SME/HPC

Erasmus+

Co-funded by the European Union

Erasmus + KA2: Cooperation for innovation and the exchange of good practices

Knowledge Alliances

SME/HPC

Enabling SMEs to gain competitive advantage from the use of HPC

Material for self-study by staff at HEIs from regions in Ireland, Slovenia and Romania

Prepared by: P2 UNN

Version: 2.0 (Final)

Deliverable: D4/1

SME/HPC

Erasmus+

Co-funded by the European Union

Document reference

Project Acronym	**SME/HPC**			
Project Number	588372-EPP-1-2017-1-IE-EPPKA2-KA			
Project URL	www.smehpc.eu			
EU Project Officer	Silvia.DE-BENEDETTI@ec.europa.eu			
Project Coordinator	**Name**	Prof. Bill O'Gorman	**Affiliation**	WIT
	Email	wogorman@wit.ie	**Phone**	+353.872429537
Document Name	Materials for self-study by staff at HEIs from regions in Ireland, Slovenia and Romania			
Deliverable Number	D4/1			
Type	Document	**Distribution Level**	Public	
Responsible Author(s)	P2: University of Northumbria at Newcastle (UNN)			
Contractual Date of Delivery	Month 16			
Status	FINAL			
Quality assurance readers	Sergio Botelho Junior	Friedrich Meckel		

SME/HPC

Erasmus+

Co-funded by the European Union

Change history

Version	Date	Status	Author (Unit)	Description
0.1	08th April 2019	Working	University of Northumbria at Newcastle (UNN)	Higher Education Institution (HEI) Staff Capacity Building
1.0	30th April 2019	Draft	University of Northumbria at Newcastle (UNN)	Module descriptions are completed, substantially more content is added to the modules in this version
2.0	31st May 2019	Final	University of Northumbria at Newcastle (UNN)	Following updates are made: • Review comments / feedback from concerned project partners accommodated • Added section 3.2.3 illustrating a few HPC case studies specific to SMEs • Section 3.2.1 elaborated with more details on EU SMEs • Other minor corrections and formatting changes
2.1	30th June 2019	Final	University of Northumbria at Newcastle (UNN)	Added names of the Quality Assurance readers

Table of Contents

Introduction

HPC is a rapidly developing field of expertise, where new systems and technologies are emerging with immense speed. HEIs from less developed regions have limited experience of teaching short-term training courses on HPC since their regions are not at the forefront of HPC development and also, owing to their weaker university-enterprise linkages. Therefore, staff capacity building through trainings that address existing gaps in relation to HPC expertise and utilization. This document is designed and developed as part of HEI staff capacity building work package and aims to provide self-study material for training staff of participating SME/HPC HEIs from regions in Ireland, Slovenia and Romania. The self-study material is to be used as a preparation to staff training workshop conducted by UNN and on-the-job experience at Arctur.

Study Learning Objectives

The purpose of this course is to increase participating HEI staffs' knowledge of HPC domain, including the importance of HPC, application areas, technology components and systems, and how HPC could potentially address SME business computing problems.

After completing this self-study course, you will be able to:

- Describe the key concepts of HPC and articulate its importance
- Identify the application areas of HPC
- Recognise the different forms of HPC solutions
- Describe the HPC technology components and HPC architecture
- Identify the key concepts of Cloud Computing and describe how it relates to HPC
- Identify the global market landscape for HPC, including market drivers, trends, segments and key players
- Recognise the key HPC adoption challenges for SMEs
- Provide examples of some regional and international initiatives on HPC for SMEs
- Explain how you can position HPC to address SME computing, product/process development and business needs fostering economic growth and competitiveness

Erasmus+

Co-funded by
the European Union

Organisation of this Study Material

This self-study course is organized in three distinct modules. Module names and the expected time for self-study of the module content are mentioned below. The aims and desired learning outcomes for each module are described at the individual module level.

Module Number	Module Name	Expected Time to Complete
1	Introduction to HPC	2 Hours 30 minutes
2	HPC Building Blocks	2 Hours 30 minutes
3	HPC for Small and Medium Enterprises	3 Hours

SME/HPC

Erasmus+

Co-funded by
the European Union

Module 1: Introduction to HPC

SME/HPC

Erasmus+

Co-funded by the European Union

Aims of this Module:

This module aims to describe / explain:

- What does a High Performance Commuting (HPC) system mean?
- History and evolution of modern HPC
- Why use of HPC is important?
- What are the scientific and business application areas of HPC?

Prerequisites:

You don’t need to have any previous knowledge of High Performance Computing (HPC) but familiarity with basic computing concepts is desired.

Learning Outcomes:

At the end of this module you will be

- able to understand the concepts of HPC
- familiar with the history and evolution of HPC
- able to explain the importance of HPC for scientific and business applications
- able to identify the potential application areas of HPC

Module 1

- ✓ What is High Performance Computing?
- ✓ How is HPC different from regular desktop computing?
- ✓ Why is HPC important?
- ✓ History of HPC
- ✓ Application areas for HPC
- ✓ Reflections on this module learning
- ✓ Test Your Knowledge
- ✓ Recommended Further Readings / Resources

1. Introduction to High Performance Computing (HPC)

1.1. What is High Performance Computing?

High-performance computing (HPC) involves the use of "supercomputers" and massively parallel processing techniques to solve complex computational problems through computer modelling, simulation, and data analysis (Techopedia, 2016). High-performance computing brings together several technologies, including computer architecture, programs and electronics, algorithms, and application software under a single system to address advanced scientific and business computational needs quickly and effectively. HPC technology focuses on developing parallel processing algorithms and software so that programs can be divided into small independent parts, and can be executed simultaneously by separate processors by incorporating both administration and parallel computational techniques. HPC systems have the ability to deliver sustained performance through the concurrent use of computing resources. The terms "high performance computing" and "supercomputing" are sometimes used interchangeably.

1.2. How is HPC different from regular desktop computing?

While a desktop computer / workstation / laptop today typically contains a dual or quad core processor, an HPC system essentially represents a network of processors, each of which contains multiple computational cores as well as its own local memory to execute a wide range of software programs (NICS, 2016). To put it into perspective, a laptop or desktop with a 3 GHz processor can perform around 3 billion calculations per second. While that is much faster than any human can achieve, it pales in comparison to HPC solutions that can perform quadrillions (1 quadrillion = 10^6 billion) of calculations per second.

One of the best-known types of HPC solutions is the supercomputer. A supercomputer contains thousands of compute nodes that work together to complete one or more tasks through parallel processing, which is similar to having thousands of PCs networked together, combining compute power to complete tasks faster. The software programs that coders write to run on supercomputers are divided up into many smaller independent computational tasks, called “threads,” that can be executed simultaneously on these cores. Modern supercomputers can consist of over 100,000 “cores” or more and for supercomputers to operate effectively; the 'interconnection' between the cores must be well designed. For example, Cray Titan, one of the fastest supercomputer in the world, contains just fewer than 300,000 cores, which are capable of operating more than 6,000,000 concurrent threads (Shimpi, 2012).

In other words, a supercomputer can be compared to tens of thousands of workstations performing together like a symphony orchestra to process billions and trillions of bits of data every second, sometimes for hundreds of users simultaneously. Some supercomputers are general or multipurpose machines that perform diverse tasks such as modelling and simulation or advanced business data analytics; others may be dedicated to specific tasks, such as operating cloud-based services, such as music streaming or managing telecommunications infrastructure (Carey, 2015).

1.3. Why is HPC important?

The Digital Economy is developing rapidly worldwide. It has been cited as the single most important driver of innovation, competitiveness and growth. Digital innovations such as supercomputing are an essential driver of innovation and spur the adoption of digital innovations across multiple industries and small and medium-sized enterprises, fostering economic growth and competitiveness. Applying the power of supercomputing combined with Artificial Intelligence (AI) and the use of Big Data provide unprecedented opportunities for transforming businesses, public services and societies.

Modelling and simulation have come to complement theory and experiment as a key component of the scientific method, and many of our scientific findings and technological advances rely on models simulated on high performance computers. The importance of this approach becomes clear with the following statement made by Andy Searle, Head of Computer Aided Engineering, Jaguar Land Rover. “It costs £500,000 to do each physical test of a car crash, and it’s not repeatable. It costs £12 to run a virtual simulation of a car crash, and it’s fully repeatable, so it can be used to optimise the design of a vehicle.” (BIS, 2013). This cost case is based on a study which is already around 6 years old and costs for the same quality of simulation have dropped since then. High-performance computing is mature technology mainstream for 40 + years, driven by academia and research, and it is based on well-understood mathematical, scientific and engineering principles, for solving advanced problems and performing research activities through computer modelling, simulation and analysis. However, HPC application areas are broader than simply modelling and simulation and cover large-scale data processing, real-time HPC and other applications. More recently, real-time HPC systems have appeared which can assist humans in instantaneous decision taking (for example, a medical simulation that supports a surgeon performing an operation).

The performance of HPC systems is increasing, as Figure 2 illustrates, growth in supercomputer operating speeds has increased exponentially over the past two decades (doubling in just 1.5 years). The digital revolution of the past two decades has led to unprecedented demands for high-performance data-processing systems. It is anticipated that the HPC will continue to be a vital technology, and demand for greater computing power will continue. One of the areas of HPC system application is in dealing with a wealth of data amassing to Terabytes/Petabytes of information. The Big Data development is relatively recent, in the last 10–15 years, but is evolving rapidly, and still developing its foundations and scientific underpinnings. Synergy of HPC and Big Data paradigms will enable new insights and ways the research and development of risk data analysis is undertaken (Holmes and Newall, 2016). Figure 1 demonstrates how HPC can accelerate the product development process cycle and reduce the "time to market" for a company providing competitive advantage. Some of the key application areas of HPC are described later in this module (and also in module 3) with case examples. These examples will help further to understand the importance of HPC in both scientific research and business application domains, to address some of the most advanced and complex and problems.

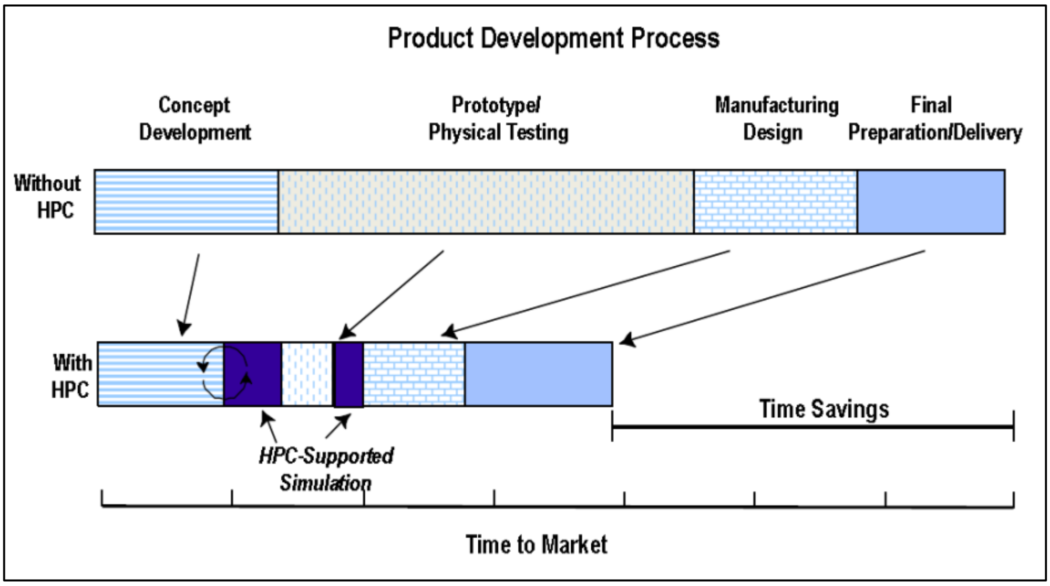


Figure 1: HPC advantage: Reduction in time to market [Source: IDC]

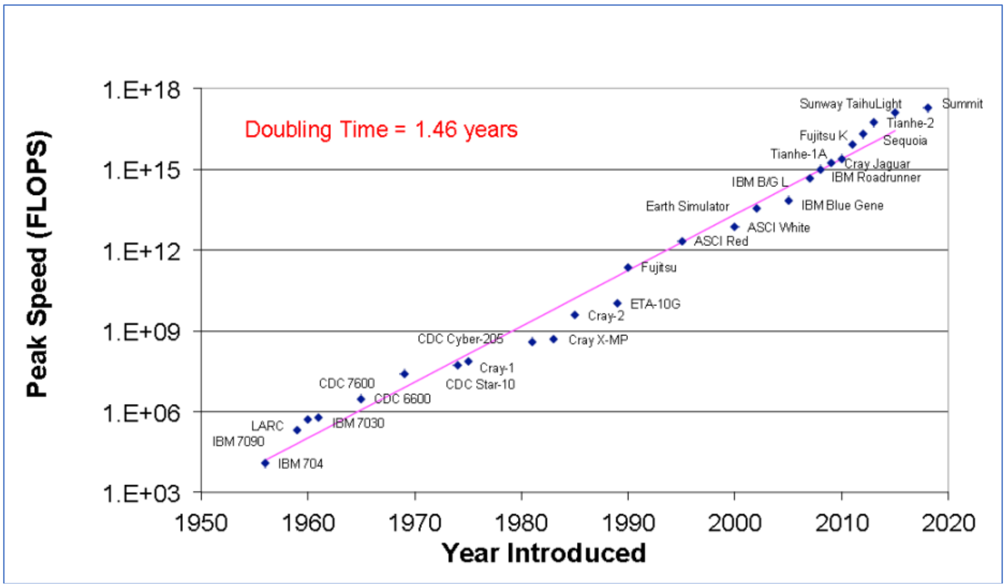


Figure 2: Growth of Supercomputing (1950-2020)

Source: https://www.top500.org/lists/2018/11/

1.4. History of High Performance Computing

The famous quote ``The Only Thing Constant Is Change" seems certainly to be true for the market of High-Performance Computing (HPC), If we look back on the last seven decades, it is clear how the marketplace for HPC has rapidly changed in terms of vendors, architectures, technologies and applications. Despite all these changes, however, the evolution of performance on a large scale seems to be a very steady and continuous process. Moore's Law is often cited to describe the progress in this context and as Figure 3 shows how well this law holds for a significant lifespan of modern computing between 1950 and 2000. On an average we see an increase in performance of 2 orders of magnitude every decade.

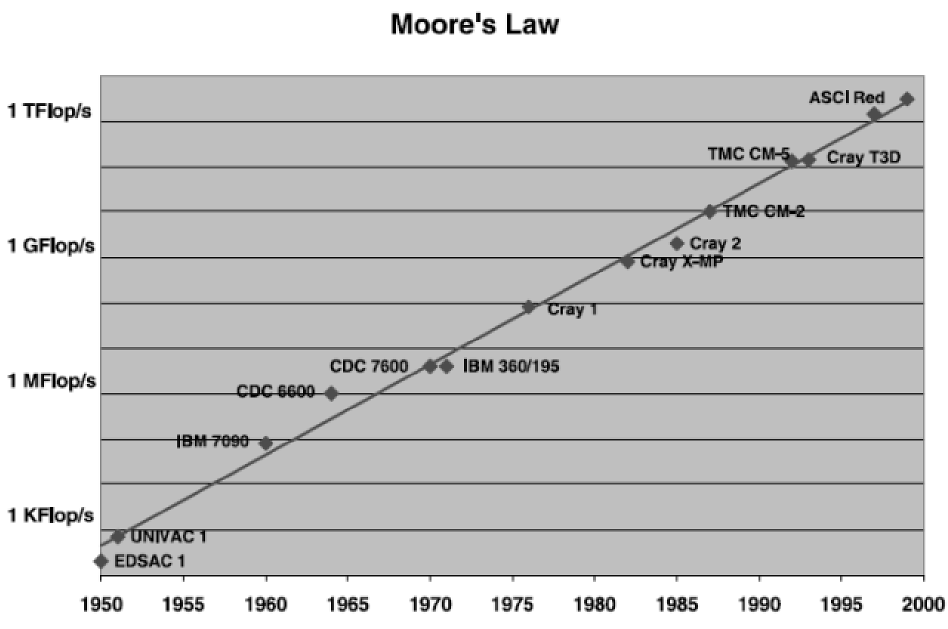


Figure 3: Performance of fastest computer systems for last 5 decades compared to Moore's Law [Source: Strohmaier et al., 1999]

1940s to 1960s: The first Computers

During this period the development of the computers was largely driven by aviation and military needs. Invented by a German scientist named Konrad Zuse, the Z3 was presented at the German Laboratory for Aviation in Berlin on May 12, 1941, as the world's first universal programmable, binary-based, electromechanical computer. Colossus was the first programmable, digital, electronic computer developed in the UK (Bletchley Park, England) in 1943 during World War II. It used vacuum valves and was designed for a single task of breaking Nazi codes like Enigma. In USA, ENIAC (Electronic Numerical Integrator and Computer) was the first stored-program electronic computer designed and built at the University of Pennsylvania from 1943 to 1946. During Cold War period, supercomputing was primarily used for design of Nuclear weapons, Aircrafts, Submarines etc., intelligence gathering and processing, and code breaking.

1970s to 1990s: Vector Era (Shared Memory Computers)

During this period speed in supercomputers was primarily achieved through two mechanisms:

- Vector Processors: these were designed using pipeline architecture to rapidly perform a single floating point operation on a large amount of data. Achieving high performance depended on data arriving in the processing unit in an uninterrupted stream.
- Shared Memory Multiprocessing: a small number (up to 8) of processors with access to the same memory space. Interprocess communication took place via the shared memory.

