SERVICES AND INFRASTRUCTURE FOR VIRTUAL RESEARCH ENVIRONMENTS – FOR USE BY SCIENCE AND BUSINESS

Assoc. Prof. Emanouil Atanassov PhD.[1], Prof. Aneta Karaivanova PhD.[1], Assoc. Prof. Todor Gurov PhD.[1]
Institute of Information and Communication Technologies – Bulgarian Academy of Sciences, Bulgaria [1]

(emanouil,anet,gurov)@parallel.bas.bg

Abstract: *The European Commission has supported substantial investment in building and integrating electronic research infrastructures comprising of High Performance Computing resources and data warehouses with large storage capacities and advanced features. The Bulgarian supercomputer system Avitohol has been deployed and reached number 332 in the Top500 list in July 2015 with a theoretical performance of 416 TFlop/s. In order to maximize the use of such expensive equipment it is necessary to deploy an integrated set of services that provide flexible and transparent access. In this paper we describe the main services that provide a virtual research environment that fulfils these needs. We discuss the main features of such services and their benefits for industrial applications, especially when SMEs collaborate with HPC centres.*

Keywords: HIGH PERFORMANCE COMPUTING, CLOUD, VIRTUAL RESEARCH ENVIRONMENT

1. Introduction

The European Union invests heavily in the development of electronic research infrastructures, based on supercomputers and HPC clusters, interconnected using open-source software and providing a rich set of services.

Recently a supercomputer using state-of-the-art hardware has been deployed at the Institute of Information and Communication Technologies of the Bulgarian Academy of Sciences. Avitohol comprises of 150 servers, each one equipped with dual Intel Xeon E5-2640 v2 CPUs, 64 GB RAM, two Intel Xeon Phi 7120P co-processors with 16 GB RAM each, all of them interconnected with non-blocking low-latency InfiniBand. The theoretical performance of the system is 412.3 TFlop/s in double precision, while the actual tests that put it in the top500 list achieved 264 TFlop/s performance [1, 2].

The VI-SEEM project [3] aims to deliver integrated Virtual Research Environment, including supercomputers like Avitohol, HPC clusters and advanced storage facilities. The next table show the computing and storage resources provided for the project:

Country	Grid			Cloud		HPC				Storage		
	Sites	Site Cores		Sites	Site VMs	Sites	Site Cores			Sites	Site TBs	
		CPU	GPU				CPU	GPU	Phi		Disk	Tape
Greece	6	851	0	1	10 000	1	8 520	0	0	1	1 000	3 000
Cyprus	0	0	0	0	0	1	1 392	16 128	0	1	260	0
Bulgaria	1	2 400	0	1	2 400	1	2 400	0	18 300	1	20	0
Serbia	3	832	0	0	0	1	1 696	54 272	0	1	96	0
Hungary	0	0	0	0	0	2	2 112	129 024	0	2	1 087	6 800
Romania	0	0	0	1	400	2	4 496	3 136	0	1	50	0
Albania	0	0	0	0	0	1	144	0	0	0	0	0
Bosnia	0	0	0	1	60	0	0	0	0	1	1	0
FYROM	1	768	0	1	436	1	768	0	0	1	36	0
Montenegro	1	32	0	0	0	0	0	0	0	0	0	0
Moldova	1	24	0	1	12	0	0	0	0	1	4	0
Armenia	1	128	0	1	96	1	128	2 880	0	1	8	0
Georgia	1	64	0	0	0	0	0	0	0	0	0	0
Egypt	0	0	0	0	0	1	1 040	0	0	1	5 200	0
Israel	1	60	0	1	560	0	0	0	0	2	50	0
Jordan	4	120	20 576	0	0	0	0	0	0	0	0	0
TOTAL	20	5 279	20 576	8	13 964	12	22 696	205 440	18 300	14	7 812	9 800

Table 1: *Computing and Storage Resources in VI-SEEM*

Some of the applications within the project foresee industrial use. For example, the drug research-related applications aim to develop new workflows for drug development and testing using compute power. The research in the domain of climate change and its impact on the environment can be important for making business plans in the medium to long term, in domains like agriculture or energy production.

