May, 2016

Some ideas about the launch of a HPC Center in emerging countries

Félix Díez Sacristán.
CEO of InterHost

Introduction to High Performance Computing

High performance computing (HPC) encompass the entire market for computer server systems and related software, storage, networking and services employed by scientists, engineers, analysts and others to address computationally intensive and data intensive simulation and analytics problems. HPC activities can be found in industry and commerce, government and academia. Industrial/commercial activities include automotive and aerospace product development, oil and gas exploration, drug discovery, weather prediction and climate modeling, complex financial modeling, consumer product design and optimization, advanced 3D animation, advance business analytics and others.

HPC systems range from commodity clusters (<10.000 €) to 100 million euro (or plus), one-of-a-kind supercomputers (a continuous spectrum from entry-level to high-end machines). HPC is in contrast to enterprise computing as used for business operations such as accounting, payroll, sales, customer relations, transaction processing, human resources, and purchasing. HPC may be carried out on premise in dedicated HPC data centers, or in private, public or hybrid cloud environments. HPC methodologies include modeling and simulation, advanced data analytics, visualization, and others.

For the emerging countries HPC and other related technologies may help reaffirm regional and economic leadership. The use of HPC in fields such as industrial production, high value-added services, the development of new technologies and the operating of a Supercomputing Center shall result in a significant increase of productivity in the midterm.

Building a Supercomuting Center implies on one hand, big investments to acquire an HPC and auxiliary operating equipment, software development, improving business processes; and strong revenues with the sale of computing cycles on the other hand. The final net balance, in terms of return on investment (ROI) is highly positive and it is worth mentioning that an HPC will help the national scientific, academic and technological community integrate to the world's scientific and technologic mainstream and the digital market in its most extensive sense (i.e Big Data). Definitely, the development and operating of an HPC is mainly strategic to bolster innovation, digital industry development and data economy (for example, the analysis of the vast amount information generated by the Internet of Things (IoT).

Workload of HPC's applications

(From *European Commission, Directorate-General of Communications Networks, Content & Technology. Digital Agenda for Europe. High Performance Computing in the EU: Progress on the Implementation of the European HPC Strategy,* conducted by IDC)

Workload: Set of related scientific, engineering or analytical applications, along with any associated infrastructure or support applications (e.g., visualization, data management). Each application workload is characteristic of a one or more market segments. Hence, the set of application workloads for biological sciences differs in important respects from the set used in chemical engineering and other segments.

- **Biological sciences (*BioSci*)**. This workload centers around applications such as genomics, proteomics, pharmacogenomics, pharmaceutical research, bioinformatics, drug discovery, bioanalytic portals, ASP-type service providers, and agricultural research. Computational techniques include database searching and management, molecular modeling, and computational chemistry. These workloads appear in commercial, academic, and institutional research environments.
- **Chemical engineering (*Chem Eng*)**. This workload centers around applications such as molecular modeling, computational chemistry, process design, and chemical analysis. It includes all chemistry applications that are not directly related to biosciences research and development. These workloads appear in commercial, academic, and institutional research environments.
- **Computer-aided design and drafting (*CAD*).** This workload centers around applications such as mechanical CAD; 2D, 2.5D, and 3D design and drafting; 3D wireframe; and civil engineering design. Design and drafting applications require graphics capability but are less compute intensive than design engineering and analysis applications. CAD tasks are typically done by designers and drafters. Users are found primarily in discrete manufacturing industries such as automotive, aerospace, heavy machinery, and consumer goods.
- **Computer-aided engineering and mechanical design and analysis (*CAE*).** This workload centers around applications such as finite element modeling and analysis, mechanical computer-aided engineering, civil engineering, structural analysis, computation fluid dynamics (CFD), crash, NVH, and solid modeling. Like CAD applications, these CAE tasks are used to design automobiles, aircraft, running shoes, ski equipment, sail boards, beer bottles, and other everyday items. Workloads include those tasks generally accomplished by engineers, not drafters.
- **Digital content creation and distribution (*DCC&D*).** This workload centers around applications such as 2D and 3D animation, film and video editing and production, and multimedia authoring for both CD and Web pages that utilize sophisticated graphics content. This category also includes servers used for image rendering, content management, and distribution of finished products for areas such as film, TV, commercial

animation, advertising, product styling, and industrial design. These workloads are developed in large part in concert with scientific visualization research and technologies. In addition, the creation of special effects and animation for motion pictures requires significant amounts of computational capacity. Thus this category is included in technical computing based on a combination of historical affinity and computationally intensive applications.