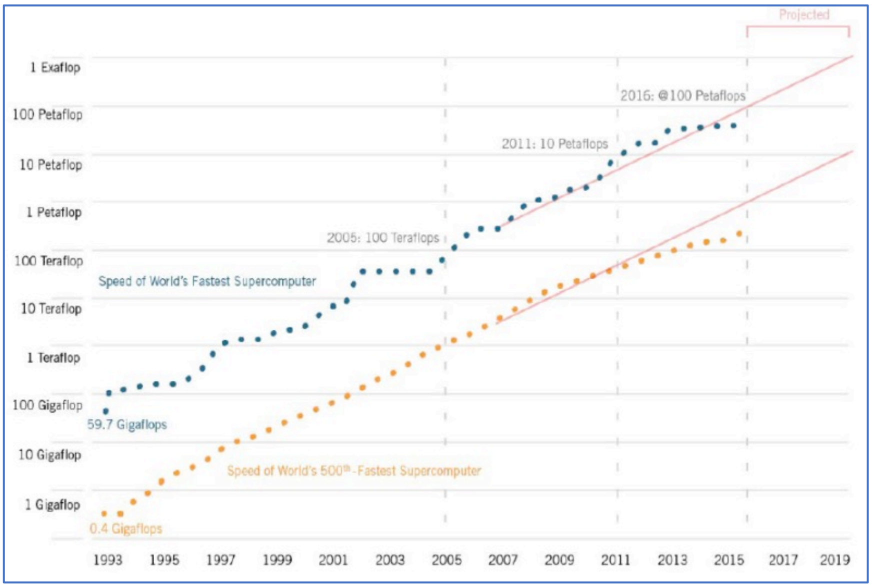


Figure 4: Speed of world's fastest and 500th-fastest Supercomputer (1993-2015)
Source: Ezell and Atkinson (2016)

1990s to 2000s: Cluster (Distributed Memory Computers)

The age of truly effective parallel computers had begun, but was already limited by access to shared memory. Memory contention was a major impediment to increasing speed; the vector processors required high-speed access to memory but multiple processors working simultaneously created contention for memory that reduced access speed. Vector processing worked well with 4 or 8 processors, but memory contention would prevent a 64 or 128 processor machine from working efficiently. The alternative to shared memory is distributed memory, where each processor has a dedicated memory space. The challenge became implementing effective processes communication - processes can't communicate with one another by writing data into shared memory; a message must be passed. During the 1990s there began to be a lot of interest in distributed memory computers.

2000 to Present: GPU and Hybrid Cluster era

During the 2000s the trend of increasing processor speeds was replaced with increasing the number of processor cores. This led to hybrid clusters with a large number of processors, each with a small number of core sharing RAM and some cache space. With the development of special purpose processor units such as GPU (Graphics Processing Unit) and other general purpose accelerator hardware, today's top supercomputers are hybrid clusters with a large number of standard processor nodes where each node has a multicore processor with some individual cache, some shared cache, and RAM shared by all cores.

Latest development in computing and internet technology has led to Utility computing - the new Cloud Computing paradigm. This on-demand computing or service computing is a concept of connecting to the external computing resources via Internet. It is defined as: a large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted virtualised, dynamically-scalable, managed computing power, storage, platforms and services are delivered on demand to external customers over the internet (Foster et al., 2008). Currently, a significant research focus lies on designing HPC solutions that integrates capabilities of High-performance computing with pay-per-use model of Cloud computing (Li, 2015).

Analysts measure the speed of computers by their ability to calculate floating-point operations per second (or "flops"). As Figure 4 illustrates, growth in supercomputer operating speeds has increased exponentially over the past two decades.

Name	Peak Speed	Developed By	Location
Summit	122.3 PFLOPS	IBM	Oak Ridge National Lab, USA
Sierra	94.6 PFLOPS	IBM	Lawrence Livermore National Lab, USA
Sunway TaihuLight	93.0 PFLOPS	NRCPC	National Supercomputing Centre in Wuxi, China
Tianhe-2	61.4 PFLOPS	NUDT	National Super Computer Centre in Guangzhou, China
Piz Daint	21.2 PFLOPS	Cray Inc.	Swiss National Supercomputing Centre (CSCS), Switzerland

Table 1: List of the top 5 fastest Supercomputers available currently
Source: https://www.top500.org/lists/2018/11/

1.5. Some Application Areas for HPC

In section 1.3, it has been mentioned how HPC is used for solving advanced problems and performing research activities through computer modelling, simulation and analysis. While HPC has traditionally been implemented for scientific research and computational science through use of supercomputer, recently the focus is gradually shifting from supercomputers to computing clusters.

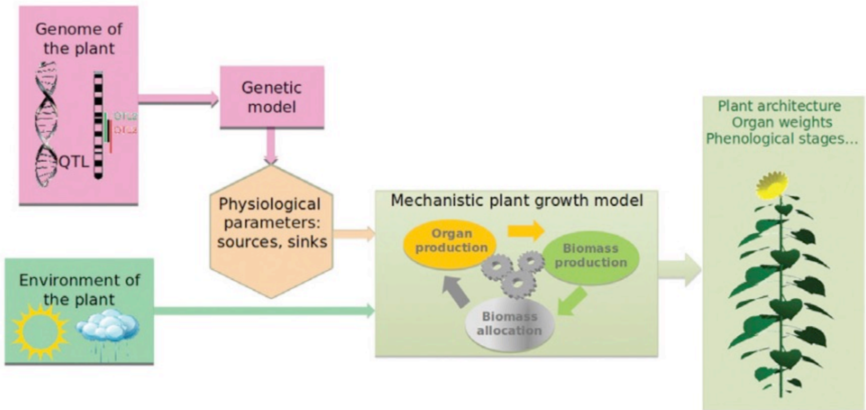
HPC is currently used to solve complex modelling problems in a spectrum of disciplines, addressing goals for both scientific research and business applications. Some of the key disciplines where HPC is been heavily used are:

- Artificial Intelligence and Machine Learning
- Climate Modelling and Weather predictions
- Automotive Engineering
- Agricultural Production
- Financial Markets
- Cryptographic Analysis
- Personalised / Precision Medicines

- Molecular Biology and Genetics
- Nuclear /Plasma /Quantum/Geo Physics
- Physical Oceanography
- Product Design

In this section, nine example case studies are illustrated to highlight the diverse application areas of HPC. However, scope of these case studies typically concern “mega” computational problem situations that demand dedicated and powerful supercomputers performing complex calculations at a grand scale. While the case studies make interesting reads and are useful in explaining the usage, power and effectiveness of HPC solutions, those might be a bit overwhelming and out of relevance from the perspective of SME businesses. Accordingly, some useful case studies that specifically focus on the computational problems related to the SMEs are discussed in section 3.2.3 under Module 3.

1.5.1. HPC for Precision Agriculture Production



Agriculture is the principal means of livelihood in many regions of the developing world, and the future of our world depends on a sustainable agriculture at planetary level. High Performance Computing is becoming critical in agricultural activity, plague control, pesticides design and pesticides effects. Climate data are used to understand the impacts on water and agriculture in many regions of the world, to help local authorities in the management of water and agricultural resources, and to assist vulnerable communities in the region through improved drought management and response.

The demand for agricultural products has increased globally and meeting this growing demand would have a negative effect on the environment. Increased agricultural production needs the use of 70% of the world's water resources and a rise in greenhouse gas emissions.

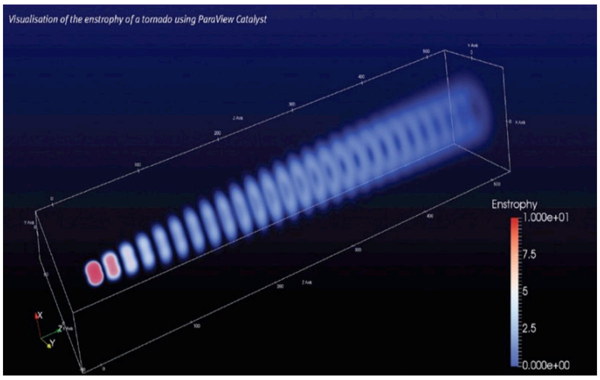
To be able to reduce the negative impact to the ecosystem, seed companies are on the lookout for new plant varieties that yield more produce. Companies normally find such new varieties through field trials. These field trials are a simple observation method but they cost a lot of money and are time consuming taking years to find the best ones.

Using High Performance Computing (HPC), it has become possible to provide the most efficient solution to this problem. HPC enables numerical simulations of plant growth that help seed companies to achieve superior varieties instead of doing field trials which are more expensive and harmful for the environment.

For example, if a farmer wants to know what the conditions are for a plant to grow best in (its genetic parameter), they would have to test its growth rate under various conditions to select the best parameter corresponding to the specific environment of the region. With the help of HPC, the estimation of these parameters is made more accurate and simpler by simulating plant growth. The simulation models take into account, the plant's interaction with the environment. It reduces the number of field trials by a large percent, for example, instead of 100, 10 field trials would be enough to estimate the best genetic parameter.

Cybele Tech, the French company has used High Performance Computing to enable farmers to produce more with less and know what exactly their plants need to get a better yield. [Source: European Commission (2018)]

1.5.2. HPC for Accurate Weather Forecast



In the current volatile climate conditions that characterize our planet, weather forecasts are very necessary to provide early warnings on weather extremes. Understanding and forewarning of extreme weather conditions such as impending storms, avalanches or tsunamis are of utmost importance for societies that are vulnerable to it to be more prepared and resilient.

Due to its increased computing capacity, High Performance Computing (HPC) can provide a more accurate representation of the climate models. Using HPC can enable and enhance prediction of extreme weather conditions, and also the reasons and impacts.

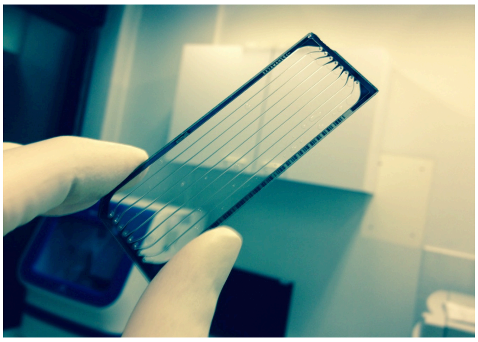
Weather modelling is used to improve our understanding of how processes work in land, oceans and the atmosphere. It also helps us to understand the evolution and impact of climate changes over the years. Weather modelling also helps predict weather between days and seasons as well as weather projection for the future.

Predicting Hurricanes' paths

More accurate models are needed to predict much in advance the path and the effects of the increasingly devastating hurricanes such as Irma and Harvey. The weather model from the European Centre for Medium-Range Weather Forecasts (UK) proved substantially more accurate than U.S.A. models in predicting the path of Hurricane Sandy that devastated America's East Coast in 2012. The MET Office, the UK's National Weather Service, relies on more than 10 million weather observations from sites around the world, a sophisticated atmospheric model and a £30 million IBM supercomputer to generate 3,000 tailored forecasts every day.

European Centre for Medium-Range Weather Forecasts predicted the path of the Hurricane Harvey fairly well also. The EU model predicted the movement of Harvey with high precision for the next 5 days on whether it would stall or move further, the direction of movement and other such details which are very helpful to find the right solutions. The European modelling centre has used supercomputing facilities to devise the best system to incorporate real-time meteorological observations into its model which implies that the accuracy of initial conditions is quite high. [Source: European Commission (2018)]

1.5.3. HPC for Genetic Sequencing, Diagnostics & Personalised Medicine



High Performance Computing is enabling major advances in new therapies: scientists heavily rely on HPC for understanding the nature of diseases, for discovering new drugs, and for moving to precision medicine, customising therapies to the specific needs of a patient.

➢ **Genetic Diagnostics**

HPC is used in simulating chemical reactions, to better understand neurological disorders and develop precision treatments. Molecular simulations help us to understand the structure and functionality of biomolecular systems. For example, simulations of chemical reactions involved in the metabolism of neurotransmitters help us understand how variation in the activity of monoamine oxidase (MAO) (enzymes that control the neurotransmitters levels) is connected to neurological disorders such as depression and autism. It also shows how oxidative stress (resulting from reactions in this metabolism) influences the development of neurodegenerative diseases such as Parkinson or Alzheimer. Getting this very detailed insight into the chemical processes, combined with genomic medicine data, requires extensive computing capability and can be used to accurately predict the varying activity of an enzyme due to genome variations in humans.

Identification of genetic causes of human disease is the cornerstone of precision medicine and enables preventive, predictive and personalised management for patients. In the last decade, human disease genetics has been transformed by novel genomic technologies that now permit accessible sequencing of whole human genomes. Slovenia has rapidly introduced these technologies in routine medical diagnostics, revolutionizing genetic testing in patients suffering from rare diseases. In order to assure rapid analysis of large amounts of data generated by these novel technologies, there was a need for collaboration with the national supercomputing facilities. Slovenian doctors used supercomputing infrastructure to massively accelerate genetic diagnostics, passing from more than one month to less than a few days, sometimes just a day. The use of supercomputers also allowed more comprehensive analysis of genetic material, which is crucial for diagnostics of patients with severe epilepsy, of critically ill newborns, in pre-natal diagnostics and for precision treatment of people with rare diseases. These novel approaches have also been a driver for novel discoveries in the field of human genetics, including identification of over a thousand of novel genes for human diseases in the recent years.

- **Personalised Medicine**

Innovative data processing technologies allow for new forms of medicine, called Personalised or Precision Medicine (PerMed), which is using information about a person's genes, proteins, and environment to prevent, diagnose, and treat diseases. Each type of cancer has its own genetic makeup, giving each tumour cell and tissue a unique character with specific tendencies and vulnerabilities. PerMed has a big impact on the early detection of rare diseases. Currently, the diagnosis takes a lot of time. Moreover, patients receive often maladjusted treatments because a precise analysis of a specific rare disease is extremely difficult. Most of the identified rare diseases are still not curable, primarily due to the lack of an exact understanding of the underlying disease mechanisms. The analysis of the whole genome is seen as key to the disease mechanisms but requires the computing of large amounts of data.

Generating results quickly and accurately from medical processes is today a primary concern. The work of medical doctors is depending more and more on computer systems that can support their decisions on the right treatment while providing direct access to the underlying reasoning. The use of decision support tools is increasingly complemented with data provided from mobile devices and sensors. They bring complementary information on the patient's profile and help disease management between visits. Real-time analysis of data was used in a pilot involving patients suffering from Parkinson disease. The pilot took place in Luxembourg. The data generated by the sensors from the smart shoes they were using showed changes in the movement pattern before the patients suffer from the typical freezing that leads to many falls and injuries. The extremely fast analysis of this data could provide warnings to patients in time to avoid accidents.

In addition, PerMed applications have a significant impact in the planning and evaluation of clinical trials. Several research and development initiatives have been launched to build data and software resources able to combine text analysing and genome mining tools. This approach is considered the most promising strategy in the research for new treatment methods. All these approaches of PerMed require the processing of large amounts of data and often rely on systems requiring a fast in- and output. Benchmarking tests at the Luxembourg Centre for Systems Biomedicine (LCSB) have shown that the use of high performance computing can reduce the time needed to process a full genome - as needed for rare diseases - from one day to 20 minutes. Using big data in the medical field has more advantages: it leads to more efficient data handling and new security technologies for a trusted management and transfer of clinical data.

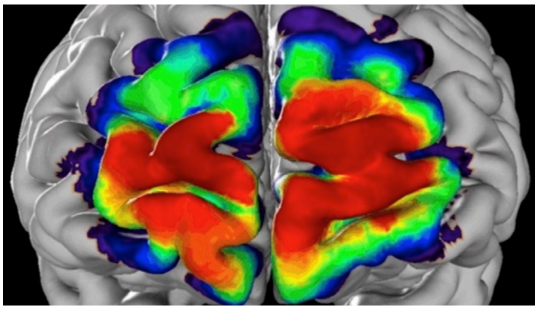
- **New Drug Development**

The time required to develop a new drug ranges between 10 and 17 years - if it ever makes it. There are also rising costs that make this increasingly unaffordable for both companies and patients. The testing of drug candidate molecules can be greatly accelerated by using High Performance Computing (HPC). HPC can also help to repositioning existing drugs for new diseases. This will benefit the treatment of the patients while strongly reducing the costs of the process. Dr. Zoe Cournia from the Biomedical Research Foundation of the Academy of Athens (BRFAA), Greece, extensively uses High Performance Computing (HPC) services from providers in Greece and the EU to investigate how mutations in proteins can lead to cancer. Analysing data from simulations, scientists can design candidate drugs explicitly for mutated proteins. These drugs should not affect normal proteins thus reducing side effects. Using HPC resources, Dr. Cournia's lab is able to identify cavities on proteins, where small molecules can bind, and then design candidate drugs for those specific cavities. Candidate drugs are then tested by the lab's experimental collaborators. This work is based on the concept of 'precision medicine' that targets very specific cancer types to treat patients more effectively. By taking advantage of computational simulations executed on Europe's most advanced HPC nodes, the time and cost of designing new candidate drugs has been significantly reduced.

With access to Europe's most advanced HPC systems through the most recent PRACE call, Dr. Cournia and her colleagues will now be able to map the full dynamics of PI3Ka, a commonly mutated protein in cancer, and use this information for designing new drug candidates against the mutation. The planned computational approach is of unprecedented scale and is expected to explore uncharted territory in the field, providing unparalleled information and insights on designing new drugs for this oncogenic protein.

Dr. Cournia declared: "I believe that our work is a good example of how access to a performant computational infrastructure helps develop products that have the potential to save millions of lives worldwide. Using HPC resources from Greece, we were able to rank #1 in an international drug design competition, Grand Challenge 2, demonstrating the accuracy of our approach."

1.5.4. HPC for Brain Research



- **Deciphering the functioning of the Human Brain**

The Human Brain Project (HBP) is the biggest ever brain research initiative at European level and it is building, using High Performance Computing (HPC) capabilities, a unique ICT infrastructure to help understand the organisation and functioning of the brain and its diseases.

The Human Brain Project (HBP) scientists use High Performance Computing (HPC) capabilities to conduct collaborative research mainly in simulation and big data analytics. For this they need the support of giant computing power because the human brain is so complex that a normal computer is not enough to simulate even a fraction of it.

HPC is essential for the Human Brain Project for producing multi-scale and high-resolution simulation and modelling of the brain. Among other examples, we can name brain molecular simulations, cellular simulations (for example of the hippocampus), simulations of cortical columns and simulations operating at system level.

The enormous computing power needed to store big amounts of data and to run the simulations requires the largest available supercomputers in Europe. These computers are located in the five Tier-0 supercomputer centres, (all members of HBP): GCS (Germany), BSC (Spain), CEA (France), CINECA (Italy) and CSCS (Switzerland).

HPC is key to conduct collaborative research in the Human Brain Project. Researchers perform complex data analytics that have become even more complex with the application of deep learning techniques. They need the largest memory capacities, massive data storage and fast data access in the Petabytes range and not only in their own computers but to exchange data on a European scale. Consequently, HBP researchers throughout Europe have increased their demand of use of the five European supercomputers, which limits the availability of use for each of the supercomputers. To overcome this limitation, HBP is developing an infrastructure with cloud-based tools that will interconnect the five supercomputers and provide access to their enormous data storage and simulation power to all users of the infrastructure, no matter their location.