The services within the project are developed in several layers. First of all we have the raw hardware, accessible directly or through Grid or Cloud middleware. Then we have the application-specific services that are built on top of it.

In order to build and maintain an integrated infrastructure that spans multiple administrative domains and provides access to heterogeneous resources, it is necessary to adhere to common policies and to accept a single system for authentication/authorization and accounting. In the next chapter we proceed to describe the services that we have used for the creation of Virtual Research Environment (VRE) of the VI-SEEM project. Then we give more information about the policies and procedures that are used by the leading European HPC centers when dealing with industrial users and applications. We conclude with the benefits of using infrastructure and services provided by these centers compared with alternative approaches.

2. Services forming a Virtual Research Environment

When establishing electronic infrastructures for the needs of scientists it is customary to use open-source software and tools and to adhere to established protocols in order to allow the infrastructure to grow in time following the evolving needs of researchers and the new hardware and software developments. The need to protect the investment from past projects and activities dictates certain conservatism. That is why the authentication and authorization framework adopted in the VI-SEEM VRE has several components that deal with different technologies and practices already tested in the communities.

The figure below shows how different sources (stand-alone Identity Providers, eduGAIN Identity Providers, VI-SEEM VHO) may provide authentication information, different sources may provide authorization (VI-SEEM VHO, HEXAA) and then through VI-SEEM Proxy service this information is forwarded to the actual services. Web-based services consume this information directly, while the services that access large-scale hardware resources directly use the Token Translator.

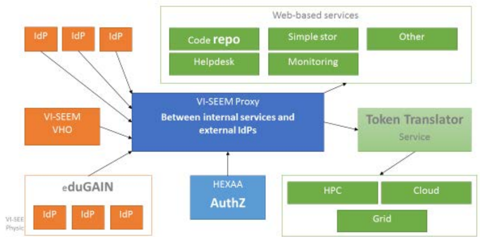

Figure 2: *Authentication information provided by difference sources*

The underlying hardware infrastructure consists of a number of supercomputers, high-performance clusters with advanced interconnects as well as some more generic clusters, mainly providing resources for Grid jobs. In the next table one can see the resources provided and the dedication to the project.

Most of the applications within the project require HPC resources for computation, as opposed to Grid and Cloud resources. In practice this means that the developers obtain direct user-level access to the clusters or supercomputers. However, the applications that require large amounts of storage in most cases are not interested only in the raw storage, but also in advanced interfaces for storing, accessing and querying data.

The storage services deployed in VI-SEEM follow practices widely accepted within the Cloud provider community. The Simple Storage services is oriented towards individual researchers, providing the substantial amount of storage useful mostly for documents and codes and to a limited extent for storing actual research data. It is expected that most of the scientific data will be stored at the VI-SEEM repository, where each research team will have up to 10 TB storage capacity. The distributed nature of the implementation of these services allows for failover and redundancy. Long term storage is provided through the archival service, which shall ensure data will remain available for validation or re-use for 5 or more years.

Some application domains have established paradigms and workflows that lead to the creation of domain-specific services that provide better user experience with easy-to-use interfaces and hide the complexity of the underlying computational and storage infrastructure. Because of the limited resources of the project a small initial set of such services has been deployed.

For example, the ChemBioServer can be used for computer-aided drug design with Schrodinger Suite. When performing virtual screening for computer-aided drug design and application with the Schrodinger Suite software, the generated virtual screening results are then post-processed (filtered) for toxicity, chemical similarity and validity of docking pose using the ChemBioServer software.