- **Economic and financial modeling (*Econ Fin*).** This workload centers on applications such as econometric modeling, portfolio management, stock market and economic forecasting, and financial analysis.
- **Electronic design and analysis (*EDA*).** This workload covers all electrical/electronic tasks, including schematic capture, logic synthesis, circuit simulation, PCB routing, and system modeling.
- **Geosciences and geoengineering (*GeoSci*).** This workload includes earth resources–related applications such as seismic analysis, oil services, and reservoir modeling. These applications are used in both institutional research and commercial enterprises. Geoscience can also include such areas as mining, natural resource management, geographic information systems (GIS), and mapping.
- **Government laboratories and research centers (*Govt Lab*).** This workload centers on government-funded research and development institutions. These organizations are generally funded at a national or multinational level and may combine both purely scientific research with research in areas of national priority (e.g., cancer research) and/or research for defense-related programs). These users are less bound by strict economic constraints than those performing applications in product development environments. These centers don't normally offer degree programs for students.
- **National defense (*Defense*).** This workload centers around applications such as surveillance and signal processing; encryption; command, control, communications, and intelligence (C3I); geospatial image management and analysis; defense research; weapons design; and other national security applications. In addition, we believe that national security organizations are fielding applications that work to identify and track potential security threats through database-oriented pattern-matching applications. Although these applications may not always be numerically intensive, they will be developed and used by organizations that are firmly rooted in technical computing markets. In addition, we believe that these applications will be run in conjunction with traditional security applications such as cryptography and image analysis.
- **Software engineering (*Software*).** This workload centers around the development and testing of technical applications and middleware targeted (at least initially) at HPC users. The segment includes technical computers used by third-party software developers. These applications are typically performed by low-end and midrange systems, particularly low-cost technical workstations, and in some cases even PCs.

- **Technical management (*Tech Mgmt*).** This workload centers around the support of tasks such as tracking, documenting, and controlling the product life-cycle chain and the scientific research process. Tasks include product data management, maintenance records management and analysis, revision control, configuration management, network management, and project management.
- **University and academic (*Academic*).** This workload centers around scientific research and engineering R&D efforts conducted at public or private institutes of higher educations and includes systems sold for both research and educations activities. Privately funded and/or nonprofit research institutes that have a strong academic mission (i.e., work to extend the bounds of public knowledge) are also included in this segment. Applications are typically compute or data intensive and often require high-performance graphics. These users are less bound by strict economic constraints than those performing applications in product development environments.
- **Weather forecasting and climate modeling (*Weather*).** This workload centers around applications such as atmospheric modeling, meteorology, weather forecasting, and climate modeling. This segment includes systems dedicated to these tasks primarily in the government and defense segments.
- **Other (*Other*).** Any technical computing workloads not otherwise specified by the above definitions. Falling into this category are advanced business intelligence/business analytics workloads from the growing number of commercial companies that have been adopting HPC for the first time in order to perform computing tasks that exceed the capabilities of enterprise server technologies.

Technical framework for action

A suggested strategy to develop supercomputing in an emerging country shall attain the following fundamental aspects:

1. *Basic and technological services*

- Architecture of HPC system

  Singular architectures (not conventional), that use accelerating subsystems and are highly parallel. The interconnection bandwidth is critical.

  The type of architecture may cause the reengineering of applications.

- Basic software for operation and management

  Software for HPC's are complex and of hierarchical nature, with high scalability and concurrency support. A management software for hardware (clusters), resources and sophisticated tasks are required.

  HPDA functionalities shall always be considered, specially its data-centric nature and its real time availability.

  It is important to promote immersive support environments (visualization and virtual reality).

- Programming environments

  HPC scalability. Programming environments shall optimize the moving of data guaranteeing concurrency.

  It shall also be considered different environments for different application domains.

  Support mechanisms of checkpoint/restart to increase resilience.