- **Human Brain mapping**

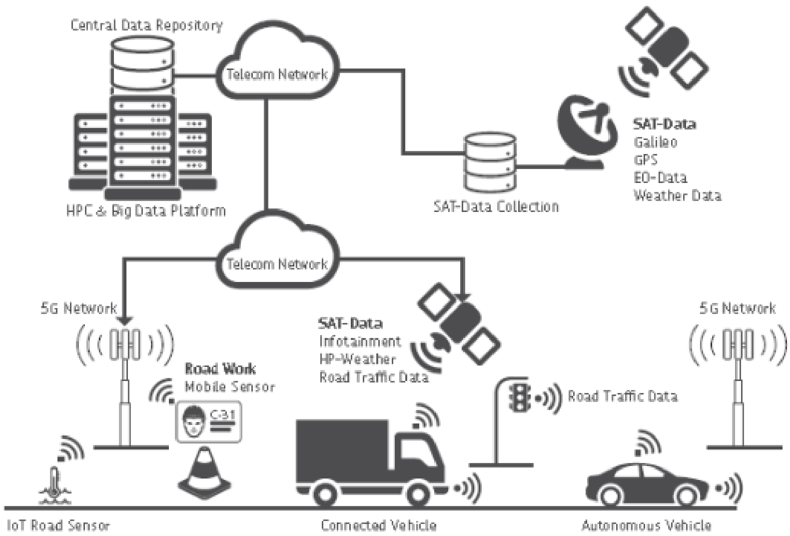
The human brain is one of the most complex systems that scientists face today. With a unique alignment of neuroscience research, large-scale simulation and multi-scale modelling, the Horizon 2020 FET Flagship "Human Brain Project" (HBP) is working towards an understanding of this complex organ.

A central aspect of this effort is a detailed 3D map of the brain, derived from many thousands of histological brain slices imaged at ultrahigh resolution with modern microscopes. Mapping brain areas is a very time consuming, semi-automatic process that involves analysing complex patterns of cell distributions in different independent subjects. Scientists at the research centre in Jülich/ Germany aim at creating a new generation of brain mapping tools that exploit the most advanced High Throughput Imaging devices, Machine Learning algorithms, and High Performance Computing (HPC) infrastructures available today.

They have trained a deep convolutional neural network (CNN) to classify texture in microscopic scans of brain tissue into different brain areas. The network learns precise texture features from existing annotations in microscopic images, and combines them with information from existing atlases.

The use of this modern HPC infrastructure enables the algorithm to process many large chunks of image data in order to capture both the cellular detail and spatial context. Without HPC, running the network would be almost infeasible. To this end, the neuroscientists and data analysts have worked closely with the JUELICH supercomputing centre (JSC) to run the application at scale on the GPU-accelerated clusters JURECA and JURON.
[Source: European Commission (2018)]

1.5.5. HPC for Future Smart Mobility- Driverless cars



Autonomous vehicles will generate and use a large variety of data to permanently analyse their geographical position, condition of the road, state of the vehicle, passenger comfort and safety. To manage all this data, we need High Performance Computing.

Driverless cars will be equipped with a large number of sensors, embedded cameras, in-car computers, high precision GPS and satellite receivers, short-range wireless network and 5G interfaces to connect to the Internet. These vehicles will permanently exchange data with management and supervising systems and will sync up with large data-bases that are constantly feeding them with real-time information about the local environment, traffic situation, emergency alerts and weather conditions. The transmitted information will be used by predictive driving functions to avoid road hazards and increase passenger safety.

When this type of vehicles will go main-stream, the amount of data generated will grow exponentially. According to Intel, driverless vehicles will send more than four terabytes of data (approximatively 1000 DVDs) in about an hour and a half of driving to the cloud. In the future, cars will create significantly more data than today the entire Internet community. Only next generation exascale High Performance Computing (HPC) and Big Data capabilities can deliver the required computing power to implement predictive decision support systems based on Artificial Intelligence to evaluate this enormous amount of data.

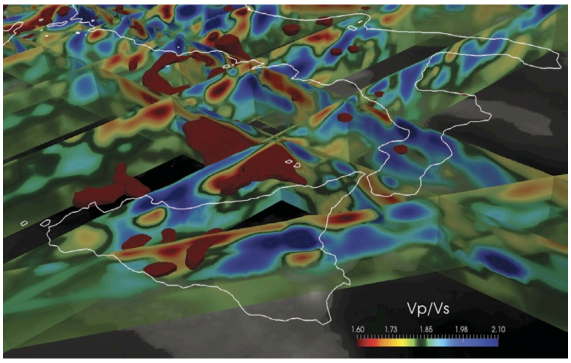
If the transportation-as-a-service business model (mobility solutions consumed as a service) will become a reality, it is essential that European companies stay in the race.

Luxembourg understands the importance of the big data and already adopted a digital transformation strategy for the entire country called the "The Third Industrial Revolution Luxembourg". This strategy aims to make the existing economic model more sustainable and interconnected for future generations by working with digital technologies, energy and transport as part of an intelligent network. The entire country as such will become a testing ground to try out the different innovative strategies on a real scale and on a national level. In the domain of intelligent transport and mobility systems, Luxembourg wants to test alternative transport systems like "electric and autonomous vehicles".

The "Smart Space - Mobility Application" is a new collaboration with the countries involved in the IPCEI: Italy France and Spain, to set up a consortium to develop together new concepts in these domains. It will combine "Connected Car" and "Big Data" technologies through the application of data cross-fertilization technics to gain new insights required for innovative Intelligent Transportation Solutions.

When using real-time High Performance Data Analysis, multiple data streams coming from Earth Observation and Galileo satellites, car and road sensors, weather models or traffic management, it will be possible to processed and converted them into meaningful information that can be delivered to people, vehicles and roadside infrastructures to improve on-road interactions, transport safety and travel comfort. [Source: European Commission (2018)]

1.5.6. HPC for Earthquake Simulation



High Performance Computing (HPC) is a powerful resource for improving our knowledge of geophysical processes and the structure of the interior of the Earth, for simulating and understanding earthquakes and protecting our society from their effects.

Analysing how the ground can move during seismic events, as well as understanding the crucial mechanisms that can lead to earthquakes, are difficult tasks that rely both on accurate images of the geological structure of the interior of the Earth at multiple scales and on the use of information buried in seismic "big data" records.

Using HPC resources, IMAGINE_IT project has developed a model of the lithosphere below the entirety of Italy based on highly accurate seismic wave imaging, providing a greater understanding of earthquakes in the region.

The international team lead by Dimitri Komatitsch used data gathered from 163 seismic events that have occurred across the Italian peninsula from 2005 to 2014, pushing the boundaries of 3D seismic velocity modelling. No model of this type has been created for an entire country at such resolution before. The model provided an accurate way of calculating where a seismic event has happened in the Earth and how big it was.

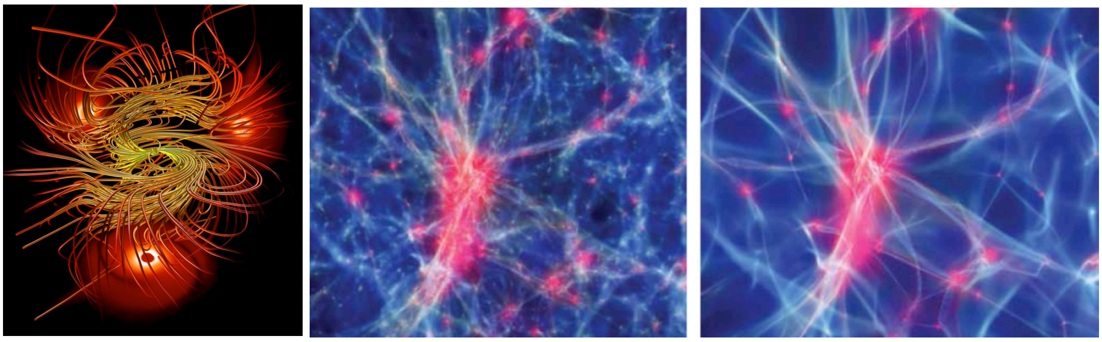
Italy, positioned at the meeting point of the Eurasian and African tectonic plates, is an area of high seismic activity, and as such has invested heavily in recording instruments that have generated huge amounts of high-resolution data. To properly analyze this data, the only feasible option was to apply to use one of the Tier-0 supercomputers offered by PRACE, with the team eventually being allocated 40 million core hours on GENCI supercomputers due to its architectural suitability to the computational techniques involved.

"What we cannot do is predict when an earthquake is going to happen; this is not possible with current techniques. But, thanks to HPC, by having more detailed information about the geological structures beneath the Earth and how they affect the movement of seismic waves, we can identify with greater precision the areas where seismic energy can be trapped and cause problems. All this information was and can used by anyone, from scientists and engineers to rescue people, workers and industries. In the future, this will help improve the monitoring of events and imaging of what is happening in near real time for emergency management situations." says Dimitri Komatitsch of the University of Marseille / CNRS.

All the conclusions and software generated by the project are available to the general public, providing benefits to the wider community inside and outside of academia.

The project has used big data heavily, and the techniques used could provide benefits in some surprising areas. Komatitsch believes that the methods they have used to handle big data in the project are easily transferrable to the lab's other areas of work, for example to the fields of medical imaging and non-destructive acoustic testing of materials or for measuring the structural integrity of a bridge or the walls of a nuclear power plant. The main difference is the scale and the source of the vibration being used. [Source: European Commission (2018)]

1.5.7. HPC for Astrophysics and Cosmology



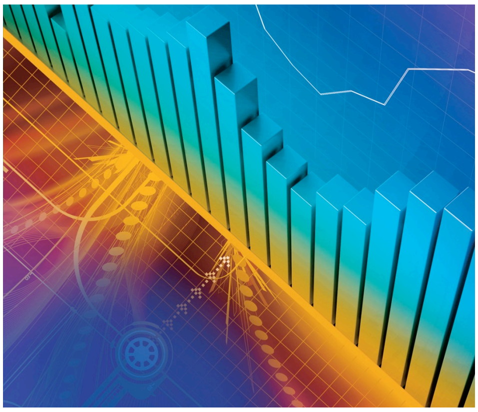
High Performance Computing (HPC) modelling and simulation techniques and data analytics approaches are key for understanding phenomena and finding innovative solutions in many scientific domains for high-impact science, like for example in astrophysics.

Plasmas are ubiquitous in the universe. Their intricate and complex dynamics have an important impact on solar flares (which are critical to the understanding of space weather). Plasmas are also important in harnessing controlled nuclear fusion and for the development of the next generation of ultra-compact accelerators and light sources. And they are also essential for the most intense lasers, or in extreme astrophysical objects such as pulsars or magnetars. To address some of the scientific and technological challenges associated with plasma kinetics in these scenarios, Instituto de Plasmas e Fusão Nuclear at Instituto Superior Técnico Lisbon have developed OSIRIS. OSIRIS is a fully relativistic, electromagnetic particle-in-cell (EM-PIC) framework. With it, they've performed some of the largest numerical plasma simulations in supercomputers worldwide.

OSIRIS has been successfully employed in the numerical modelling of several relevant kinetic plasma scenarios - ranging from astrophysical shocks and plasma shell collisions to high-intensity laser-plasma interactions.

OSIRIS is being used in more than 30 institutions by more than 150 researchers worldwide. Written in Fortran 2003 and C, its code is massively parallel, with demonstrated parallel scalability up to 1.6 million cores and sustained petascale performance in production runs. [Source: European Commission (2018)]

1.5.8. HPC for Financial Markets



The use of high-performance computing (HPC) in the financial services industry has experienced explosive growth in recent years. This is due in part to the emergence of highly complex products—particularly options, or derivatives—that require more computationally intensive financial models. Since 2008, new risk controls imposed on financial firms by government regulators in the wake of the global economic crisis have accelerated this trend.

We examine the principal applications of HPC on Wall Street and the algorithms, hardware, and programming languages these systems use. We also explore the challenges programmers face in moving to parallelization to take advantage of new computers' capabilities.

Table 1. HPC applications in the financial services industry.

Category	Low-latency jobs	Batch jobs
Algorithms	Simple math and complex logic	Partial differential equations (PDEs) and Monte Carlo simulations
Hardware	CPUs and FPGAs	GPUs and CPUs
Programming	VHDL, Verilog, C++, C#, and F#	C++

OPTIONS TRADING

Since Fischer Black, Myron Scholes, and Robert C. Merton published their groundbreaking work on derivative pricing in 1973,1,2 options trading has gone from a small business to a quadrillion-dollar industry. As the name suggests, options are financial contracts that give the buyer the right, but not the obligation, to execute some other transaction during a certain period of time or by a specified date. For example, a call option lets the buyer obtain an asset at a predetermined date and price. If at that date the asset price is higher than the agreed-upon option terms, it is advantageous for the buyer to "exercise" the option; otherwise, the buyer can purchase the asset in the open market at a cheaper price.

Options pricing and trading rest on a simple assumption: portfolios can be adjusted as the value of assets evolves. This allows option sellers to constantly neutralize the risks they bear by building portfolios in which losses on some assets (possibly options) are compensated for by profits on other, related assets. Using this assumption of continuous adjustments, traders can use the Black-Scholes-Merton formula and its variants to price many different derivatives. The calculation requires only transforming some probability measures and then performing numerical computations associated with the resulting stochastic processes.

Derivatives have become increasingly complex over the years. For example, the number of observations per asset as well as the number of assets that options reference have grown steadily, while options' contractual features have multiplied. New assets such as electricity, bandwidth, and volatility—a theoretical rather than a physical asset—have appeared, along with new dependencies: hybrid derivatives combine different types of assets into one product, and credit products have evolved from portfolios of simple asset-backed loans to "tranches" of such portfolios, requiring a much deeper understanding of correlated defaults on seemingly unrelated loans. Such complexity has led to ad hoc, or "over the counter" (OTC), transactions becoming the norm in some markets.

Collectively, these pressures have caused the computational requirements for financial instruments to balloon. As the risks associated with derivatives have increased along with their size and complexity, Wall Street has had to accelerate simulations. The fundamental assumption of continuous portfolio adjustments requires continuous analysis and valuation. While strict real-time systems are not necessary, response time remains a key metric of system capability. The bigger the risks, the stronger the need for a rapid response; but, unfortunately, the products are more numerous and complex as well. Larger financial risks require a faster response while making that response harder to achieve.

It soon became apparent that simply throwing more hardware at the problem was infeasible and that computer scientists would have to work alongside mathematicians to keep systems sufficiently responsive. This led to the emergence of computational finance and the quantitative developer profession, tasked with tackling classic computing issues such as scalability and software bloat (or the avoidance of Wirth's law3) in the context of financial services.

RISK MANAGEMENT

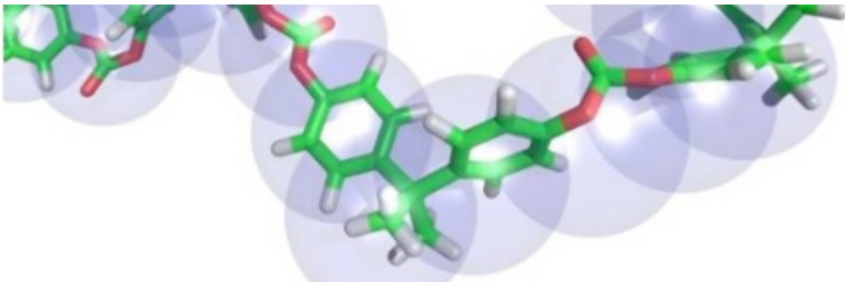
Financial crises are a recurring, even frequent, phenomenon.4 However, the 2008 meltdown impacted the global economy and hence triggered a rare, coordinated regulatory response. Financial institutions must now integrate "new" risks that have always existed but were frequently ignored—like insufficient liquidity—in the measures they use and report to regulators.

The increasing risk burden to financial institutions has motivated the improvement of stochastic market models. The approach traders and fund managers use depends on the objective—for example, algorithmic traders or "algos" focus on empirical modelling for short-term predictions, while asset managers focus on empirical modelling for long-term risk diversification. Derivative traders seek to predict the short-term evolution of option prices as functions of the underlying assets' prices, relying on risk-neutral hypothetical probabilities.

Risk should be understood as the latitude for mismatch between predicted and actual market activity. Even if they cannot predict values, financial institutions are interested in the relationships between, for example, asset prices and option prices, and they rely on such relationships to offset profits and losses among assets (and hence avoid gigantic bets against the market or clients). In this context, risk relates to the possibility of those relationships breaking down.

Wall Street firms are continually adding new risk measures. Since 2008, financial models must specifically consider liquidity issues, the simultaneous collapse of "house of cards" financial counterparties, multimarket dependencies, and other factors tied to the current crisis. Each new measure increases the need for computational power. Wider and deeper coverage is also required: not only are risk systems required to handle many more market scenarios than they did three years ago, they must compute prior scenarios over larger ranges of possible inputs, and they must do so as quickly as before. [Source : Spiers and Wallez (2010)]

1.5.9. HPC for Product Discovery (Manufacturing)



Unilever is taking advantage of the UK STFC (Science and Technology Facilities Council) Hartree Centre's expertise in high performance computing (HPC) to model how key ingredients of typical home and personal care products combine to structure everyday liquids.

<u>The challenge</u>

Inventing, making and selling home and personal care products is more complex and time consuming than often imagined. The level of complexity of Unilever's product portfolio has been compared to that of designing a Boeing 747. Just one example is the challenge of formulating a fabric conditioner. This product tends to be unstable, especially when it is shipped to very cold or very hot countries. Traditional stability tests, on the laboratory bench, tend to be boring and very time consuming, typically taking 8 to 12 weeks. However, the comparable test on a supercomputer takes only about 45 minutes.

<u>The solution</u>

Unilever now has a base at the Hartree Centre at STFC Daresbury Laboratory, home of the UK's most powerful supercomputer. The partnership with the Hartree Centre gives Unilever R&D a competitive edge by harnessing the power of supercomputers to accelerate the product discovery process. For example, a computer formulation tool will help scientists at the bench pre-screen a number of possible ingredients, so they can focus on fewer and better experiments when designing a new product. The high performance computing capabilities at the Hartree Centre are helpfully coupled with a specialist 3D visualisation suite, which Unilever product developers can use to explore the data and see correlations that are otherwise elusive to the eye.

The benefits

For a fast moving consumer goods company, speed is all that matters, especially when it needs to put on the market hundreds of new products every year. Today, "to out-compute is to out-compete". Speed is what gives a company like Unilever the competitive advantage.

[Source: STFC (2015)]

1.6. Reflections on this module learning

HPC solutions have evolved from giant supercomputers used only in laboratories by scientists / engineers, to distributed systems providing more accessibility and affordability in solving today's business computing challenges

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

Some business opportunities (or problems) faced by SMEs where HPC could help

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

1.7. Recommended Further Readings / Resources

1. Eadline, D. (2009). High Performance Computing for Dummies: Wiley Publishing, Inc.
2. Dowd, K., & Severance, C. (1998). High Performance Computing. CA, USA: O'Reilly Media.