3. Policies and procedures for accessing the Virtual Research Environment

The idea of the Virtual Research Environment is to create an integrated environment that serves the needs of a wide set of researchers. The innovative potential of such virtual labs makes them interesting for industry, especially for drug research and research related to agriculture and the impact of climate change. The researchers can obtain two types of access. Those that apply and obtain access through the Open Calls run by the project will get direct access to the resources and services and will be able to request modifications and extensive support. The other scientists will get access through web portals and other similar interfaces, so that they would benefit from running the applications and workflows, but without direct control of the execution. In this way the use of the infrastructure and the services developed is maximized and the technology barrier to using HPC for scientific research is lowered, as the hard work is done by a few dedicated developers and system administrators.

However, since large parts of the VRE framework are domain-agnostic, they are also useful for solving problems related to engineering and manufacturing. It should be noted that within the European Union there is considerable expertise in offering access to HPC and data resources, providing consultancy and similar services, especially for small and medium enterprises (SMEs). The SESAME Net is a network of HPC competency centres [4] that strive to improve their policies and procedures so that the SMEs can benefit from HPC technology following a stream-lined process.

Contrary to our experience with scientists who seem to prefer direct access to the HPC resources in order to have full control of the running of their applications, we observe that SMEs prefer to use Cloud-type interfaces because of the improved security and perhaps also because of their acquired expertise in using Cloud technology in general. That is why many of the HPC competency centres provide Cloud access.

While not all use cases actually require the SME to manually perform computations on the HPC system, we describe a common way of providing direct access.

Typically, the HPC centre will provide access to the system at user-level, meaning without any kind of administrative rights, and through an interactive command-line interface for general purpose access. The most typical scenario would proceed as follows: An account request is submitted to the HPC centre via a web interface or similar. The account is validated and confirmed. A username and password or alternative login credentials are provided to the user. The user can then login to a server which is connected to the Cluster (login node) via a program implementing the Secure Shell (SSH) protocol, also called an SSH client. SSH client software is available for all popular operating systems, including mobile systems such as Android. The SSH protocol also accommodates file transfers, so data can easily be transferred to and from the HPC system. Software to provide more comfortable file transfers through SSH exists for several operating systems as well, e.g. the free WinSCP software for Windows. Optionally, to use programs with a graphical user interface on the HPC system, the client computer will need the installation of additional software. The HPC systems typically use the "X Window System", which requires a counterpart on the client PC to function over network connections. The method of displaying a GUI window of an application running on the remote server on your PC is called "X-forwarding", and typically needs to be enabled in the SSH client. Usually Linux PCs provide this functionality out-of-the-box, and there are free applications that provide such functionality on MS Windows. Once connected to the HPC system, the user can interactively discover what kind of software packages and libraries are installed. If the particular software intended to be used is not readily available on the HPC system, it can be deployed by the HPC centre's staff. Typically, the centre provides web-based documentation, as well as a support helpdesk for these kind of questions and general help.

Because the HPC systems may support numerous users at the same time, they usually provide a way to "enqueue" the execution of programs in a kind of waiting line system. A single call of an application in such a waiting system is then called a "(compute) job" or old fashioned "batch job". Applications that require real-time response can be accommodated by the HPC centre by special reservations.

4. Benefits from using HPC technology for SMEs

We studied many use cases of successful use of HPC technology by SMEs, mainly in the engineering and manufacturing domain, and we could summarize the main benefits of using these technologies as follows:

4.1 Need for large computing power and storage capacity:

Some SMEs already experience the lack of computing power and storage on their enterprise infrastructure to run simulations or other demanding computer applications. Broadly speaking, HPC technology is useful in all cases where large amounts of computations are necessary or large amounts of data need to be processed. Even when a compute infrastructure can easily be extended, it will be difficult to reach significant performance in-house compared to the capacity of HPC centres. However, most of the advantages of HPC are related to the kinds of computations that cannot be accomplished on a regular infrastructure. Due to the advancements in applied sciences, accurate simulations and modelling of new products and processes increasingly require HPC-type equipment. The size of the HPC market is projected to grow from USD 28.08 Billion in 2015 to USD 36.62 Billion by 2020, at a Compound Annual Growth Rate (CAGR) [5] of 5.45%. The use of HPC technologies is a testament of the innovative character of the SMEs and brings substantial Return On Investment (ROI). Various studies have been performed on the ROI for HPC. One pilot study from IDC indicates that one can expect $356.5 on average in revenue per dollar of HPC invested and $38.7 on average of profits (or cost savings) per dollar of HPC invested [6]. The European project EESI concluded that HPC has clearly added great economic value across Europe and the ROI results in an average of $867 revenue dollars generated for each dollar invested in HPC in this study and $69 dollars in profits (or in cost savings) for each dollar invested in HPC [7]. Obviously, the actual results will depend on the type of industry and type of use of HPC technology, but the potential is obviously interesting. It is important to understand that there are synergetic effects for a country's economy from increased use of HPC.

4.2. Need for shorter development cycles, e.g., because of competition on the market.

Typically, new products are developed with a combination of physical experiments and simulations. Having large computing power for large and complex simulations opens possibilities for testing new ideas, innovation and new products in pharmacy, production, automotive components, etc. By affording to run multiple simulations with different input parameters, HPC systems for example offer the possibility to adapt the simulation closer to the experiments or to optimize products through simulations. Large industrial enterprises, e.g. from the automotive industry, are said to have operated their own HPC clusters since a decade. HPC centres like the ones collaborating in SESAME Net also enable small and medium enterprises to access such powerful resources. HPC can add tremendous value to any SME who develops products via:

- Creation of large, high-fidelity models that yield accurate and detailed insight into a design's performance.
- Extremely accurate simulations to predict real-world conditions to convince a customer.
- In engineering HPC enables high mesh densities for improved accuracy, numerous geometric details, or sophisticated treatment of physical phenomena.

HPC is not limited to a few commercial sectors, non-traditional and smaller companies are coming around to the idea that HPC can have a major impact on their business. Among others you will find use cases of the film industry or from the growing number of start-ups in the field of bio-informatics. Last but not least, there might also be new business perspectives for a company to provide services in the HPC eco system. Examples of innovative products with high impact can be seen in the HPC innovation and ROI awards [8].

4.3. Economic reasons

The use of HPC may reduce costs and/or improve benefits. For example, by reducing designing and prototyping costs; detecting design errors in early stages of product development. If prototyping has a major role in the SME's chain value then HPC would save the costs of building many expensive physical prototypes which now can be built virtually, minimizing the number of physical prototypes to be built.

Since all the decisions taken in an enterprise are intended to have a measurable impact on its economic balance, in order to maximize profit marketing, research, innovation reasons should at last have an economic impact on the enterprise balance, but in this section we are going to focus in the economic reasons (the ones related more directly to economic balance) to integrate HPC as a key aspect in the enterprise value chain.

The three key economic reasons to consider integrating HPC in SMEs (and also in big enterprises) are:

- Cost reduction: Prototyping (wind tunnel, physical testing) and design costs;
- Reducing design times, reducing time to market, (translating this time reduction in money);
- New product/service development that is not possible to develop without HPC.

Cost reduction: HPC can help the SME save costs, for example in prototyping and design procedures, as well as in operational costs. Physical prototyping means to create a physical prototype [9]. This process is very time and cost consuming, because of the time needed to produce an item outside of an assembly line and the cost of prime matters, and this only gets worse when multiple prototypes are needed, especially in case of destructive physical testing when prototypes can only be used once. Physical prototypes can be substituted by virtual prototypes, cheaper to develop and that can be used again and again without destroying them. Virtual prototypes once developed can be used to simulate the physical behaviour or interactions under many circumstances: aero and hydrodynamics, substituting the need of complexity and expensiveness of wind tunnels or fluid tests. Design costs can be reduced using HPC, since HPC brings us the power to design prototypes of products without the need of any physical prototype and begin to acquire knowledge about it in the very early stages of design. This also improves the design process allowing detecting design errors or misconceptions in these early stages of design, where they can be corrected easily and cheaper than in later stages. Operational costs of the SMEs can also be saved by accurate and extensive simulation, modelling and optimization of certain kinds of operations. For example, HPC can be used in optimizing public transport or logistics. It is widely used in the search of oil and gas or in energy production problems.