- Operating algorithms

  Parallel, robust and library-framed algorithms.

  Advancements in this field show strong opportunities for collaboration with HPC manufacturers. In addition, software developers (ISV) and software providers (SaaS) will benefit of studies in this matter.

2. *Applications*

- Conventional :

  Bio-Sciences
  Computer Aided Engineering, CAE
  Chemical Engineering

Distribution Cost Charge, DCC & Distribution
Economic/ Financial analysis
Electronic engineering (Electronic Data Access, EDA)
Geosciences
Mechanical design
Defense
Government NatLab Research and Development (R&D)
University/Academic
Weather
Numeric calculations (Computational Fluids Dynamics, CFS) & and (Finite Element Method, FEM)
Audiovisual and animation industry (media and entertainment)

- New (such as the analysis of Big Data, HPDA)

  In such environments, the data is recollected and not generated so it has to be processed by systems with a lot of memory capacity, high storing capacity and high parallelism with short coupling. We are facing a new paradigm where distributed computations are imposed from global locations differing from the scientific simulations the centralized data is managed locally.

  The convergence of HPC and Big Data analytics is being driven by HPC users and the growing contingent of commercial firms that are adopting HPC solutions to tackle data analytics jobs that are too complex or time critical for enterprise IT resources to handle efficiently and cost effectively.

3. *Services*

- To industries
- To academia
- To ISV's
- To SME
- To Public Administrations
- To education

4. *Operating*

- Energetic efficiency

  Reduction in energy loss thanks to the reduction of data transfers. Eventually a reduction in the precision of calculations can be approached too.

  Failure-prediction tools and compliance with resilience.

  Cogeneration.

- Balance between processes, I/O, storage and networks

The increase in complexity and parallelism shall evolve parallel to the increase in storage subsystems (object oriented storage for HPC). The main objective is to ensure a proper balance (absence of bottlenecks) between Input/Output (I/O) and computing.

Use storing hierarchy-rules with solid-state technologies revealing a very high-performance in the vertex.

Performance objectives, architecture and applications (as in May 2016)

1. Install and operate a **Petascale HPC** system supercomputer with a peak power (**R_{peak}**) of approximately **1 PetaFlop/s (10^{15} flop/s).**
2. Alternatively: sufficient computing power to enter the list of TOP500 (www.top500.org) of 2017 (June) in positions between 350-500.
3. A cluster-type architecture shall be attained (easiest environment for programming means), with multicore conventional processors (e.g. **Intel Xeon** or **IBM PowerX** type), coprocessors (e.g. **Intel Phi** or **NVDIA** type) and high performance interconnections between nodes (e.g. **Myrinet** or **Infiniband**).
4. These technical specifications are necessary to have a mature equipment that can enable the offering of a vast catalogue of applications and that provides an easy learning curve (use, programming).

5. Quantitative objectives

   - **R_{peak}** of maximum **1 Petaflop/s** (1 PF.s^{-1})
   - Energy efficiency: more than **1 Gflop.s^{-1}/W**
   - Less than **1 MW of total power**
   - Minimum memory capacity: **0.25 TB/TF.s^{-1}** (this is, Terabyte/Teraflop. s^{-1})
   - Interconnection bandwidth: Maximum of **0.05 TB.s^{-1}/TF.s^{-1}**
   - Cost: **between 5,000 and 10,000 USD/TF.s^{-1}.** Depending on the services and their scope (installation, maintenance, integrated training, operating support, updating commitments, etc.)

5. Applications

   Although the objective is to acquire a supercomputer, some specific fields of activity must be strongly and decidedly approached in the phase of promotion and commercialization of computational hours.

   Action fields:

   - **Scientific calculations**. Aimed at Universities and investigation centers (normally public) that focus on basic sciences: chemistry, computational, genomic, applied mathematics, materials engineering, tissue engineering, etc.
   - **Simulations**. Aimed at the automobile industry (Finite Elements Simulation), oil industry (seismic), aeronautics (Computational Fluid Dynamics, CFD), meteorological, video and entertainment (rendering and animation), etc.
   - **Big Data and analytics**. This is meant to become a new frontier for supercomputing. There is a big amount of data generated by Internet, IoT and M2M that needs to be analyzed (correlated) and exploited.