Videos / Podcasts

1. https://www.youtube.com/watch?v=bkLVuNfiCVs (Introduction)
2. https://www.youtube.com/watch?v=A_i5kOlj_UU (HP Video)

SME/HPC

Erasmus+

Co-funded by
the European Union

MODULE 2: HPC Building Blocks

SME/HPC

Erasmus+

Co-funded by
the European Union

Aims of the Module

This module aims to describe/ explain:

- Parallel computing and its importance
- Different types of HPC solutions
- Clusters, its key components and benefits
- Grid Computing and Cloud Computing
- Terminology used around HPC systems / solutions

Prerequisites

Self-study of Module 1

Learning Outcomes

At the end of this module you will be

- familiar with the concepts of parallel computing and why it is important
- able to identify different forms of HPC solutions
- familiar with the concept of a cluster and its benefits
- able to identify the various components of a cluster
- familiar with the concepts of Grid Computing and Cloud Computing
- able to understand the various terminology in use for HPC

Module 2

- ✓ A brief introduction to Parallel Computing
- ✓ What are different types of HPC solutions?
- ✓ What is a cluster and what are its benefits?
- ✓ What are the key components of a HPC Cluster?
- ✓ HPC and Cloud Computing
- ✓ Demystifying terminology on HPC
- ✓ Reflections on this module learning
- ✓ Test Your Knowledge
- ✓ Recommended Further Readings / Resources

2. Building Blocks of High Performance Computing

2.1. Why Parallel Computing Matters

For over half a century, Information Technology (IT) industry has been driven by the dynamics of Moore's Law (Gordon Moore's revolutionary prediction made in 1965 that the number of transistors on a chip would double every 12 to 18 months, and thus, roughly, so would computer processing speeds). In the previous module, as we have noted, Moore's Law has been proven to be consistently valid, as over the past 50 years computer processing speeds have increased over one million-fold, unleashing a wave of innovation across industries ranging from aerospace to life sciences that have played a transformational role in driving the global economy and improving quality of life for citizens throughout the world.

However, it is now established that the ability to pack more transistors onto a single processor is beginning to reach its physical limits and experience diminishing returns. The dominant silicon-based complementary metal–oxide semiconductor (CMOS) architecture could start to hit physical limits that threaten to compromise Moore's Law unless a leap can be made to radically new semiconductor chip architectures or radically new systems architectures. Accordingly, foundational innovation in semiconductor electronics and systems architecture is needed from both the public and private sectors to ensure computing power continues to advance and improve our future digital economy. One way to deal with the increasing challenge of making individual chips more powerful is to effectively link those chips together in massively parallel computer systems, so more chips can work together in tandem to solve complex computational problems. In other words, using existing CMOS architectures, engineers are not likely going to be able to get individual microprocessors to go much faster, but they may be able to get more chips to work together simultaneously and also to position computing functions in different parts of the system itself. Such "parallel computing" and "distributed computing" approaches push processing capability out to other system components, even to the storage platforms, instead of having it all centrally focused on the CPU. In this new distributed architecture approach, instead of just driving information in and out of the CPU itself, engineers consider which tasks they are asking the machine to perform and where can they be more efficiently performed within the system. HPC systems are breaking new ground in distributed computing architectures, which may provide an avenue to sustaining the continued increases in computer processing speeds users have come to expect thanks to Moore's Law. However, while this may be true for large, massively scalable parallel and distributed computing systems, the challenge of developing faster chips for use in robotics or cell phones, for example, will continue.

2.2. A brief introduction to Parallel Computing

Generally speaking, high performance computing researches the parallel algorithm and develops the application related to parallel computing. By definition, parallel computing is a type of algorithm to execute multiple instructions simultaneously, operating on the principle that large problems can often be divided into smaller ones. Its purpose is to increase the computing speed, solve complex computing tasks. As simple as it may sound, splitting efficiently a large problem into smaller ones so that they can be truly solved in parallel, is almost an art. While "parallelly" may seem equivalent to "concurrently", both are quite different things. Solving a problem parallelly implies splitting the problem into smaller, completely independent tasks. There are two categories of parallelism, time parallelism and space parallelism respectively. Time parallelism refers to pipelining while space parallelism means concurrent computing by multiple processors. For example, pipelining may be compared to the assembly line technology in the factory. There are three steps of the assembly line: production, detection, and packaging. If not introduced the assembly line, the next product starts to be produced only after the first product completing all the steps, which costs more time and affects the efficiency. This is the time parallelism. Executing two or more operations at the same time greatly improves computational performance. As for space parallelism, for instance, Tom needs to clean the two rooms, which costs two hours. And if his friend helps him, it only takes 1 hour to clean both rooms. It is the same concept of space parallelism, splitting a large task into multiple same sub-tasks to speed up the problem solving. There is no unified computing model for parallel computing, but there are several valuable reference models. The most common one ought to be PRAM (parallel random access machine).

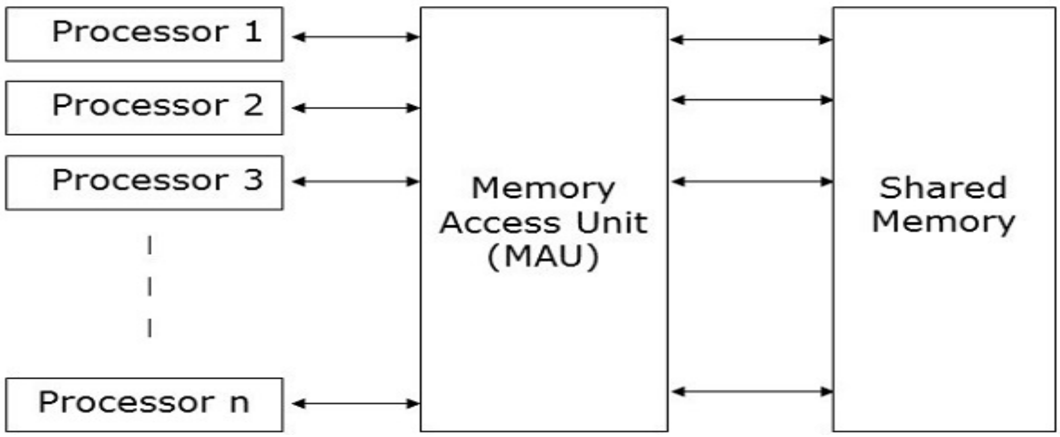


Figure 5: PRAM Reference Model for Parallel Computing

The PRAM model (Figure 5) is a type of model with shared storage in the SIMD (single instruction stream multiple data) parallel machine. It assumes that there are the infinite capacity of shared memory and shared storage unit that can be accessed at any time by the processors. However, this model is usually used for the theoretical analysis because of the infinite capacity storage does not exist.

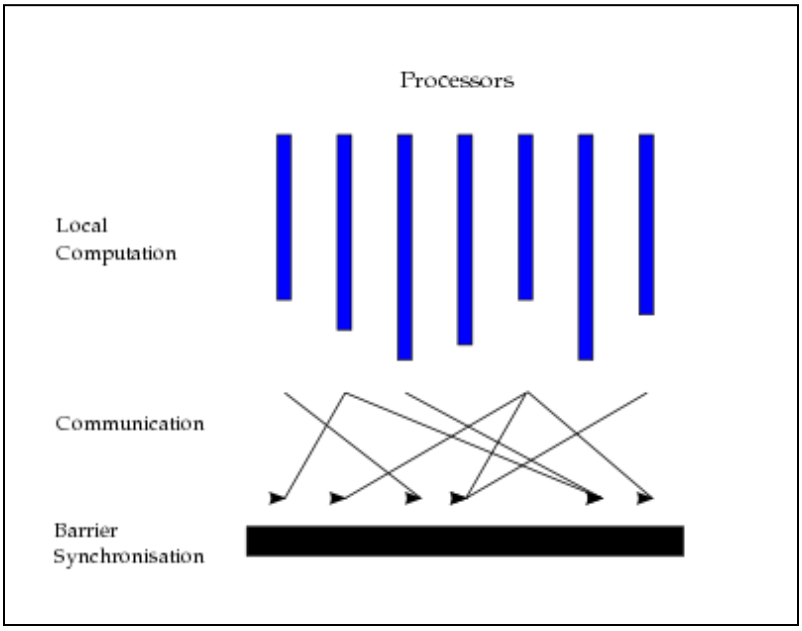


Figure 6: Bulk Synchronous Parallel (BSP) Computing Model

The other model is BSP (Bulk Synchronous Parallel) computing model (see Figure 6) that is developed by Viliant from Harward University and Bill McColl from University of Oxford. The BSP is carried out as a transition model, researching parallel computing between hardware devices and applications. A parallel computer based on BSP model, consists of a set of memory units linked by the communication network. There are three main elements: a set of distributed processors with local memory, global data communication network, and a mechanism that supports global synchronization between process units. It is a theoretical model for general architectures and scalable parallel performance software development.

2.3. Different Types of High Performance Computing

HPC solutions can be provided through four different forms. These types are described below.

2.3.1. Dedicated Supercomputers

In the past, the dedicated supercomputer was the only way to throw a large number of compute cycles at a problem. Supercomputers are still produced today and often use specialized non-commodity (often proprietary) components. Depending on your needs, the supercomputer may be the best solution although it doesn't offer the commodity price advantage.

2.3.2. Commodity HPC Clusters

Since the 1990s, there has been an increasing trend to move away from expensive, specialized, and often proprietary supercomputers to cost effective, general purpose systems for high performance computing. Such drive towards commoditization led to development of HPC cluster based architecture that is affordable and cost effective yet scalable and reliable in terms of performance.

2.3.2.1. What is a HPC Cluster

The idea of HPC Clusters stems from the idea of harnessing massive computing power by distributing the workload across many computers and combining the power of these multiple PC-based "commodity servers"- instead of investing in a giant, expensive supercomputer. In order to build a high-performance computing architecture, a large number of computer servers (often called 'nodes') are networked together into a cluster. Software programs and algorithms are run simultaneously on the servers in the cluster. The cluster is networked to the data storage to capture the output. Together, these components operate seamlessly to complete a diverse set of tasks. To bring a simple analogy, a cluster can be compared to a swarm of sharks (vis-à-vis a giant shark representing a mainframe or supercomputer) moving in tandem (see Figure 7).

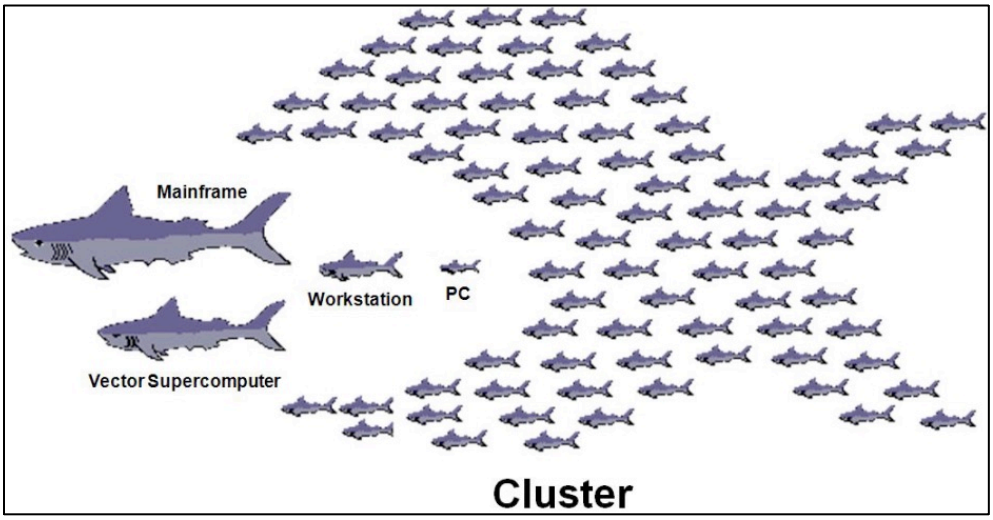


Figure 7: How Cluster compares to other forms of computing

2.3.2.2. What are the components of a HPC Cluster

A HPC cluster has four main components (see Figure 8 below):

1. Computer Server
2. Software
3. Networking
4. Storage

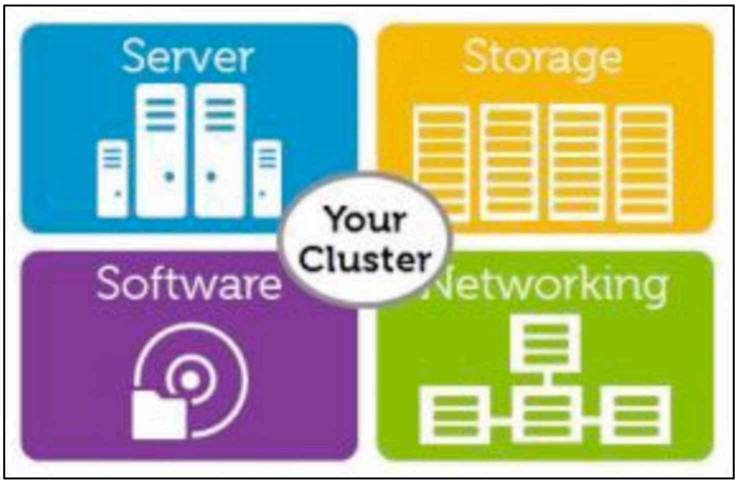


Figure 8: Components of a HPC Cluster [Source: Dell Inc.]

In a HPC cluster, each server has its own hardware (processor, memory etc.), software (operating system, and application software) and hardware-network interface to enable the servers to be connected to a common cluster interconnection communication network (see Figure 9 below.)

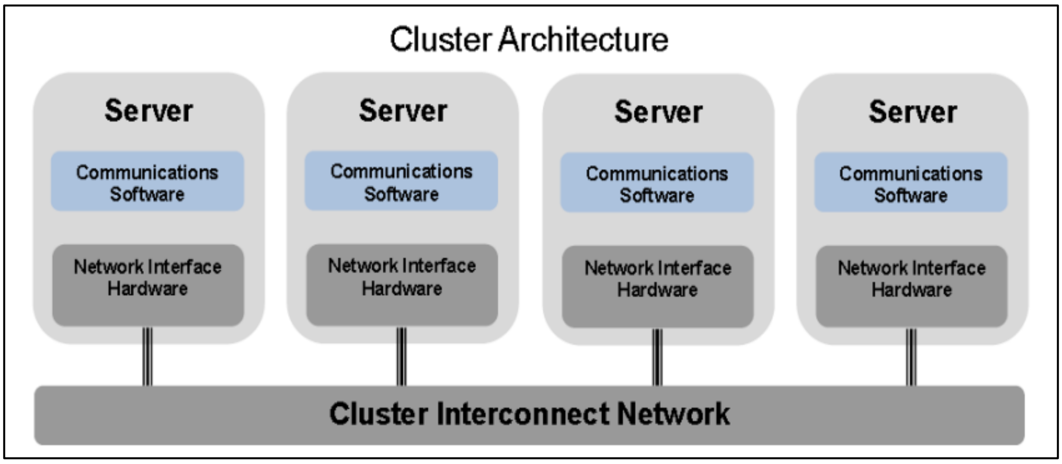


Figure 9: A typical Cluster Architecture

Figure 10 below provides an alternative view of a HPC Cluster depicting user terminal and data storage components as part of the cluster architecture.

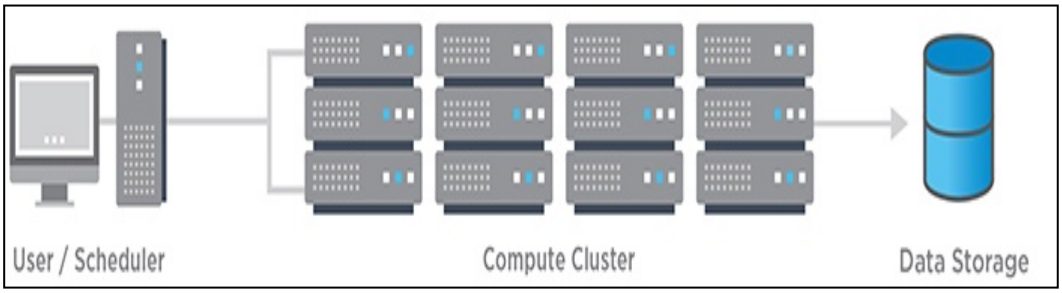


Figure 10: Another view of a HPC Cluster with data storage and user terminal

2.3.2.3. How does a Cluster work?

A computer cluster consists of a set of loosely connected or tightly connected computers that work together so that in many respects they can be viewed as a single system. On most modern computer clusters, computers are loosely connected. They have some way of talking with each other but work and behave independently. The advantages and disadvantages of using a computer cluster over using a single computer are strictly bound to the computer program that is being run, its dataset size and the operation types. HPC clusters will typically have a number of computers often called 'nodes' and, in general, most of these nodes would be configured identically. The idea is that the individual tasks that make up a parallel application should run equally well on whatever node they are dispatched on. However, some nodes in a cluster often have some physical and logical differences. Though from the outside the cluster may look like a single system, the internal workings to make this happen can be quite complex. To operate at maximum performance, each component must keep pace with the others. For example, the storage component must be able to feed and ingest data to and from the compute servers as quickly as it is processed. Likewise, the networking components must be able to support the high-speed transportation of data between computer servers and the data storage. If one component cannot keep up with the rest, the performance of the entire HPC cluster suffers.

Figure 11 provides a schematic view of a basic HPC cluster. Most of the nodes (servers) in the cluster are ordinarily compute nodes. With a specific end goal to give a general arrangement, a compute node can execute one or more tasks, taking into account the scheduling system. Since clusters are complex environments, and administration of each individual segment is essential, a master node or administration node is required that provides numerous capacities, including: observing the status of individual nodes, issuing administration orders to individual nodes to right issues or to give orders to perform administration capacities, for example, power on/off. One cannot underestimate the importance of cluster management. It is an imperative when trying to coordinate the activities of a large numbers of systems. Applications that keep running on a cluster, compute nodes must have quick, dependable, and concurrent access to a storage framework. Storage gadgets are specifically joined to the nodes or associated to a brought together the storage node that will be in charge of facilitating the storage demands.

To run a HPC cluster successfully, cluster software is a vital element of the architecture. There are several types of software tasks that need to be running, including administration, programming, debugging, job scheduling and provisioning of nodes. In order to take advantage of the combined power of a HPC cluster, the software programme needs to support running parallel tasks on multiple cores. Commercial and open source tools implementing Message Passing Interface (MPI) offer such parallel programming capabilities.

