec.europa.eu/information$_{s}ociety/newsroom/cf/dae/document.cfm?doc_{i}d = 3488$
4. http://ec.europa.eu/digital-agenda/news/digital-agenda-eu-task-force-advise-how-promote-ehealth-help-patients-and- healthcare-systems
5. http://ec.europa.eu/digital-agenda/en/news/big-data-what-it-and-why-it-important
6. http://healthstartup.eu/2012/05/top-big-data-opportunities-for-health-startups/
7. http://ec.europa.eu/information$_{s}ociety/newsroom/cf/dae/document.cfm?doc_{i}d = 3488$
8. Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux et al. NCBI GEO: archive for functional genomics data sets update. Nucl. Acids Res. (1 January 2013) 41(D1): D991-D995
9. Yuichi Kodama, Martin Shumway et al. The sequence read archive: explosive growth of sequencing data. Nucl. Acids Res. (2012) 40(D1): D54-D56
10. Diabetes UK. Awareness Campaign. [Online]. Available from: http://www.diabetes.org.uk/awareness-campaign
11. National Institute for Health and Care Excellence. Preventing type 2 diabetes: risk identification and interventions for individuals at high risk. NICE public health guidance 38 [Online] 2012. Available from: http://www.guidance.nice.org.uk/ph38
12. Maxeler Technologies, http://www.maxeler.com
13. MaxCompiler, http://www.maxeler.com/products/software/maxcompiler/
14. J. Dean and S. Ghemawat, MapReduce: Simplified Data Processing on Large Clusters, Proc. Sixth Symposium on Operating System Design and Implementation (OSDI04), pp.137-150 (2004).
15. Apache Hadoop, online available ¡http://hadoop.apache.org¿
16. Richard M. Yoo, Anthony Romano, Christos Kozyrakis, ”Phoenix rebirth: Scalable MapReduce on a large-scale shared- memory system,” 2013 IEEE International Symposium on Workload Characterization (IISWC), pp. 198-207, 2009 IEEE International Symposium on Workload Characterization (IISWC), 2009.
17. M. Grund, J. Kru ger, H. Plattner, A. Zeier, P. Cudre-Mauroux, and S. Madden, HYRISE - A Main Memory Hybrid Storage Engine, in VLDB, 2012.
18. M. Stonebraker et al, The End of an Architectural Era (Its Time for a Complete Rewrite), in VLDB, 2007.

19. Vijayshankar Raman, Gopi Attaluri, Ronald Barber, Naresh Chainani, David Kalmuk, Vincent KulandaiSamy, Jens Leenstra, Sam Lightstone, Shaorong Liu, Guy M. Lohman, Tim Malkemus, Rene Mueller, Ippokratis Pandis, Berni Schiefer, David Sharpe, Richard Sidle, Adam Storm, and Liping Zhang. 2013. DB2 with BLU acceleration: so much more than just a column store. Proc. VLDB Endow. 6, 11 (August 2013), 1080-1091.
20. Peter Bakkum and Kevin Skadron. 2010. Accelerating SQL database operations on a GPU with CUDA. In Proceedings of the 3rd Workshop on General-Purpose Computation on Graphics Processing Units (GPGPU '10). ACM, New York, NY, USA, 94-103.
21. Christopher Dennl, Daniel Ziener, and Jurgen Teich. 2012. On-the-fly Composition of FPGA-Based SQL Query Accelerators Using a Partially Reconfigurable Module Library. In Proceedings of the 2012 IEEE 20th International Symposium on Field-Programmable Custom Computing Machines (FCCM '12). IEEE Computer Society, Washington, DC, USA, 45-52.
22. Mohammadreza Najafi, Mohammad Sadoghi, and Hans-Arno Jacobsen. 2013. Flexible query processor on FPGAs. Proc. VLDB Endow. 6, 12 (August 2013), 1310-1313.
23. Rajesh K. Gupta and Giovanni De Micheli. 1993. Hardware-Software Cosynthesis for Digital Systems. IEEE Des. Test 10, 3 (July 1993), 29-41.
24. Yi Shan, Bo Wang, Jing Yan, Yu Wang, Ningyi Xu, and Huazhong Yang. 2010. FPMR: MapReduce framework on FPGA. In Proceedings of the 18th annual ACM/SIGDA international symposium on Field programmable gate arrays (FPGA '10).
25. Xydis, S.; Bartzas, A.; Anagnostopoulos, I.; Soudris, D.; Pekmestzi, K., "Custom multi-threaded Dynamic Memory Management for Multiprocessor System-on-Chip platforms," Embedded Computer Systems (SAMOS), 2010 International Conference on , vol., no., pp.102,109, 19-22 July 2010
26. Anagnostopoulos, I.; Xydis, S.; Bartzas, A.; Zhonghai Lu; Soudris, D.; Jantsch, A., "Custom Microcoded Dynamic Memory Management for Distributed On-Chip Memory Organizations," Embedded Systems Letters, IEEE , vol.3, no.2, pp.66,69, June 2011
27. S. Xydis, I. Stamelakos, A. Bartzas, and D. Soudris. Runtime tuning of dynamic memory management for mitigating footprint-fragmentation variations. In Proc. of PARMA Workshop, pages 2736. VDE Verlang, 2011.
28. Dionysios Diamantopoulos, Sotirios Xydis, Kostas Siozios and Dimitrios Soudris, Dynamic Memory Management in Vivado-HLS for Scalable Many-Accelerator Architectures, in Proceedings of 11th International Symposium on Applied Reconfigurable Computing (ARC 2015), Springer Lecture Notes in Computer Science, 2015.
29. Endert, Alex, M. Shahriar Hossain, Naren Ramakrishnan, Chris North, Patrick Fiaux, and Christopher Andrews. "The human is the loop: new directions for visual analytics." Journal of Intelligent Information Systems (2014): 1-25.
30. Xydis, S.; Palermo, G.; Zaccaria, V.; Silvano, C., "SPIRIT: Spectral-Aware Pareto Iterative Refinement Optimization for Supervised High-Level Synthesis," Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on , vol.34, no.1, pp.155,159, Jan. 2015
31. Sotirios Xydis, Kiamal Pekmestzi, Dimitrios Soudris, and George Economakos. 2013. Compiler-in-the-loop exploration during datapath synthesis for higher quality delay-area trade-offs. ACM Trans. Des. Autom. Electron. Syst. 18, 1, Article 11 (January 2013), 35 pages.
32. O. Lindtjorn, R. G. Clapp, O. Pell, O. Mencer, M. J. Flynn and H. Fu, Beyond Traditional Microprocessors for Geoscience High-Performance Computing Applications, IEEE Micro, vol. 31, no. 2, March/April 2011.