Reducing design times: In some sectors like aeronautics, engineering or energy, where the interaction of fluids with: aircrafts, shipbuilding, turbines, wind or offshore energy generators is a key factor of efficiency, HPC can be a key to success because it can reduce design times drastically. The times needed to construct a virtual prototype and discover the full interactions with their surrounding or inner contained fluids are much shorter than the time needed to get this knowledge from physical ones. This is an incontestable advantage that can make the difference between a successful and a ruinous project, as well as enable a project that would otherwise never start because HPC possibilities were not on mind or simply were unknown, or because tight schedules are present and standard design times are not possible [10].

4.4. Innovative reasons

Most of the advantages of using HPC technology can be obtained by SMEs that have a substantial Research & Development department, since the access to HPC resources can improve the possibilities to develop innovative products and services. By collaborating with an established HPC Competency Centre, the SME will not necessitate costly on-premise infrastructure. The costs of owning and operating such infrastructure are a heavy burden for most SMEs. By using flexible contracts SMEs can have a low-risk, low-effort, pay-as-you go access to HPC and leverage state-of-the-art software and tools to speedup their development process.

Apart from the cost-related reasons, there are some other main advantages of using HPC:

- obtaining results much faster by employing high amount of compute power;
- achieving a much more accurate representation of the product and its working environment;
- exploring much wider search space for the parameters of the envisaged product.

By collaborating with a leading HPC Competency Centre, the personnel at the SME's Research & Development department acquire precious expertise that is also useful in other, more standard situations. For example, the increased use of multi-core devices in all kinds of products makes knowledge of parallel computing paradigms an important asset. The need to achieve a given result with as little use of energy as possible stresses the importance of using optimal algorithms and scalable programming approaches.

Overall, the use of HPC for completing simulations that would normally run for days or weeks on a desktop workstation, in a matter of hours enables much faster time-to-market and thus first-mover advantage for the SME.

It should not be underestimated that being in close contact with the academia and HPC experts in industry, the SME may obtain advanced knowledge of new developments in HPC technology or applied sciences that will enable cross-fertilization and rapid incorporation of the new technologies in their product portfolio.

4.5. Marketing reasons

An important marketing reason for the use of HPC is to present the enterprise as a cutting-edge technological firm. This has a positive effect in its market position and on the perception of its products and services by customers. And of course, it results in the subsequent positive economic impacts, e.g. more sales, price premiums, and so forth.

More specifically, the marketing reasons to use HPC can be listed as follows:

Leading brands partnerships are a key: If, when selling a product one wants to reach a market where a “brand” or “organization” is well known and with good reputation, it is a good idea to associate it with this leading “brand” or “organization”. So a product gets an improved perception over other products that do not own this advantage. SESAME Net is composed of very well-known and reputable partners, so the collaboration with SESAME HPC partners can lead to a win-win situation, where the new product is associated with a well-known brand and, on the other hand SESAME Net gets an improved visibility.

Best product build with best technology: In our society, most of the products get an improved perception by clients when they are built with the latest technology. Even if the use of this technology only involves a very low percentage of improvement, there are clients that always want to own the “state-of-art” product with the last improvements, while others are more concerned about environment and want to have “green” products, with more efficiency and low power consumption. HPC technology can deliver all these properties, since it is an innovative technology with relatively low use by SMEs because it is not generalized yet. Innovative technologies can lead to innovative products, and for SMEs willing to improve their products it is a competitive advantage. HPC can be applied successfully to improve products in order to make them more “green”. Under the umbrella of FORTISSIMO there are some examples of products optimized using HPC that gained special characteristics that can be valuable for some products or differentiate them from the product´s competence.