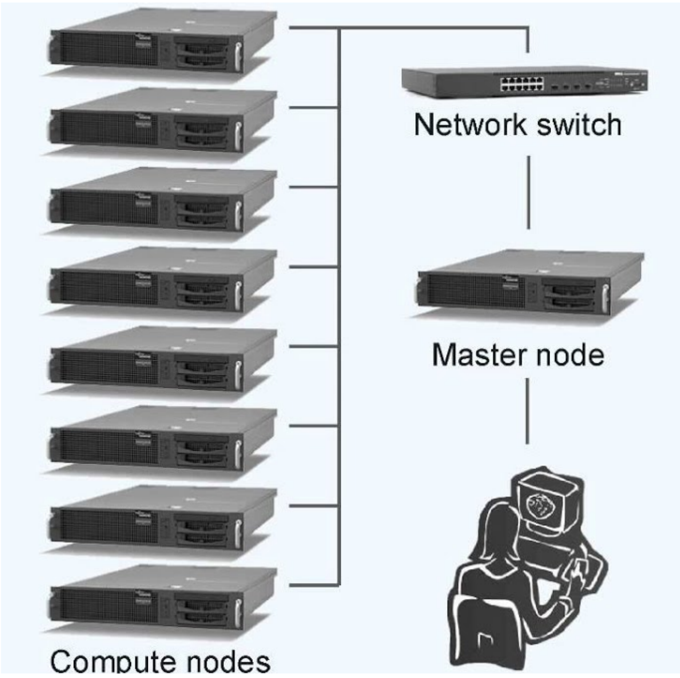


Figure 11: Schematic view of a basic HPC Cluster with Master Node

2.3.2.4. Benefits of a Cluster

A. Reduced Cost

The price of off-the-shelf consumer desktops has plummeted in recent years, and this drop in price has corresponded with a vast increase in their processing power and performance. The average desktop PC today is many times more powerful than the first mainframe computers.

B. Processing Power

The parallel processing power of a high-performance cluster can, in many cases, prove more cost effective than a mainframe with similar power. This reduced price-per-unit of power enables enterprises to get a greater ROI (Return on Investment).

C. Scalability

Perhaps the greatest advantage of computer clusters is the scalability they offer. While mainframe computers have a fixed processing capacity, compute clusters can be easily expanded as requirements change by adding additional nodes to the network.

2.3.3. Grid Computing

In the mid-1990s, the term Grid was coined to describe technologies that would allow consumers to obtain computing power on demand. Computing experts suggested that by standardizing the protocols used to request computing power, we could spur the creation of a Computing Grid, analogous in form and utility to the electric power grid. Researchers subsequently developed these ideas in many exciting ways, producing for example large-scale federated systems (TeraGrid, Open Science Grid, Earth System Grid among others) that provide not just computing power, but also data and software, on demand. Relevant standards were also defined by Standards organisations such as OASIS. Also, the term was leveraged by industry as a marketing term for clusters.

GRID Computing offers virtualization and distributed architecture techniques that can be applied to better utilize existing resources that have been installed in different locations for geographic or historic reasons and can be integrated as a global resource (Brochard, 2006). Grid computing is primarily used in academic projects where local HPC clusters are connected and shared on a national and international level. Some computational grids span the globe while others are located within a single organisation. Using such technologies each department, branch or subsidiary of the enterprise can access independently of its location all the resource of the enterprise increasing the productivity and the resilience of the HPC service. GRID computing technology enables new calculation to be performed because a community is putting together a set of resources than can then be used to perform some large HPC calculation. SETI@HOME (http://setiathome.ssl.berkeley.edu/) and WorldCommunityGrid.org (http://www.worldcommunitygrid.org) are two good examples of such communities on the Internet. This paradigm can also be applied to the Intranet of a company. Pharmaceutical company Novartis did boost its bioinformatics research using the cycles of its unused PC is an example of Intranet GRID computing. Web portals also can enable new users to access HPC resource through simple web interfaces.

Grid computing uses a processor architecture that combines computer resources from various domains to reach a computational objective. In grid computing, the computers on the network can work on a task together, thus functioning as a supercomputer. One advantage of grid computing is that it allows one to share computer resources across networks. This can both increase the computational power available to programs and reduce the number of machines needed by an organisation. As an example, Grid Computing instances often have hundreds or maybe thousands of processors. Desktop computers have an Operating System like Windows, Mac OS, or Linux.

2.3.4. HPC on Cloud Computing

Owning and maintaining traditional enterprise business systems have always been very complicated and expensive. The amount and variety of hardware, networking and software required to run them are daunting. You need a whole team of experts to install, configure, test, run, secure, and update them. When you multiply this effort across dozens or hundreds of applications, it's easy to see why the biggest companies with the best IT departments aren't getting the apps they need. Small and midsize businesses don't stand a chance. Cloud computing is a kind of outsourcing of computer programs. Using cloud computing, users are able to access software and applications from wherever they are; the computer programs are being hosted by an outside party and reside in the cloud. This means that users do not have to worry about things such as storage and power, they can simply enjoy the end result. With cloud computing, you eliminate those headaches that come with storing your own data, because you're not managing hardware and software — that becomes the responsibility of an experienced vendor like Amazon, Microsoft or Sales force. The shared infrastructure means it works like a utility with a pay-per-use model: you only pay for what you need, upgrades are automatic, and scaling up or down is easy.

Cloud Computing can be defined as a large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet (Foster et al., 2008). As it comes from the above definition, Cloud Computing is a specialized distributed computing paradigm; it differs from traditional ones in that:

- it is massively scalable,
- it can be encapsulated as an abstract entity that delivers different levels of services to customers outside the Cloud,
- it is driven by economies of scale, and
- services can be dynamically configured (via virtualization or other approaches) and delivered on demand

Currently, governments, research institutes, and industry leaders are rushing to adopt Cloud Computing to solve their ever increasing computing and storage problems arising in the digital economy. The five key benefits of a Cloud Computing system are:

1. **On-demand self-service:** On-demand self-service allows end users to access computing resources and experience computing facilities such as server services and network storage, as and when on demand, according to their needs, without additional authorization or the intervention of the system administrator.
2. **Rapid elasticity with consistent quality of service:** The service scope can be changed automatically to adapt to dynamic change of load. Overloading or redundancy of the server performance, leading to decrease service quality and waste resource, can be avoided in the cloud computing.
3. **Location-independent and shared resource pooling:** All computing resources are managed as a form of the shared resource pooling. These resources include networks, servers, storage, applications and services. The use of virtualization technology, shared resources technology and resource management, are all transparent to the end users.
4. **Ubiquitous network access:** The user can take great advantage of various terminal devices, such as mobile phones, laptops, computers, personal digital assistants, and so on, to access cloud computing via the Internet at any time anywhere.
5. **Measured service with flexible pricing model:** The cloud computing service here can be provisioned and billed to the users just like common utilities such as water, electricity or gas.

In terms of deployment, Cloud Computing can typically be deployed using P**ublic, Private, and Hybrid Clouds** (see Figure 12 below). The P**ublic Cloud** usually refers to the Cloud, owned and provided by a third party organisation, and is available to the general public or a large industry group. Typical examples of Public Cloud services are Amazon Web Services (AWS), Google Cloud and Microsoft Azure. A Public Cloud can be used via the Internet that may be free or low-cost. A core attribute of the public cloud is shared resource services. The **Private Cloud** is built for an individual client or company. The company or client can manage the deployment of the application based on the infrastructure. It is one of the most effective methods to control data traffic, data security and service quality. Private Clouds can be deployed within the firewall of the enterprise data centre, or in a safe place to host. A core attribute of a Private Cloud is the resource owned. The H**ybrid Cloud** is a blend of the Public and Private Cloud and shares many features of both these cloud models. In recent years, the hybrid computing is the main model of development. For data security reasons, companies are more willing to store data in a Private Cloud. At the same time, companies expect to gain access to the Public Cloud computing resources. The Private Cloud and Public Cloud maintain the distinctive entities, but are mixed and matched by standardized or proprietary technology in the Hybrid Cloud to achieve the best performance. Furthermore, the Hybrid Cloud breaks through the limitations of the hardware of the Private Clouds. Using the extensible ability of Public Cloud, the Hybrid Cloud can obtain higher computing power. When enterprises transfer the non-confidential data to Public Clouds, the needs and pressure of the internal Private Clouds can be reduced. Moreover, the Hybrid Cloud can effectively reduce the cost. The enterprise applications and data will be placed on the most appropriate platform to obtain more interests in the combination method.

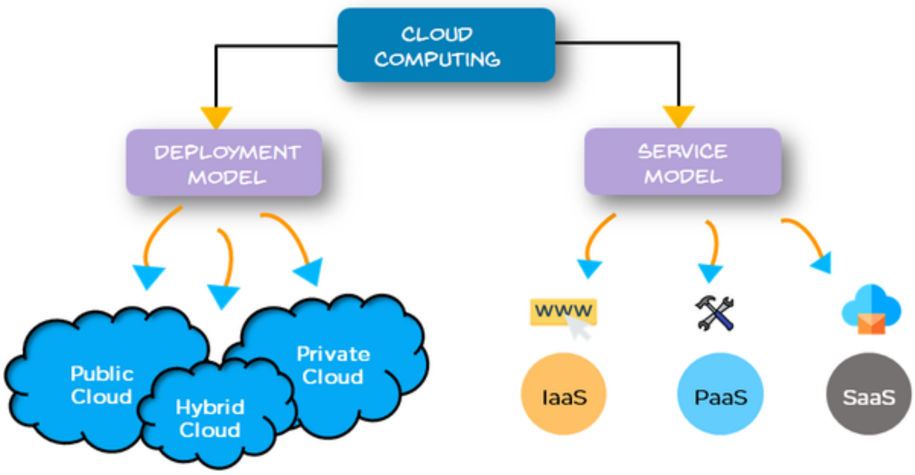


Figure 12: Cloud Computing Deployment and Service Models

Cloud Computing offers three service models that are categorised as

A. **Software-as-a-Service** (SaaS)

B. **Platform-as-a-Service** (PaaS)

C. **Infrastructure-as-a-Service** (IaaS)

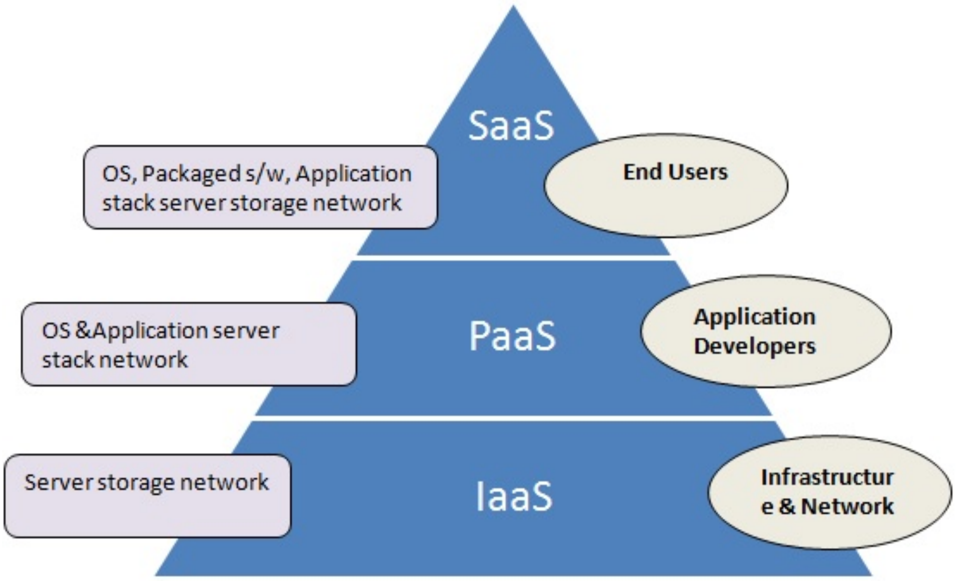


Figure 13: Cloud Service Models

Next, we will discuss these three Cloud service models (see Figure 13 above) in detail.

Software-as-a-Service (SaaS)

- SaaS is known as **'On-Demand Software'**
- It is a software distribution model. In this model, the applications are hosted by a cloud service provider and published to the customers over Internet
- In SaaS, associated data and software are hosted centrally on the cloud server
- User can access SaaS by using a thin client through a web browser
- CRM, Office Suite, Email, Games, etc. are the software applications which are provided as a service through Internet
- The companies like Google and Microsoft provide their applications as a service to the end users.

Advantages of SaaS

- SaaS is easy to buy because the pricing of SaaS is based on monthly or annual fee and it allows the organisations to access business functionalities at a small cost, which is less than licensed applications
- SaaS needed less hardware, because the software is hosted remotely, hence organisations do not need to invest in additional hardware
- Less maintenance cost is required for SaaS and do not require special software or hardware versions

Disadvantages of SaaS

- SaaS applications are totally dependent on Internet connection. They are not usable without Internet connection.
- It is difficult to switch amongst the SaaS vendors

Platform-as-a-Service (PaaS)

- PaaS is a programming platform for developers. This platform is generated for the programmers to create, test, run and manage the applications
- A developer can easily write the application and deploy it directly into PaaS layer
- PaaS gives the runtime environment for application development and deployment tools
- Google Apps Engine (GAE), Windows Azure, SalesForce.com are the examples of PaaS

Advantages of PaaS

- PaaS is easier to develop. Developer can concentrate on the development and innovation without worrying about the infrastructure
- In PaaS, developer only requires a PC and an Internet connection to start building applications

Disadvantages of PaaS

- One developer can write the applications as per the platform provided by PaaS vendor hence the moving the application to another PaaS vendor is a problem

Infrastructure-as-a-Service (IaaS)

- IaaS is a way to deliver a cloud computing infrastructure like server, storage, network and operating system.
- The customers can access these resources over cloud computing platform i.e. Internet as an on-demand service
- In IaaS, you buy complete resources rather than purchasing server, software, data centre space or network equipment
- IaaS was earlier called as Hardware as a Service (HaaS). It is a Cloud computing platform based model
- HaaS differs from IaaS in the way that users have the bare hardware on which they can deploy their own infrastructure using most appropriate software.

Advantages of IaaS

- In IaaS, user can dynamically choose a CPU, memory storage configuration according to need
- Users can easily access the vast computing power available on IaaS Cloud platform

Disadvantages of IaaS

- IaaS cloud computing platform model is dependent on availability of Internet and virtualization services

As digital technologies grow ever more powerful and available, apps and cloud-based platforms are becoming almost universally widespread. Businesses are taking advantage of new SaaS and PaaS capabilities to further outsource tasks that would have otherwise relied on local solutions. This is all made possible through advances in cloud computing. Delivering HPC solutions over Cloud essentially is often termed as "High-performance computing as a service (HPCaaS)", which essentially means the provision of high-level processing capacity to customers through the Cloud. Recently, this mode of HPC solution has gained significant traction. HPCaaS provides dynamic and scalable resources (and possibly virtualization) to the end-user allowing a user to request remote access to HPC service on-demand. Although Clouds can be cost effective and allow HPC solutions to be purchased as an expense and not a capital asset, it also places some layers between the user and hardware that may reduce computing performance. Additionally due to the nature of the Cloud services, the users lacking knowledge or expertise of HPC might find it challenging in contrast to "on-premise" services rendered by national or HEI HPC centres.

2.4. Demystifying some terminology around HPC

2.4.1. Exascale Computing

Exascale computing refers to computing systems capable of at least one exaFLOPS, or a billion billion (i.e. a quintillion) calculations per second. Such capacity represents a thousand fold increase over the first petascale computer that came into operation in 2008. Exascale computing would be considered to be a significant achievement in computer engineering, for it is estimated to be the order of processing power of the human brain at neural level. It is, for instance, the target power of the Human Brain Project.

2.4.2. Hyperscale Computing

Hyperscale computing is a distributed infrastructure that can quickly accommodate an increased demand for internet-facing and back-end computing resources without requiring additional physical space, cooling or electrical power. Hyperscale computing is characterized by standardization, automation, redundancy, high performance computing (HPC) and high availability (HA). The term is often associated with cloud computing and the very large data centres owned by Facebook, Google, Amazon and Netflix.

So technically, HPC can be considered to be a vital element of hyperscale computing paradigm. Currently, there is a lot of interest in hyperscale computing because the open source software and architectural changes created for hyperscale data centres are expected to trickle down to smaller data centres, helping them to use physical space more efficiently, consume less power and respond more quickly to user's needs. Hyperscale innovations currently being adopted by smaller organisations include software-defined networking (SDN), converged infrastructure and micro-segmentation.

2.5. Reflections on this module learning

How would you compare HPC solutions delivered by "on-premise" (HPC centre) with the same delivered "on-Cloud"?

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

2.6. Test Your Knowledge

1. Point out the wrong statement from the options below
   a. All applications benefit from deployment in the Cloud
   b. With Cloud computing, you can start very small and become big very fast
   c. Cloud computing is revolutionary, even if the technology it is built on is evolutionary
   d. Cloud computing employs the Internet as a basis for new service models of high performance computing

2. Fill in the blank with the most appropriate choice

   _________ as a Service is a Cloud Computing service model that offers a development environment upon which applications could be built.

   a. Infrastructure
   b. Service
   c. Platform
   d. All of the mentioned

3. Fill in the blank with the most appropriate choice

   _________ as a Service is a Cloud Computing service model in which hardware is virtualized in the Cloud.

   a. Infrastructure
   b. Service
   c. Platform
   d. All of the mentioned

4. Point out the wrong statement from the options below
   a. HPC Clusters work on a “divide and conquer” approach in solving large computing problems

b. A HPC Cluster can employ thousands of distributed servers working in tandem

e. All nodes (servers) in a HPC cluster need to have similar configuration and capacity

f. A HPC Cluster solution provides better scalability in compared to a supercomputer based solution

5. In the Computing world, FLOPS refers to

   a. A type of processor configuration

   b. Size of a supercomputer

   c. A measure of computer performance

6. Performance of a HPC Cluster typically depends on

   a. Number of nodes (servers) in the cluster

   b. Number of cores and processor speeds in the individual nodes

   c. Memory of the individual nodes

   d. All the above

7. Parallel File Systems were developed in order to

   a. Distribute the storage load across multiple, separate storage devices

   b. Allow parallel file access by multiple tasks across a Cluster network

   c. Support HPC solutions that require access to large files and massive volume of data

   d. All the above

2.7. Recommended Further Readings / Additional Resources

1. Foster, I., Zhao, Y., Raicu, I., & Lu, S. (2008). Cloud computing and grid computing 360degree compared. Published in: IEEE conference proceedings of Grid Computing Environments Workshop 2008: 1–10. IEEE, New York.
2. Sridhar, T. (2015). Cloud Computing - A Primer [online]. Available from: https://www.cisco.com/c/en/us/about/press/internet-protocol-journal/back-issues/table-contents-46/124-cloud2.html
3. https://whatis.techtarget.com/definition/hyperscale-computing

Videos / Podcasts

https://www.youtube.com/watch?v=FU3tUKkKgzQ (Amazon AWS and Cloud)
https://www.youtube.com/watch?v=11Z_RRFe6Rg (Part 1)
https://www.youtube.com/watch?v=zmYzF-x91JE (Part 2)

SME/HPC

Erasmus+

Co-funded by
the European Union

MODULE 3: HPC Solutions for SMEs

Erasmus+

Co-funded by the European Union

Aims of this Module:

This module aims to describe / explain:

- Global market landscape for HPC
- Who the leading providers of HPC solutions are
- How HPC is relevant to SME businesses (illustrated with case studies)
- Key challenges for HPC adoption by SMEs
- Some examples of HPC initiatives for SME sector at regional and international level
- HEI HPC Infrastructure for SMEs – Exploring potential business models

Prerequisites:

Self-study of Module 1 and 2

Learning Outcomes:

At the end of this module you will be

- familiar with the global HPC market landscape, including market drivers, trends, segments and key players
- able to define an SME and get a glance of European SME sector
- able to articulate the value proposition of HPC for SME businesses
- able to gather some examples of HPC applications applied in SME sector, at both regional and international levels
- able to identify the key challenges for adoption of HPC by SMEs
- aware of potential business model options for HPC offering from HEIs

Module 3

- ✓ HPC Global Market Landscape
- ✓ Value Proposition of HPC for SME Businesses
- ✓ HPC Adoption Challenges for SMEs
- ✓ Initiatives on HPC Adoption by SMEs: Regional and International Perspectives
- ✓ HPC Centres in HEIs- A Business Model Perspective
- ✓ Reflections on this module learning
- ✓ Test Your Knowledge
- ✓ Recommended Further Readings / Resources

3. HPC for Small and Medium Enterprises

3.1. HPC Global Market Landscape

Some key market statistics and trends are shown below (Intersect360 Research, 2018; Grand View research, 2018)

- Global HPC market size was valued at **USD 35.4 billion** in 2017
- HPC market is expected to expand at a **CAGR of 7.2%** from 2018 to 2025 with a projected market size of **USD 59.6 billion** by 2025
- Growth is driven significantly by **commercial HPC solutions**
- HPC market includes **Servers, Storage, Software, Networks, Services and Cloud**
- Revenue-wise **Commercial HPC solutions** constitute 56% of the market, followed by solutions for **Government** (26%) and **Academic** (18%) projects
- While USA is the **leading market for HPC** (30% of global market share), fastest growth is happening in Asian countries (China, India, Indonesia etc.)