Management, exploitation and governance models

General idea

We suggest the creation of a public/private business that is open to a strong participation of the State. The Government's participation will provide stability, credibility and future projections to the Supercomputing Center. In Europe, for instance, the consortium models are very vast and serve as a clear example (Barcelona Supercomputing Center, Spanish Supercomputing Network, CESGA, INRIA in France, etc.).

It is also very interesting to position the Supercomputing Center as a center that provides advanced computational sciences services providing for support to cybersecurity.

Exploitation

Itis recommended to provide a national-coverage infrastructure service (Infrastructure as a Service, **IaaS**) to facilitate the hosting and internet access of information systems. This initial offering can generate a big flow of revenues and clients to be complemented immediately by a High Performance Computing as a Service (**HPCaaS**) offering.

HPC public cloud computing has been mainly limited to workloads that are not latency-sensitive or IO-bound, but some cloud vendors now support applications as challenging as structural analysis and fluid-structures interactions. Data security/data loss in public clouds remain important concerns for HPC users and should be evaluated.

The advantage of the abovementioned model is that it can facilitate the establishment of a potential base of supercomputers users which will become familiar with the network-access to applications, which will believe in the security of this centers and its features and will also become familiar with the pay-per-use policies.

Governance model

The model of governance to be adopted shall be more corporate-oriented. Nevertheless, privileged access to learning institutions (universities) and governments with a non-profit view shall also be considered as a counterpart to public financing.

An excellence center shall be promoted to achieve a competitiveness increase in the country and an upgrading of the scientific-technical community without unattendingg the need to participate in a hyper-connected society which generates massive amounts of data.

Governance structure
(*Suggested by PRACE, Partnership for Advanced Computing in Europe*)

The Council is the deliberative body of the Centre and decides on all matters of the Centre. As a general rule, decisions of a purely scientific nature are subject to majority vote, while decisions related to provisioning and usage of funding and resources require a qualified majority based on partner contributions.

The Board of the Council. It consists of a Chairman, a Vice-Chairman and a Secretary.

The Scientific Steering Committee (SSC) is composed of leading researchers that are responsible for advice and guidance on all matters of a scientific and technical nature which may influence the scientific work carried out by the use of the Centre’s resources.

The Access Committee (AC) gives advice to the Board of Directors concerning the allocation of resources. The AC is composed of researchers experienced in areas of science, industry engineering and supercomputing.

The Industrial Advisory Committee (IAC) is composed of industry representatives representing different industrial sectors, e.g., Aeronautics/Aerospace, Automotive/Transport, Energy, Engineering/Manufacturing, Oil & Gas, Renewable Energy, Telecommunications/Electronics, ISV, HPC Vendors, Life Sciences, and Finances. They provide with advice on HPC usage for the benefit of country´s competitiveness and economic growth.

The Board of Directors (BoD) is the executive body of the Centre and is generally responsible for managing and representing the Centre.

Impact of HPC use

(From *European Commission, Directorate-General of Communications Networks, Content & Technology. Digital Agenda for Europe*. *High Performance Computing in the EU: Progress on the Implementation of the European HPC Strategy,* conducted by IDC).

Examples of the impact of supercomputers in the European industry:

- HPC has enabled **automakers** to reduce the time for designing new vehicle platforms from 60 months to 24 months or less, while greatly improving safety, passenger comfort and environmental friendliness. At Rolls-Royce, single components and sub-systems design/analysis/optimization through to whole engine modelling all rely heavily on HPC.
- Total recently tripled the power of its supercomputer to develop more complete visualizations of **seismic landscapes** and run simulations at 10 times the resolution of existing oil and gas reservoir models. This new capability will enable more efficient **upstream oil and gas exploration**, as well as the discovery of reserves under more challenging geological conditions.
- Swiss **pharmaceutical** giant Novartis and Schrödinger, a global life sciences and materials science software company with offices in Munich and Mannheim, Germany, hired Cycle Computing, a cloud computing services company, to test 21 million drug candidate molecules on the Amazon public cloud, using a new technical computing (HPC) algorithm Schrödinger developed.
- The Center for Pediatric Genomic Medicine at Children's Mercy Hospital, Kansas City, Missouri, has been using HPC to help **save the lives of critically ill children**. In 2010, the center's work was named one of Time magazine's top 10 medical breakthroughs.
- GE used a supercomputer to reveal a **new aspect of turbine behavior** that is already providing GE with a competitive advantage in fuel efficiency. Every 1% reduction in fuel consumption saves users of these products $2 billion/year.
- BMI utilized supercomputers and computational models to design components that could **save 1.5 billion gallons of fuel** and $5 billion in fuel costs per year.
- Researchers from the Centers for Disease Control (CDC) created a far more detailed **model of the hepatitis C virus**, a major cause of liver disease. Annual health care costs associated with this virus are $9 billion in the U.S. alone.
- For the first time researchers developed a computer model that comprehensively **simulates the human heart down to the cell level**. This innovation has significant potential for saving health care costs by reducing heart disease and improving heart health. This research has strong potential for helping to reduce coronary heart disease, which costs the United States over $100 billion each year.

- But that's just part of the story. Without supercomputers, detecting today's sophisticated **cyber security breaches**, insider threats and electronic fraud would be impractical. In short, high performance computing has become indispensable for both maintaining national security and economic competitiveness.

These are some of the powerful reasons that explain why powerful institutions like China, USA, Japan, EU and Russia have created national programs and have invested a lot of money in the development of new supercomputers (exascale projects). That’s why, for a regional scope, similar efforts in projects on HPC (petascale projects) will also results in outstanding results.

Expected Return On Investment (ROI)

(From *IDC's HPC market report* presented at SC15 Conference, Austin, TX, USA; *IDC HPC Update at ISC'14*; *High Performance Computing in the EU: Progress on the Implementation of the European HPC Strategy, SMART number: 2014/0021* and *Implementation of the Action Plan for the European High-Performance Computing strategy. European Cloud Initiative - Building a competitive data and knowledge economy in Europe*).

Revenue/Sales per Euro Invested in HPC

Each euro invested in HPC on average returned an impressive 867,00 € in increased income. For the academic projects, the ROI was a more modest, but still substantial 30,00 € in additional income (i.e., funding for research grants and other sources) per 1,00 € invested.

For industrial projects, the average ROI in revenues generated was a whopping 974,00 € per 1,00 € invested. These ROI figures are higher, but in line with those obtained for U.S. HPC projects (see IDC's 2013 HPC ROI pilot study for the U.S. Department of Energy (DOE).

Even if the average ROI gains were to decline somewhat if a larger sample were taken, it is easy to see from the European (and U.S.) findings why HPC is increasingly recognized as a transformational, game-changing technology by more and more governments around the world.

Revenue/Sales per Euro Invested in HPC, by Market Sector

The most modest revenue ROI gains occurred in the evaluated telecommunications projects—but these still averaged 10,00 € in increased top-line revenue/sales per 1,00 € invested in HPC. The largest average ROI gains per 1,00 € invested in HPC were reported by financial services firms (1590,00 €), transportation firms (1180,00 €), and oil/gas companies (312,00 €),

The major differences in the sizes of average financial returns by industry sector are related, not surprisingly, to the role HPC plays in each sector.

- In the financial services industry, even a slight competitive advantage provided by HPC to back-office "quants" can result in large, nearly immediate ROI.
- Similarly, an HPC advantage in upstream exploration can enable oil and gas companies to find and extract energy sources more efficiently, with large, fairly quick financial returns.
- In manufacturing, by contrast, HPC plays a vital but less stellar role. HPC-related ROI is correspondingly less dramatic.

ROI in the form of bottom-line gains

That is, cost savings or corporate profits related to HPC investments made for projects. Here again, the academic ROI was impressive—averaging 29,00 € per 1,00 € invested in HPC—even though not as large the 75,00 € average ROI figure for the industrial projects

The largest ROI gains in that respect were reported by oil/gas companies (268,00 € per 1,00 € invested in HPC) and financial services firms (67,00 € per 1,00 € invested in HPC). The smallest gains, though still substantial, occurred in the telecommunications companies (10,00 € per 1,00 € invested in HPC) and the insurance firms (7,00 € per 1,00 € invested in HPC).