4.6. Competitive advantage reasons

Technologically capable firms have inherent advantages compared to other ones. E.g. the reduced time to market enabled by HPC use, allows a bigger manoeuvrability to adapt to changing modern markets.

Being able to bring new products to the market, thanks to the HPC edge on research and development process, allows the SME to gain a competitive advantage over their competitors. This first-mover advantage, thanks to the technological leadership, has several positive impacts on the firm. For example there is brand recognition; this means the automatic association in costumers’ minds of the new type of product with its initial manufacturer. Another example of first-mover advantage is the pre-emption of scarce assets, which allows the first-mover firm which has superior information to purchase assets at market prices below those that will prevail later in the evolution of the market [11].

Additionally, HPC capable firms are able to reduce the product cycle time, - the period for a manufacturer to complete the development and production of a new, or modified, product. It does not only reduce costs, but an improvement in the time to market response also allows for a bigger manoeuvrability to adapt to the changing conditions of modern markets. The use of HPC enables the design of a more accomplished product, with less defects or covering a wider spectrum of consumer needs.

In the economy of today it is increasingly more important to find the consumer niche and to engage the potential clients. With HPC technology the SME can employ advanced algorithms and processes that will increase their market share and reach a wider consumer audience.

By establishing collaboration, possibly through a pilot project, not only with HPC resource providers, but also with Independent Software Vendors and research software providers, application experts and consultants, the SME is immersed in a network with high innovation potential and thus will have increased flexibility and high potential for sustainable development of innovative products [12].

5. Conclusions and recommendations

The modern approach to building integrated electronic infrastructures with a rich set of services is being used extensively for scientific research in Europe. Gradually such services are extended for use by industrial partners to the HPC Competency Centres, especially Small and Medium Enterprises. In order to fully benefit from these possibilities, the SMEs must have substantial Research and Development Teams and be innovation-oriented.

Acknowledgments

This work was supported by the European Commission under H2020 projects - VI-SEEM (Contract Number: 675121) and SESAME Net (Contract Number: 654416).

Bibliography

[1] E. Atanassov, T. Gurov, A. Karaivanova, S. Ivanovska, M. Durchova, D. Georgiev, D. Dimitrov, Tuning for Scalability on Hybrid HPC Cluster, Mathematics in Industry, Cambridge Scholar Publishing, pp. 64-77, 2014, ISBN (10): 1-4438-6401-3.

[2] A. Radenski, T. Gurov, et al., "Big Data Techniques, Systems, Applications, and Platforms: Case Studies from Academia", in: "Annals of Computer Science and Information Systems", Volume 8, pp. 883-888, Sep, 2016.

[3] VI-SEEM project, https://vi-seem.eu/, (retrieved 2016).

[4] SESAME Network partners, https://sesamenet.eu/centres/, (retrieved 2016)

[5] HPC Market by Components Type, Markets&Markets, http://www.marketsandmarkets.com/Market-Reports/Quantum-High-Performance-Computing-Market-631.html, (retrieved 2016)

[6 IDC Analyse the Future, https://hpcuserforum.com/roi.html, (retrieved 2016)

[7] EESI project, http://www.eesi-project.eu/wp-content/uploads/2015/05/EESI2_D7.4_Final-report-on-HPC-Return-on-Investment.pdf, (retrieved 2016).

[8] HPC User Forum Innovation Awards, http://www.hpcuserforum.com/innovationaward, (retrieved 2016).

[9] Blackwell, A. H.; Manar, E., eds. (2015). "Prototype", UXL Encyclopedia of Science (3rd ed.), Retrieved 13 July 2015.

[10] HPC-Cloud-based design of centrifugal pumps: http://www.fortissimo-project.eu/experiments/418, (retrieved 2016)

[11] First-mover advantage, https://en.wikipedia.org/wiki/First-mover_advantage, (retrieved 2016).

[12] Success Stories, https://sesamenet.eu/resources/success-stories, (retrieved 2016).