- By deployment of HPC solutions, **“on-premise”** segment currently holds the top position. The **“on-cloud”** deployed solutions are gaining higher adoption and this segment is currently exhibiting fastest growth.

3.1.1. Key drivers for HPC market growth

In this section, we will try to understand the factors that are currently driving demand for HPC solutions and consequently, fuelling substantial growth in the HPC market.

- **Scientific research** in the areas of space exploration, energy, weather forecasting etc. that demands high performance computing. For example, the National Nuclear Security Administration (NNSA) and the Office of Science, both of which fall under the U.S. Department of Energy, have embarked upon a collaborative effort called the Exascale Computing Project.
- HPC is finding **widespread application in diverse scientific and engineering areas** such as military defence, national security, training and simulation, navigation technologies, high-resolution image processing, cryptographic analysis, satellite mapping, autonomous vehicle design, among others.
- **Rising popularity of cloud computing coupled with the digitization initiatives** adopted by several governments would drive the market in coming years. For instance, the Government of United Arab Emirates (UAE) has launched digital transformation projects, such as Smart Abu Dhabi and Smart Dubai. Such projects would encourage adoption of cloud computing and trigger the demand for HPC systems.
- The ability of HPC systems to **process large volumes of data at higher speeds** is prompting government agencies, defence agencies, academic institutions, energy companies, and utilities to adopt HPC systems. Enterprises are generating massive amounts of data, which also creates the need for computational solutions and capabilities to manage or handle them. Delivering **“Big data” solutions with HPC systems** is currently one of the top focus areas in today’s computing world
- Scientific research as well as business applications involving **Artificial Intelligence (AI) and Machine Learning** technologies necessitate high performance computing solutions

- The emergence of **Exascale Computing** (Fiore et al., 2018) is expected to open new opportunities in the HPC market in the future. Exascale computing refers to computing systems capable of at least one exaFLOPS, or a billion billion (i.e. a quintillion) calculations per second. It is considered to be a significant achievement in computer engineering, for it is estimated to be the order of processing power of the human brain at neural level and accordingly, is being used in the Human Brain Project.
- Systems not running true HPC applications, but requiring HPC technologies to run services at supercomputing scale (e.g., Google, Amazon, Baidu, etc.) – referred as **"Hyperscale systems"**

3.1.2. HPC Market Segments

At a broad level, HPC market can be categorized into following two segments:

High-Performance Technical Computing (HPTC):

Applications in Science and Engineering that typically includes following areas:

- Bio/Life Sciences/ Pharmaceuticals
- Manufacturing and Industrial Design
- Geosciences and Geoengineering
- Meteorology
- Energy
- Defence and Aerospace
- National Security
- Government, Academic and Not-for-profit

High Performance Business Computing (HPBC)

Non-scientific Business applications that typically includes following areas:

- Financial Services
- Media and Entertainment
- Retail
- Transportation and Logistics

- Healthcare
- Business Analytics

Table 2 below illustrates these two market segments - HPTC and HPBC, with some example application areas within each segment.

High Performance Technology Computing (HPTC)	High Performance Business Computing (HPBC)
University/Academic - Basic and Applied research	Economics/Financial: Risk analysis, Fraud detection, automated trading, portfolio management, pricing etc.
Government Labs – Basic and applied research	Digital Content Creation and Distribution: Computer-aided graphics in film, media and entertainment (gaming)
Geosciences and Geoengineering – Oil & Gas exploration and Reservoir modelling	Retail: Customer analytics, product design, supply chain management
Chemical Engineering: Process and Molecular design	Transportation
Defense and Energy – Basic and Applied research in areas such as Nuclear, Aerospace, Alternative energy etc.	Training and Simulation
Weather forecasting: Near term and Climate / Earth modelling	Information Technology (IT) and Telecommunication
Computer-aided Engineering: Automotive design and testing, transportation, structural, mechanical design	Healthcare- Data analytics,
Electronic Design and Automation: Electronic component design and verification	
Mechanical Design & Drafting: 2D/3D design and verification, mechanical modelling	
Bio-sciences and the Human Genome : Drug discovery, disease detection / prevention	

Table 2: HPTC and HPBC Segments

Figure 14, 15 and 16 are shown with an aim to provide a glance of the HPC market landscape, in terms of market revenues by verticals (Industry / Government / Academic), leading vendors for HPC servers (on % market share) and companies (on % of global revenue earnings from HPC solutions). However, these data could be a bit out-of-date and also, might not be accurate due to the mechanism the market research firms categorise the HPC solutions and calculate market performance.

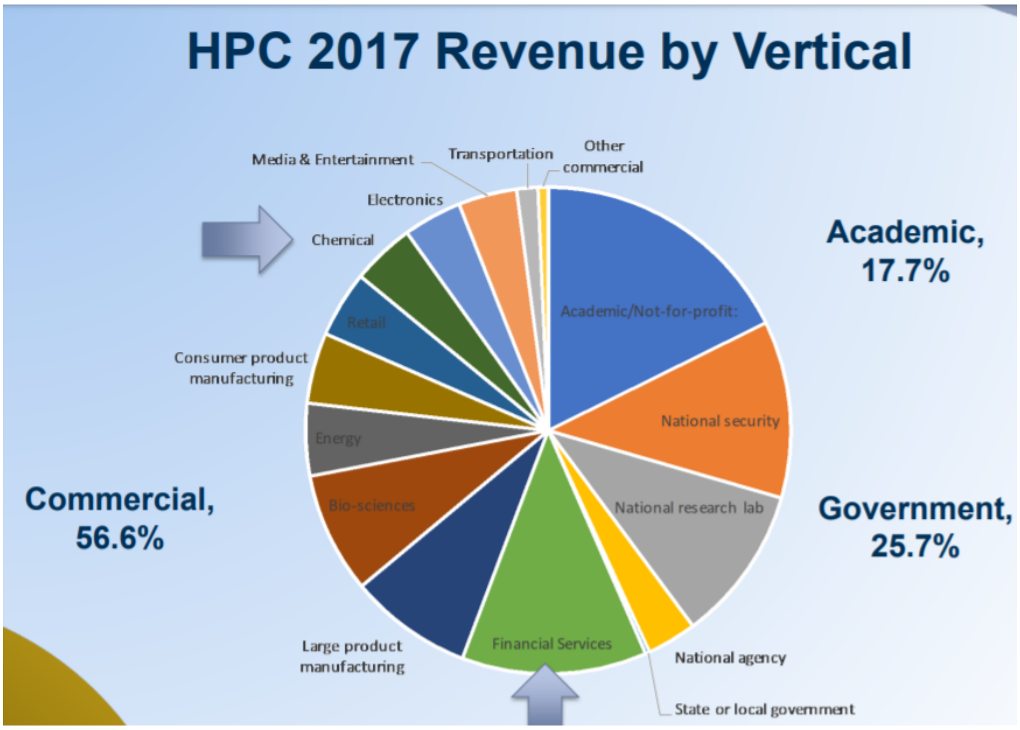


Figure 14: Global HPC Market Revenue by Verticals 2017
[Source: Intersect360 Research, 2018]

3.1.3. Key Players in HPC Market

Globally, the top HPC manufactures / solution providers are:

- Hewlett Packard (HP) Enterprise (US)
- Dell (US)
- Lenovo (China)
- IBM (US)
- Cray (US)
- Fujitsu (Japan)
- NEC (Japan)
- Atos SE (Europe, France)
- Silicon Graphics International (SGI), US

- Advanced Micro Devices (AMD), US
- Intel Corporation, US
- Cisco

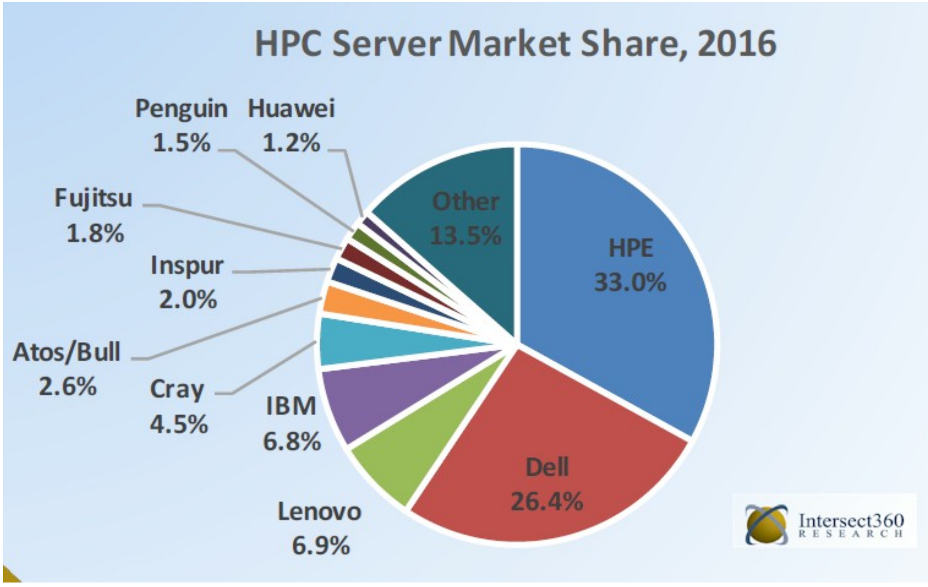


Figure 15: HPC Server Market Share 2016 [Source: Intersect360, 2017]

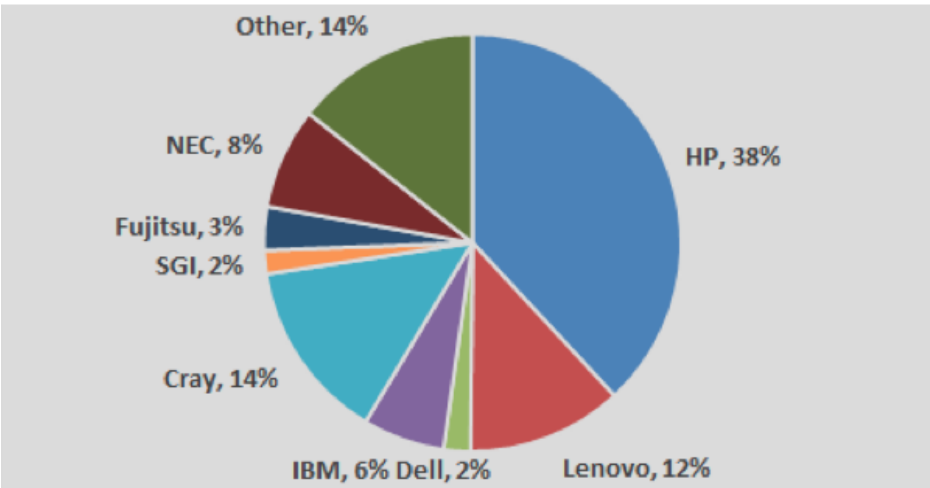


Figure 16: Company share of Global HPC Revenues Q1 2015
[Source: IDC]

3.2. Value Proposition of HPC for SME Businesses

3.2.1. What is meant by an SME?

An SME is a small or medium-sized enterprise. As per European Union (EU) definition, an SME is a business with **fewer than 250 employees**, and a turnover of **less than €50 million**. EU also makes definitions of the three categories of business within the SME umbrella.

According to the EU recommendations:

- A **medium-sized business** has fewer than 250 employees and either a turnover of up to €50 million or a balance sheet total of up to €43 million
- A **small business** has fewer than 50 employees and either a turnover of up to €10 million or a balance sheet total of up to €10 million
- A **micro-business** has fewer than ten employees and either a turnover of up to €2 million or a balance sheet total of up to €2 million

In recent times, EU ran a consultation with a view to potentially changing its SME definition. The responses were collected earlier in 2018, and the latest progress can be seen on the EU's website. SMEs make up for a significant proportion of all the businesses operating in the EU countries, and therefore SME sector is enormously important to the EU economy. Millions of people work in SMEs, and they are generally seen as the key engine of innovation, growth and sustainability.

In 2015, more than 22.3 million SMEs in the European Union constituted 99.8% of all non-financial enterprises. Often referred to as the backbone of the European economy, they employed around 90 million people and generated around 58% of total added value (€ 3.9 trillion). Over the past five years SMEs have created around 85% of new jobs and provided two thirds of total private sector employment in the EU, ranging from 53% in the United Kingdom to 86% in Greece. An average SME has four employees. Most European SMEs are independent enterprises and do not belong to an enterprise group. However, in the sub-segment of medium-sized SMEs, the entities are very often part of a group, mostly in manufacturing and to a lesser degree in knowledge-intensive business services. These dependent enterprises generate a large proportion of the total growth generated by SMEs. Overall, SMEs create a higher proportion of value added in the services sector than large enterprises.

The five most important sectors for SMEs are (i) manufacturing, (ii) construction, (iii) business services, (iv) accommodation, and (v) food and wholesale and retail services, which together account for 79 % of all SMEs, 71 % of SMEs' value added and 78 % of SME employment. Every year around 200 000 SMEs go into bankruptcy, affecting 1.7 million workers.

3.2.2. How could HPC solutions benefit SMEs?

Making HPC solutions accessible to SMEs can be a tremendous differentiator, as it can boost productivity and innovation in the SME sector to an unprecedented level and provide businesses huge competitive advantage. In addition to running simulations, modelling, HPC can also enable SMEs perform data analytics such as "Big Data" services in a cost-efficient way. An international study of industrial value chains has shown that reputed OEMs (original equipment manufacturers, the firms such as Boeing or GM) ensure that SMEs (located in the value chain) have adequate access to and facility with OEM’s high-performance computing environment. One of the key reasons for this decision is to be certain that OEM’s key supplier base can interface with their product development systems and also produce the most innovative and cost-effective parts and components of their own.

In this section, we will illustrate some real-world examples from three European SMEs providing services across different industry sectors – Industrial Design, Media and Information Technology. The examples should provide an understanding of how HPC resources can help these SMEs in driving innovation that can have a profound positive impact on not just on national economies but at a larger level, the societies, and the quality of human life.

Example #1 – RECOM Services [source: European Commission, 2015]

RECOM Services, a 12-person SME based in Stuttgart, Germany, started out doing combustion modelling for large-scale power plants and now does designs and process optimization for a wide range of industrial furnaces and boilers. This minute company could not justify buying an HPC system and instead performs simulations on supercomputers at the nearby High Performance Computing Centre Stuttgart. This has enabled RECOM to grow revenue quickly, including recent moves into the U.S. and Asian markets.

Example #2 – Sicos / M.A.R.K. 13 [source European Commission (2015)]

Sicos subsidiary M.A.R.K. 13 is a 45-person German media company focused on designing movies. This business requires swift reaction time and extremely high quality. When approached to do work for the Australian-German animated movie based on the internationally known book "Maya the Bee," M.A.R.K. 13 entered into collaboration with Germany's High Performance Computing Centre Stuttgart (HLRS) to guarantee high-quality, on-time production of 3D pictures for the 79-minute movie. Although the movie only required about 1% of HLRS resources, it could not have been done without high investment costs by such a small company. Using HLRS resources not only helped speed up the works but also substantially reduced the financial risk for the customer.

Example #3 – SME S1 in Slovenia [source: Marolt et al. (2016)]

SME S1 was founded in beginning of 1990's as internet service provider and has transformed over the years into high tech IT service and solutions provider in the field of HPC. Today it offers services of supercomputer infrastructure to their clients and provides them system administration, optimization and parallelization of code, cloud computing services, web and mobile SaaS services. SME S1 provides their clients end to end IT solutions for their problems through either offering services on their own HPC platform or they guide clients for building their own HPC infrastructure. It also offers its clients classical IT system administration, software application development services. Initially they started with leasing only infrastructure for HPC, but this did not generate enough revenue to cover the high maintenance costs, let alone to create profits. Therefore they started to lease HPC along with specialized services in the cloud and focused mostly on manufacturing SMEs. This redesign of business model helped SME S1 to adjust the value proposition to its potential customers.

3.2.3. Example HPC Case Studies for SME Businesses

Fortissimo is a collaborative project that enables European SMEs to be more competitive globally through the use of simulation services running on a High Performance Computing cloud infrastructure. The project is coordinated by the University of Edinburgh and involves more than 100 partners including Manufacturing Companies, Application Developers, Domain Experts, IT Solution Providers and HPC Cloud Service Providers from 14 countries. These partners are engaged in over 90 experiments (case studies) where business relevant simulations of industrial processes are implemented and evaluated. The project is funded by the European Commission within the 7th Framework Programme and Horizon 2020 and is part of the I4MS Initiative. More information on this project is available at https://www.fortissimo-project.eu/. In this section, five example case studies drawn from Fortissimo project are demonstrated.

In addition to the five case studies illustrated here, additional case studies are available at https://www.fortissimo-project.eu/success-stories . These case studies addressing a variety of industry sectors and application domains are highly interesting, relevant, and useful for SME businesses. You are strongly encouraged to read all of them, if time allows. The case studies can possibly provide some practical insights into the business challenges encountered by SMEs, the solutions offered by HPC, and the potential benefits / impacts of these solutions. These case studies can be further be discussed during the face-to-face workshop following the study of this self-study material.

Case Study 1: HPC-Cloud-based simulation of steel casting
Organisation (s) Involved: Ergolines, an SME, is a world leader in the manufacture of a wide range of products specifically designed for the production of speciality steels, including electro-magnetic stirrers and special instrumentation designed around the requirements of a continuous casting facility. Ergolines' goal is the development of equipment supporting the production of flawless steel alloys with metallurgical properties able to satisfy an ever increasingly quality-oriented market. In the development of such equipment, Ergolines routinely simulates the flow of liquid steel, as it becomes a solid mechanical structure, using in-house computational resources. This case study addresses the problem of slag carry-over from the ladle to the tundish which is a serious problem in steel casting and which can lead impurities in steel or poor ladle yield. Slag carry-over is a complex phenomenon which cannot be observed directly. The simulation of slag carry-over requires the use of HPC which has not previously been used by Ergolines.
The Challenge: In the field of continuous casting there is an increasing industrial demand for the development of new technologies for preventing slag transfer from the ladle to the tundish. Such an event may cause a breakout, that is the breaking of the solid skin of the solidifying cast products, which results in hazardous dispersion of liquid steel within the industrial plant. Ladle-slag monitoring is currently performed by operators on an empirical basis. Given the relevance of both safety and the economic implications of a breakout, there is a

Case Study 1: HPC-Cloud-based simulation of steel casting
significant demand for an effective, automated system for ladle-slag monitoring. While passing through the ladle shroud, liquid slag induces characteristic vibrations which can be measured. In order to develop an effective detection system, it is necessary to correlate the vibrational signal with the fluid dynamics of the system. Such a correlation requires a complex, detailed simulation, which can only be carried out on an HPC system.
The Solution: Dedicated HPC-based simulations followed by case experimental validation have provided Ergolines with key insights into the physics of the system and into different ladle-emptying mechanisms. As a result, it has been possible to establish a correlation between the shroud vibrational signal and the fluid dynamics of the system. The results obtained constitute the basis for the development of an innovative slag monitoring technology based on vibrational analysis, which would significantly contribute to both better occupational safety and greater productivity of steel plants. Previously Ergolines was using simulation in its design process. However this case study was their first experience of HPC and the benefits it could bring.
Business Benefits and Impact: Given the complexity of the phenomenon to be simulated, a very fine discretization in terms of geometry and time is needed in order to obtain accurate results. Such a fine discretization involves a significant computational load and therefore requires adequate computational capabilities. As the company does not possess the necessary computational infrastructure, the possibility of using Cloud-based HPC resources proved fundamental in addressing this specific industrial and scientific challenge. In fact, the availability of a cloud-based HPC system allowed Ergolines to exploit supercomputing resources and reduce computational times without having to sustain the high costs of a dedicated infrastructure, used for only part of the time. The use of such an HPC resource can contribute to a significant reduction in time to market and improved product design. The results attained by the HPC-based fluid-dynamic analysis set the stage for the development of a new product for automatic slag detection in steel continuous casting, a promising technology envisioned to bring significant benefits to the end-users both in terms of occupational safety and productivity of steelworks. The ability to detect slag while it is passing through the shroud would enable a steel plant to control the closing of the ladle better and so increase the steel yield. For an average ladle size of 100 tons, usually 0.5 to 1% of steel remains in the ladle. Using the proposed slag monitoring technology, 60% of that lost steel can be saved. On an average production of 1 million tonnes of per year, a medium-size factory could then save 6,000 tonnes of steel that do not need to be re-melted. The re-melting of 6,000 tonnes of steel would cost approximately €70 to €100 per tonne, namely 420,000€ to 600,000€. Additionally the loss of a further 300 tonnes of steel for a cost of 70,000€ could be avoided. This means a total saved amount up to 670,000€ saved per year per medium sized steel plant. Casting is a high energy-consuming process. It is very easy to see what this means in terms of energy saving for the re-melting of 6,000 tons of steel of each steel plant equipped with

Case Study 1: HPC-Cloud-based simulation of steel casting
the proposed monitoring technology.

Case Study 2: Cloud-based optimisation of a multi-body wave energy device
Organisation (s) Involved: Zyba is a UK registered SME with expertise in digital and numerical modelling. Since 2014, Zyba's primary goal is to engineer beautifully simple systems that work with nature. Together with BioRock Technology, Zyba is harnessing wave energy to rotect coasts from erosion, enhance coral reefs and deliver sustainable power. HPC resources and expertise in this experiment were provided by Arctur.
The Challenge: Waves are the primary cause of erosion, yet they carry a readily accessible supply of energy. Zyba's core product, CCell, is a curved wave energy converter designed to be efficient, light, and simple to assemble and operate. CCell moves with the waves to simultaneously extract their energy and reduce their impact on the beach, while generating electricity. BioRock reefs are wire mesh frames that can be built in any shape or size. Secured to the seabed, they use the safe, low-voltage electrical charge from CCell, to form limestone rock from seawater minerals. These reefs provide a habitat for corals to grow at an accelerated rate, creating an active breakwater to protect shores from erosion and enhance beaches. Optimisation of the CCell device for each site is heavily dependent on the local wave resources and bathymetry. However, modelling waves is a computationally intensive task, requiring computational simulation in excess of most SMEs' budgets. Physical testing is a potential alternative, but it is also expensive and the limited availability of facilities constrains the rate at which tests can be undertaken.
The Solution: Simulations of the CCell device in different incoming wave conditions were carried out using OpenFOAM on an HPC machine. An easy-to-use GUI was developed which allows simulations to be set up quickly, with a series of scripts/tools written to streamline the workflow on the HPC system. The increased computational power also allowed the whole CCell system to be modelled for the first time, including power-take-off hardware, software, and the intended control logic. This provided new insight into how the different pieces of the system work together.
Business Benefits and Impact: • Reduced simulation set up time from 2 hours to less than 1 minute. • Mitigated use of physical modelling, providing a nine fold reduction in cost. • Reduced HPC costs from 0,09€/core-hour to 0,05€/core-hour as all software used was open source. • Ability to concurrently run simulations improved productivity by a factor of 7 As a result of the experiment, Zyba has been able to increase its productivity, and deploy a

Case Study 2: Cloud-based optimisation of a multi-body wave energy device
series of pilots offshore. The automated and streamlined design process enabled it to rapidly optimize each design, reduce design costs and development time. With the new GUI, less experienced users can now complete design tasks, which empowers individuals within the team and increases the rate of innovation. Overall the new system has enabled Zyba to significantly speed up its market entry (saving perhaps 9-12 months), and will continue to facilitate the evolution of future products and service offerings. Provision of resources from an HPC centre transforms an imposing capital expense to a more manageable operating cost. Administrative tasks associated with IT management and upgrades are also effectively outsourced, alleviating pressure on a small SME team and reducing interruptions to design work.

Case Study 3: OptiBike Robust Lightweight Composite Bicycle design and optimization
Organisation (s) Involved: IDEC (End User) is in an SME focused in composites and new materials engineering. Its main strength is the development and industrialization of Advanced Composite Structures by Resin Transfer Moulding technology. NOESIS (HPC Expert) is an engineering partner from Belgium. UNITO (HPC Expert) is an Italian university from Turin that provided HPC expertise together with NOESIS in order to achieve better performance. ARCTUR (HPC Provider), the Slovenian HPC centre, is the HPC Provider.
The Challenge: Composite structures, and especially those incorporating carbon fibres, are much more complex than isotropic metal alloys, as they require lengthy development, significant knowledge and fine tuning. Building prototypes to test various configurations may take years to reach an optimal structure, which is, of course, not affordable. Numerical simulations can significantly reduce the time and effort required, but modelling software tools and HPC infrastructures represent a large investment and solving optimization problems with hundreds of parameters require highly skilled engineers.
The Solution: The OptiBike experiment focuses on providing SMEs with a user-friendly service to optimize the configuration of the layers of a composite material part, returning, in a reasonable timeframe, the best performing orientation of the layers that respects manufacturing constraints. In order to speed up the process, this service is deployed on a cloud-HPC infrastructure and it leverages machine learning algorithms to exploit and enhance the expert's knowledge about performance and manufacturability of the configuration. The OptiBike service, once it has defined the best-performing configuration, is able to analyse the robustness of the optimal design with respect to the uncertainty related to manufacturing tolerances. The approach is validated on a use case that involves the simulation of stresses and deformations on a composite bike frame.
Business Benefits and Impact:

Case Study 3: OptiBike Robust Lightweight Composite Bicycle design and optimization
• The end user has reduced the development time by 80% and number of physical prototypes by 75%. • €45,000 saved by the end-user per bike frame model. • 9 new customers expected by the end user within 5 years after the end of the project. The OptiBike solution leads to a reduction of 80% in the time to design and optimise a bicycle that can currently take up to 8 months. In addition, it reduces the number of the manufactured prototypes by 75%. Having this optimization workflow available, means that IDEC is capable of staying ahead of the competition despite its limited resources. This is particularly relevant in volatile sectors such as competition bike manufacture. The reduced development time, will also allow IDEC to react more quickly to modifications of the design coming from either the customer or an in-house department. On the other hand, Noesis will be able to provide this composite optimization service to customers looking for a general-purpose solution for their composite parts. Moreover, the methodology on which the service is based, enriches Noesis' offer of engineering services and is meant to attract customers that are looking for a tailored solution for very high-end composite parts. Presence in the Fortissimo Marketplace will allow Noesis to reach out to a number of customers that are not yet in its sales network. Arctur, in turn, foresees a continuation of efforts that have been established within the overall Fortissimo project, as it will continue to provide their HPC infrastructure and services through the Fortissimo Marketplace.

Case Study 4: HPC based high-resolution modelling of magnets
Organisation (s) Involved: Magneti Ljubljana is a Slovenian SME that has produced permanent metallic and systems magnets for the European market for over 60 years. These have many uses in a variety of sectors. XLAB is a Slovenian R&D company with a strong research background in the fields of distributed systems, cloud computing, system security, information visualization and image processing.
The Challenge: Magneti produces its magnets through a process called compaction, which uses a hydraulic press to apply pressure to magnetic powders until they solidify. The hydraulic press is made up of several very expensive parts which regularly wear out and must be replaced. The pressing tool needed to be optimized, so it could be used for longer and with lower material costs, but doing this requires the ability to automatically detect yielding of the tool under a given pressure. This requires many iterations of computer simulation and post-processing, which exceeds in-house capabilities.
The Solution:

Case Study 4: HPC based high-resolution modelling of magnets
To tackle the problem of pressing tool optimization, XLAB developed a set of software services based on open-source solutions. XLAB built a computer model of the pressing tool and its behaviour during the compaction process. This model is highly configurable, so Magneti can reuse it for other applications. The optimization service runs as a web application, which provides an easy-to-use interface. The application connects to Arctur's HPC system and submits an HPC job according to configuration and input parameters from the end user. This means that even inexperienced users are able to design and run experiments using HPC resources, avoiding the need for costly training. Magneti only needs to pay for the computing resources it uses, providing the company with a cost-effective solution.
Business Benefits and Impact: • Design of a better pressing tool that saves money, does not need to be replaced as often, and uses less energy. • Savings of over €100,000 per year for Magneti and ability to create new services based on improved pressing tool. • Magneti and XLAB both gained experience in simulation. The partners in this experiment have derived a number of benefits from their involvement in the Fortissimo project. For Magneti, there are several benefits. Due to the optimized geometric properties of the pressing tool, the quantity of excess material in an existing tool was reduced by around 32%, reducing material costs. This has reduced the cost of making the pressing tool by 27%, which represents an annual saving of €87k. The pressing tool is also of a higher quality, containing narrower coils which consume less power. Assuming Magneti replaces all of their pressing tools with those designed by the HPC tool; this will save another €16.2k annually. XLAB has broadened its software development expertise to the field of magnet production, and gained knowledge about developing complex software with the help of open source tools. The existing software developed for Magneti can be extended and modified to potential new customers coming from the same or similar industries. Arctur has increased its reputation in the research community, potentially allowing it to attract new customers from the magnet production industry.

Case Study 5: CAE driven design of a water treatment plant
Organisation (s) Involved: AKVOLA TECHNOLOGIES (End User) provides cost-effective and environmentally friendly solutions to clean hard-to-treat industrial wastewater containing high concentrations of oil (free and emulsified) and suspended solids. NAVASTO (Application Expert) is a SME, which offers engineering services in the field of aerodynamics ranging from the

Erasmus+

Co-funded by the European Union

Case Study 5: CAE driven design of a water treatment plant
conceptual design and implementation of fluid dynamics investigations, through the analysis of flow phenomena, to the development of complete product solutions. ARCTUR (HPC Provider), the Slovenian HPC centre, was the HPC Provider.
The Challenge: Enabling energy efficient and sustainable water cleaning is one of the main challenges of the 21st century, since clean water is already a rare resource in many regions of the world. To tackle future challenges imposed by fierce competition and increasing customer demands, Akvola as a first time HPC resource user, wanted research the viability of devising a water treatment plant entirely based on CAE technology from conception to start of production. The major goal of this experiment was to gain the ability to simulate an industrial-water treatment plant with sufficient accuracy using HPC. This would enable the end-user Akvola to improve their design process from an experiment-based workflow to a more economic CAE-based design process. The simulations have to deal with the complex three-phase flow inside the waste-water treatment device. This requires significant computing power and a validation of the numerical model against experimental data.
The Solution: During the experiment, investigations were conducted on a prototype plant to enable a calibration of the simulation process. After developing a reference cases for several aspects of the simulation, the resulting flow fields match very well with the experimental data. The validated and fully automated numerical workflow then could be used to identify regions inside the wastewater plants where the flow structures are unfavourable for the floatation-filtration process. Subsequent improvements can be evaluated with very short turnaround times. This speedup in the design process is enabled by the combination of the open-source software OpenFOAM in combination with cloud-based HPC.
Business Benefits and Impact: • The advanced hydrodynamic design of the wastewater-treatment plant increases the filtration efficiency by 20%. • The smaller size of the plants and the reduced energy and chemical consumption shortens the time to positive return on investment. • The development costs and development time for new products are reduced. Identifying the shortcomings of akvoFloat (a hybrid water treatment process consisting of flotation and filtration) and their mitigation through Computational Analysis saves time and energy for plant optimisation. This results in lowering the price of the plant and reduces the capital investment needed by end-users for new technology. Also, CFD can be used in adapting the plant to the type of the water, since every industry creates water of different qualities. Additionally, CFD-induced changes in the plant design reduce chemical and energy consumption and, in that way, reduce operational costs. These reductions translate into shorter payback times, which is the biggest driver for a sale. The increased expertise in the field of multiphase simulation, hydrodynamic simulations and simulation process design

Case Study 5: CAE driven design of a water treatment plant
will help Navasto to showcase their competencies and capacities at conferences, trade fairs and on the company's website in order to increase the visibility of the company. This can be used to reach out to new customers and increase business. Potential end-users are manufacturers of water treatment devices and companies that design or manufacture devices with complex internal.

3.3. HPC Adoption Challenges for SMEs

3.3.1. Access to Expensive HPC Platform

IDC studies during the past decade have consistently shown that HPC can accelerate mission-critical innovation and competitiveness for companies of all sizes and in a broad, expanding range of markets (European Commission, 2015). In spite of such evidence, many SMEs that might have gained from HPC technology remain unaware or inadequately informed about the contemporary realities and opportunities of HPC. It can be argued that whilst HPC is available to large companies, the costs of owning (capital expenditure) and maintaining a suitable infrastructure (operating expenditure) are beyond the financial capabilities of most of the SMEs. So metaphorically, HPC adoption by the SMB community is like a car running with both the accelerator and brake pedals depressed.

Currently, HPC systems have become more affordable with prices starting at under $100,000 (approximately €90,000). However, purchasing HPC systems is not the only option open to SMEs with high performance computational needs. In addition to the issue of affordability, consistent and effective usages of the infrastructure also pose essential business considerations for owning such infrastructure. The possibilities for accessing supercomputers and HPC infrastructure at national (government) HPC centres, HEI HPC centres, and also from public cloud providers provide cost-effective and viable options for SMEs. Public cloud computing can be an attractive alternative for SMEs that haven't invested in “on-premise” HPC resources for their workloads. Many SMEs rely entirely on public clouds for solving their business problems. Nonetheless other SMEs turn to large, national HPC data centres for more powerful computing resources and expertise. A major advantage these centres typically offer is access not only to powerful HPC systems but also to HPC experts who can help SMEs run their problems on these supercomputers.

3.3.2. Knowledge and Skills related to HPC

One of the major barriers for HPC adoption by SME businesses are cited as general lack of knowledge about how to apply HPC tools to solve engineering or business computational challenges – especially since many SMEs lack training in Information Technology (IT) and Computational Science areas such as simulations and modelling (Ezell and Atkinson, 2016). Therefore, it remains vital to deliver skills and HPC competencies necessary for SMEs to be ready to engage into knowledge-intensive and high value-added entrepreneurial activities. Higher Education Institutions (HEIs) can play a pivotal role in such skill development and knowledge transfer process. To begin with, it is necessary to conduct a systematic audit of the hard infrastructure and preparedness of both HEIs and enterprises – in order to identify the specific knowledge that needs to be transferred to SMEs. Some of the techniques that will be employed to deliver this innovative approach will be survey analysis of HEIs, analysis of enterprises' needs; identification and clarification of a knowledge-transfer needs analysis; the tailored design and delivery of knowledge transfer to all stakeholders engaging in the process; the measurement of outcomes from the knowledge transfer process; and the reiteration of the knowledge transfer process so as to continuously propagate the engagement between HEIs and enterprises in specific regions. Subsequently, HEIs can train concerned SME stakeholders through carefully designed courses and engage with SMEs so that the SMEs could recognise possibilities, potential commercialisation opportunities, and increased efficiencies and competitiveness through the use of HPC.

3.4. Initiatives on HPC Adoption by SMEs: Regional and International Perspectives

3.4.1. HPC Adoption Initiatives for SMEs: EU Countries

In February 2012, the European Union released a report titled "High Performance Computing: Europe's Place in the Global Race", which articulated a goal (and set of action plans) to achieve HPC leadership including by "acquiring at least one Exascale computer in the same timeframe as the U.S., Japan, and China (European Commission, 2012). Europe's HPC strategy recognizes that "HPC is a strategic resource for Europe's future" with "computational science already the 'third pillar' of science" and "industry relying more and more on HPC to innovate in products and services”. In accordance with this initiative, the European Union has significantly increased funding for HPC systems research and development, including launching the European Technology Platform on High-Performance Computing (ETP4HPC) in 2012, part of a €700 million public-private investment in HPC through the Contractual Public Private Partnership (cPPP) initiative. That investment has been complemented by €400 million in commitments (mostly in-kind contributions) made through the Partnership for Advanced Computing in Europe (PRACE). Reviewing Europe’s efforts since 2012 to bolster its HPC capabilities, market research firm IDC noted in late 2015 that, “The European Commission, PRACE, and ETP4HPC have done an admirable job of advancing Europe’s position in the few years since the 2012 Communication”. IDC notes that, historically, “Europe’s HPC community has been more oriented toward science than industry and industrial access to Europe’s leading supercomputers has grown but remains limited.” However, Europe is working to address this, with the PRACE, SME HPC Adoption Programme in Europe (SHAPE), and Fortissimo initiatives focused on expanding industrial use. Thus, while Europe does “have some world-leading HPC centres for collaborating with industrial users, including SMEs (such as HLRS, Teratec, SURFsara, CINECA, and LRZ)” democratizing access to HPC resources for industrial purposes remains as much a challenge in Europe as it is in the United States. The major thrust of the European HPC strategy remains positioning its HPC investments so it can become a leader in HPC software and applications development. In other words, Europe appears to be trying to position itself to lead in HPC adoption broadly, and in terms of production, to focus on the software rather the hardware side.

Among the world's premier national HPC data centres, the High Performance Computing Centre Stuttgart (HLRS) is situated in the heart of Germany's auto industry. HLRS is seeing increasing demand for HPC from SMEs. Especially in this region; SMEs serve as technology solution providers for larger companies. Increasingly, these large clients require a validation of their technology through simulation. In certain fields, simulation can play a crucial role but is not well known inside tier 1 companies. Very small companies with very special knowledge in modelling and simulation make a living in these small market niches, but they need access to large-scale systems for the computational part of their portfolio (European Commission, 2015).

Recently, the PRACE Council has decided to make SHAPE (SME HPC Adoption Programme in Europe) a permanent service. There have been several initiatives undertaken by this SHAPE programme to help make HPC use more pervasive among Europe's SMEs, in order to increase their competitiveness and ability to contribute to the European economy. This initiative addresses the widespread concern among European HPC stakeholders surveyed that greater outreach to SMEs in needed because relatively few seem to be aware of the value of HPC or the opportunities available to SMEs through PRACE. The projects the SMEs propose for HPC use would not need to be of special scientific or technical interest—they would only need to be important for advancing the SMEs' business prospects. Awards would be based primarily on this importance factor, along with the SMEs' commitment to the process. The initiative would be governed by the same groups that govern the SHAPE program and would be given access to compute cycles and expertise sufficient to accommodate, say, a dozen SMEs in the first year of the initiative's operation. The experiences would be monitored and reviewed to determine whether the initiative should be continued for a second year, with or without changes.

3.4.2. HPC Adoption Initiatives for SMEs: USA

While large manufacturers have made great progress in leveraging HPC for innovation, the penetration of HPCs into America's SME manufacturing base has been rather slow and sporadic (Ezell and Atkinson, 2016). SME manufacturers (those with fewer than 500 employees) account for about 98.5 percent U.S. manufacturing companies as of 2013. Moreover, 94 percent of all U.S. manufacturers employ 100 or fewer workers. This vast number of SMEs constitutes the so-called "missing middle" of HPC adoption in U.S. industry. This term "missing middle" refers not directly to company size but rather to a company's computing capacity; the term specifically refers to the group of HPC users between low end, mostly workstation-bound HPC users, and the kind of high-end HPC uses typically performed at national labs and some universities. Nevertheless, in industry parlance, the term has come to refer to the wide swath of small- and mid-sized manufacturers who could be leveraging HPC in their product development or manufacturing processes, but are not. For example, one 2013 study estimated that only 8 percent of U.S. manufacturers with fewer than 100 employees are using HPC. Earl Joseph, an HPC analyst at IDC, estimates that at least 25,000 U.S. manufacturers, the vast majority of these SMEs, would benefit from having access to HPC-empowered modelling and simulation tools in design, prototyping, and testing of their parts, components, and finished products.

Three major barriers that prevented America's SME manufacturers' adoption of HPC solutions are identified. First, there exists a general lack of knowledge about how to apply HPC tools to solve engineering challenges, an especially acute problem because many of the engineers working at America's SMEs simply were not exposed to computational sciences in their electrical or mechanical engineering training. Second, taking those engineers "off the line" to train them in modern modelling and simulation tools takes them away from the urgent needs of the business and represents an expense many SMEs cannot incur. Third, and more subtly, many existing modelling and simulation packages (e.g., designed to model aircraft and engines) are often too complex or overdesigned for the needs of smaller manufacturers.

In recognition to these gaps in addressing the high-performance computing needs for "missing middle" segment of SMEs and also, to leverage the opportunities for a potentially big market for HPC solutions, a few important initiatives have been launched at both regional and national levels. The initiatives primarily focus to help remedy the lack of availability, accessibility, or approachability to HPC tools for SME manufacturers. For example, the National Centre for Manufacturing Sciences (NCMS) has created a dozen centres throughout the United States (located near universities and national labs to tap into local expertise) to connect manufacturing firms with HPC resources. NCMS's network of "Predictive Innovation Centres" represents public-private collaborations providing U.S. manufacturers with HPC tools aimed at increasing product design cycles, improving manufacturing processes, and reducing the need and costs of laboratory testing of new products. Likewise, the Ohio Supercomputer Centre's (OSC's) AweSim program is a partnership among OSC, simulation and engineering experts, and industry to assist SME manufacturers with simulation-driven design to enhance innovation and strengthen economic competitiveness.

As AweSim Director Alan Chalker explains, "Simulation-driven design replaces physical product prototyping with less expensive computer simulations, reducing the time to take products to market, while improving quality and cutting costs. Smaller manufacturers largely are missing out on this advantage". AweSim levels the playing field by giving smaller companies equal access to HPC technologies. AweSim invites SMEs to bring in their technical challenges and then work with experts to understand how HPC-enabled modelling and simulation tools can help solve their problems. The National Centre for Supercomputing Applications, a hub of transdisciplinary research and digital scholarship led at the University of Illinois at Urbana-Champaign, has also played a pivotal role in helping U.S. enterprises, large and small alike, understand how they can leverage HPC tools to bolster their competitiveness. Likewise, the Chicago-based Digital Manufacturing and Design Innovation Institute (DMDII), one of the institutes within the National Network for Manufacturing Innovation (NNMI), has developed a new cloud-based system to democratize SME manufacturers' access to HPC resources. DMDII envisions its Digital Manufacturing Commons (DMC) as a free, open-source software project to develop a collaboration and engineering platform that serves as an online gateway for digital manufacturing. Akin to an "app store for manufacturing," the DMC sits in a digital services marketplace with a software development kit and collaboration platform at its core, essentially equipping SME manufacturers with the modelling and simulation tools they need to address technical design challenges as well as access to shared HPC resources.

3.5. HPC Centres in HEIs – A Business Model Perspective

3.5.1. Generic Business Models for HEI HPC Services

HPC Centres are vital part of the academic infrastructure of HEIs with critical importance for academic research. While HEIs incur high capital expenditures in running those centres, the business aspects are often missing. In their empirical research, Eurich et al. (2013) try to seek an answer to the question of ''on what business models do HPC centres run'' and they suggest four generic business models for HPC centres in higher education institutions. The four generic business models - library, shareholder, cost centre, and industry collaboration are described in Table 3. The business models primarily differ in terms of revenue streams and the stakeholders involved. In practice, however, most university HPC centres combine these generic types and run on a hybrid model - designed based on considerations of HEI's academic / research objectives and business constraints.

Business Model	Idea	Revenue Stream	Example
Library	HPC is part of HEI infrastructure	Subsidies	University of Southampton
Shareholder	Group of shareholders who need computing power pool their money to finance a large common cluster	Share	ETH Zurich
Cost Centre	HPC centre is fully self-sustained and runs like an industrial cost centre with the need to compensate for all costs	Pay-per-use	University of Cambridge
Industry Collaboration	Collaboration with OEMs, granting industrial partners access to the HPC infrastructure	Funding from OEM, revenue From industrial partners	Dell / Cambridge HPC Solution Centre

Table 3: Generic Business Model Types for HEI HPC Service
Source: Adapted from Eurich et al. (2013)

A. Library Model

In this business model, HPC is considered as being a part of the academic infrastructure of the HEI. The basic logic of this business model is similar to the one of a library – just as in a library a user can borrow a book from a basic collection without charge (provided the book is available for issue), so in HPC service provisioning a user can consume an HPC service from a basic set of services without charge (provided there is enough capacity). The business model runs on a completely subsidised approach (indirectly everyone pays for HPC) and enables exploration of unfunded ideas and research directions. The success of this model depends on the commitment of university in terms of investments and expenditures.

B. Shareholder Model

The basic idea of the shareholder model is that a group of ‘‘shareholders’’ who need computing power pool their money to finance a large common cluster. The HPC centre’s management or an IT portfolio manager conducts the acquisition, maintenance and operation of the system, which, in this model, is mostly a cluster that can be scaled according to the needs of its ‘‘shareholders’’. Each shareholder receives a share of the cluster proportional to his or her investment. Unlike the library model, competition plays a role here as users could be motivated to buy their own hardware or outsource their jobs to an external commercial vendor like Amazon or Google. While fair, proportional and prioritised distribution of services can be applied, non-users are not encouraged to experiment with HPC in this model. Problems might occur when a shareholder leaves the consortium.

C. Cost Centre Model

In this business model, an HPC centre runs like an industrial cost centre that remains fully self-sustained and compensate for all incurred costs, including capital depreciation of the machines, staff costs, power, and energy. Users of HPC services are charged on a “pay-per-use” basis. One of the key assumptions of this model is that HPC users have a choice of facilities / resources where to spend their money. Due to this competitive environment, HEI HPC centre has a high incentive to attract and retain its customers, and to provide value for money and quality services. Accordingly, there is a pressure from the customers to optimise cost and respond to their needs. Since the cost centres invest for tomorrow based on today’s earnings, technology infrastructure upgrade maps could be problematic. Also, researchers may feel worse off than their colleagues at other institutions which provide HPC services for free.

D. Industry Collaboration Model

This business model is founded on a long-term collaboration either downstream, with an original equipment manufacturer (OEM) or upstream, by granting industrial partners access to the HEI’s HPC infrastructure. A partnership with an OEM can be beneficial for both sides- while the OEMs that lack a large internal research and development base can benefit from the operational knowledge and real world experience of the HPC centre, HPC centre benefits from being subsidised by the OEM. The HPC infrastructure can also be open to the industry and to sell them particular services. HPC service sites offer an interesting value proposition to organisations that have particular but infrequent problems which require supercomputing power and knowledge about the HPC infrastructure and the applications. Standard HPC services can be provided as well in order to harmonise the utilisation of the HPC infrastructure. The advantages of this model are that the HPC centre is independent from grants and the university's commitment and that it allows for researchers to consume a subsidized service. On a downside, the HPC centre remains dependent on the OEM and the usage of industrial organisations.

3.5.2. HEI-SME HPC Collaborative Business Model

The purpose of including this section in this self-study course is primarily to stimulate ideas from the readers on a pivotal aspect of the **SME/HPC project** – **co-creation of knowledge between HEIs and SMEs** through co-design and co-implementation of HPC solutions – to drive innovation capability and capacity in both HEIs and enterprises. The ideas and perspectives from the readers (and also future participants of workshops) will be elaborated and discussed in detail during the face-to-face workshops planned following completion of this self-study course.

It can be reasoned that a business model-centric discussion should help SME/HPC partners, especially enterprises, to integrate the ideas and concrete actions into their business models- in relation to value co-creation and value realisation (commercialization). While the collaborative effort of the SME/HPC project partners affects most of the key elements of their concerned business models, such as customers, resources and capabilities, offerings and their value propositions and finance, the focus of this discussion is anchored on two elements - **resources and capabilities**; and **offerings and their value propositions**. The main obstacles to accessing and using existing expensive HPC infrastructure and vibrant cooperation between HEIs and enterprises are the lack of competencies of SMEs to fully recognise, understand and appreciate the HPC potential. This gap is further widened by the general lack of awareness of HPC as an emerging technology that can increase SME innovation competencies, capacities and capabilities to improve their enterprises' competitiveness. Brainstorming and elaboration on the two business model elements (highlighted above) should help in moving towards a solution that addresses these gaps.

3.6. Reflections on this module learning

It is quite common to find that SMEs lack mature IT infrastructure (hardware, software etc.), governance (policy and processes) and digital skills. How do these factors affect adoption of HPC solutions by SMEs?

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

What could be some of the enablers for HPC adoption by a typical SME?

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

How could SMEs spot opportunities (with business case) for HPC in their businesses?

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

3.7. Recommended Further Readings / Resources

1. https://ec.europa.eu/digital-single-market/en/high-performance-computing
2. https://www.fortissimo-project.eu/About
3. http://www.europarl.europa.eu/RegData/etudes/IDAN/2017/603967/EPRS_IDA(2017)603967_EN.pdf
4. http://www.prace-ri.eu/prace-industry-2019/
5. https://www.smallbizlabs.com/2014/08/dispatches-from-the-new-small-business-economy.html
6. Hall, J. (2018). What is the definition of an SME? Available at : https://www.simplybusiness.co.uk/knowledge/articles/2018/12/what-is-an-sme/
7. Ed Turkel, CIO, "Democratizing high performance computing," March 21, 2017. Available at : https://www.cio.com/article/3183523/democratizing-high-performance-computing.html
8. IDC, "IDC HPC ROI Research Update: Economic Models For Measuring The Financial ROI And Innovation From HPC Investments," June 2016. Available at: https://science.energy.gov/~/media/ascr/ascac/pdf/meetings/201612/IDC_DOE_ROI_Research_Update_12-16-2016.pdf
9. William Payne, Scientific Computing World, "Modular HPC goes mainstream" October 1, 2016. Available at: http://ftp.scientific-computing.com/feature/modular-hpc-goes-mainstream
10. Stan Gibson, SearchCIO, "High-performance computing use cases and benefits in business" April 2019. Available at: https://searchcio.techtarget.com/feature/High-performance-computing-use-cases-and-benefits-in-business
11. http://primeurmagazine.com/weekly/AE-PR-02-15-1.html

Videos / Podcasts

https://www.youtube.com/watch?v=p2v7aJeG7Xc

https://insidehpc.com/2018/09/video-cloud-based-hpc-optimization-manufacturing-processes/

References

Almeida, S. (2013). An introduction to high performance computing. *International Journal of Modern Physics A, 28*(22n23), 1340021.

Attaran, M., & Woods, J. (2018). Cloud Computing Technology: A Viable Option for Small and Medium-Sized Businesses. *Journal of Strategic Innovation and Sustainability, 13*(2).

Barney, B. (2014). Introduction to Parallel Computing [online]. Available from: https://computing.llnl.gov/tutorials/parallel_comp/

BIS. (2013). Department for Business, Innovation and Skills, E-infrastructure: The ecosystem for innovation. 2013, London, UK.

Brochard, L. (2006). High performance computing technology, applications and business. *Informatik-Spektrum, 29*(3), 191-200.

Eurich, M., Calleja, P., & Boutellier, R. (2013). Business models of high performance computing centres in higher education in Europe. *Journal of computing in higher education, 25*(3), 166-181.

European Commission (2012). High-Performance Computing: Europe's place in a Global Race [online]. Available from: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0045:FIN:EN:PDF [Accessed: 23 April 2019]

European Commission (2015). Study on "High Performance Computing in the EU: Progress on the Implementation of the European HPC Strategy" [online]. Available from: https://ec.europa.eu/digital-single-market/en/news/study-high-performance-computing-eu-progress-implementation-european-hpc-strategy-final-report

European Commission (2018). High Performance Computing - best use examples [online]. Available from: https://ec.europa.eu/digital-single-market/en/news/high-performance-computing-best-use-examples

European Investment Bank (2018). Financing the future of supercomputing: How to increase the investments in high performance computing in Europe [online]. Available from: https://www.eib.org/en/infocentre/publications/all/financing-the-future-of-supercomputing.htm

EPCC (2015). Bringing cloud-based HPC to business [online]. Available from: https://www.epcc.ed.ac.uk/sites/default/files/EPCC%20News%2077.pdf

Ezell, S. J., & Atkinson, R. D. (2016). The vital importance of high-performance computing to US competitiveness. Information Technology and Innovation Foundation, April, 28.

Foster, I., Zhao, Y., Raicu, I., & Lu, S. (2008). Cloud computing and grid computing 360degree compared. Published in: IEEE conference proceedings of Grid Computing Environments Workshop 2008: 1–10. IEEE, New York.

Holmes, V., & Newall, M. (2016). HPC and the Big Data challenge. Paper presented at the Safety and Reliability.

Hong, L., Zhong-hua, L., & Xue-bin, C. (2010). The applications and trends of high performance computing in finance. Paper presented at the 2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science.

Li, L. (2015). HIGH PERFORMANCE COMPUTING APPLIED TO CLOUD COMPUTING.

Livingstone, P. (2011). High Performance Innovation [online]. Available from: https://www.rdmag.com/article/2011/02/high-performance-innovation [Accessed: 23 April 2019]

Marolt, M., Lenart, G., Maletič, D., Borštnar, M. K., & Pucihar, A. (2016). Business model innovation: Insights from a multiple case study of Slovenian SMEs. *Organizacija, 49*(3), 161-171.

Needleman, T. (2009). SMALL BUSINESS: Accounting in the clouds
'Cloud' computing is all silver lining for small businesses (and their accountants, too)(News). *Accounting Today, 23*(10), 1.

Spiers, B., & Wallez, D. (2010). High-performance computing on wall street. *Computer, 43*(12), 53-59.

Sridhar, T. (2015). Cloud Computing - A Primer [online]. Available from: https://www.cisco.com/c/en/us/about/press/internet-protocol-journal/back-issues/table-contents-46/124-cloud2.html

Sterling, T., & Stark, D. (2009). A high-performance computing forecast: Partly cloudy. *Computing in Science & Engineering, 11*(4), 42-49.

STFC (2015). Accelerating the product discovery process at Unilever [online]. Available from: https://stfc.ukri.org/innovation/success-stories/accelerating-the-product-discovery-process-at-unilever/

Strohmaier, E., Dongarra, J. J., Meuer, H. W., & Simon, H. D. (1999). The marketplace of high-performance computing. *Parallel Computing, 25*(13-14), 1517-1544.

Wince-Smith, D. (2009). High-Performance Computing for All. *Issues in Science and Technology, 25*(4), 67-70.