MSR Technical Computing in Latin America

Dr. Fabrizio Gagliardi
EMEA and LATAM Director
Technical Computing
Microsoft Research
Outline

- TCI mission, a reminder
- Grid for scientific computing: the experience so far
- Issues and new trends:
  - Green and Cloud Computing
- Technical Computing at MSR
  - EMEA and LATAM
- Conclusions
Overall mission: help scientists in doing better science with less efforts

Use MSR technology, tools and MS products to engage MSR scientists in external research with scientists around the world on important problems

www.microsoft.com/science
Highly Skilled Scientist Spending to Much Time Doing Non-scientific Work—Past and Present Approach are Manually Intensive

Integrated Information Management – Contextual, Collaborative and Rich Content

The Goal... More Time For Science

Today
Non-Scientific Activities

Not Enough Science

Tomorrow
More Time on Real Science
The experience of the Grid 1/2

- Grids for e-Science: a success story so far?
  - Several mature Grid Middleware stacks
  - Many HPC applications using the Grid
    - Some (HEP, Bio) in production use
    - Some still in testing phase: more effort still required to make the Grid their day-to-day workhorse
  - e-Health applications also part of the Grid
  - Some industrial applications:
    - CGG Earth Sciences
The experience of the Grid 2/2

- Grids beyond e-Science
  - Slower adoption: prefer different environments, tools and have different TCOs
    - Intra grids, internal dedicated clusters, cloud computing
  - e-Business applications
    - Finance, ERP, SMEs
  - Industrial applications
    - Automotive, Aerospace, Pharmaceutical industry, Telecom
  - e-Government applications
    - Earth Observation, Civil protection:
      - e.g. The Cyclops project
In summary...

- Grid computing has delivered an affordable and high performance computing infrastructure to scientists all over the world to solve intense computing and storage problems within constrained research budget.

- This has also been effectively used by industry to increase the usage of their computing infrastructure and reduce Total Cost of Ownership (TCO).

- Grid is not only aggregating computing resources but also leveraging international research networks to deliver an effective and irreplaceable channel for international collaboration (EELA and EELA II).
The flip side... 

- Major issues with wide adoption of Grid computing in e-Science, e-Business, industry etc. have to do with:
  - Cost of operations and management complexity
  - Not a solution for all problems (latency, fine grain parallelism are difficult)
  - Difficult to use for the average scientist
  - Security and reliability

- Power consumption and heat dissipation are becoming a limiting factor to consumer based distributed systems

- We are observing the limits of Moore’s law
Prof. Satoshi Matsuoka, TITech

- In the late 90s, petaflops were considered very hard and at least 20 years off …
- … while grids were supposed to happen right away
- After 10 years (around now) petaflops are “real close” but there's still no "global grid"
- What happened?
What happened...

- It was easier to put together massive clusters than to get people to agree about how to share their resources.
- For tightly coupled HPC applications, tightly coupled machines are still necessary.
- Grids are inherently suited for loosely coupled apps (e.g., Monte Carlo, Parameter Sweep), or enabling access to machines and data, and the integration of the two.
  - With Gilder's Law, bandwidth to the compute resources will promote thin client approach.
  - Example: Tsubame machine in Tokyo.
New Trends (1/2):
Green Grid and pay per ...CPU/GB

- The Green Grid, IBM Big Green and other IT industry initiatives try to address current HPC limits in energy and environmental impact requirements
- Computer and data centers in energy and environmental favorable locations are becoming important
- Elastic computing, Computing on the Cloud, Data Centers and Service Hosting are becoming the new emerging solutions for HPC applications
- Many-multi-core and CPU accelerators are promising potential breakthroughs
New Trends (2/2):
Cloud computing and storage on demand

Supercomputing at less than the cost of a gallon of gas

New service allows users to offload computational work to supercomputer

Published: May 3, 2007

by Timothy Prickett Morgan

Nearly two months ago, Sun Microsystems announced the launch of Sun Grid compute and storage on demand. Although the service is intended to simplify the job of setting up and running supercomputers, the more important problem that Sun had, however, is that only one company in the world can match its grid capacity.

Starting today, however, the Sun Grid utility, which by default, is now available in 24 additional countries, is now allowing customers to log in and buy compute and storage from Australia, Austria, Belgium, Canada, China, Czech Republic, Denmark, Greece, Hungary, India, Ireland, Italy, Japan, Netherlands, Norway, Poland, Sweden, and the United Kingdom. To help spur adoption of the service, Sun is giving away 200 CPU hours of compute time to any individual who starts an account. This is a first step in what it hopes will be an eventual much larger push.

Late last month, however, Google and server maker and supercomputing giant Cray, working with some prominent American universities to set up a grid that will support those universities in teaching and developing the future of corporate computing some day.

Another Internet-based application, Google perhaps knows more about disk utilization on the fly (as workloads change) or cloud systems imply utility-style infrastructure as well as with the size of the computing resources dedicated to cloud computing, with workloads that may encompass many different kinds of utilities, with varying peaks, and a workload that may use of the computing power only a way to refer to a network of machines connected by cell phones.

For instance, a Linux-204 cluster might be a vector parallel computer might be used to run a single virtual run time and then separate virtualization system might be used in an ad hoc, large-scale parallel systems can be used to search for patterns (especially well), perhaps looking for spam lurking in wdl and forum and in business transactions.
Some examples:

- Major collaboration at the Barcelona Super Computer Centre (Prof. Mateo Valero) on development of S/W environment for support of many-multicore architectures in collaboration with MSR in Cambridge/UK (Spain)

- HPC tools and computing clusters at RAS-IKI for environmental studies and at SPb Polytechnic University for CFD applications (Russia)
Technical Computing in EMEA/LATAM 2/2

- Cancer, climate and financial research at Oxford OeRC (UK)
- Computational Chemistry with Cambridge University (UK)
- Epidemiology research at Manchester (UK)
- HPC for automotive industry at HLRS Stuttgart (DE)
- HPC tools at CoSBi in Trento, Italy (IT)
HPC tools and joint research at the Center for Bioinformatics and Genome Biology of the Foundation CIENCIA PARA LA VIDA in CHILE

Courtesy of David Holmes
Conclusions

- TCI has developed a number of collaborations mostly in computational science and demonstrated how MS and MSR technology can help scientists around the world.

- Good example also in LATAM with FCV in Chile.

- Clearly more needs to be done, event like this one are an excellent occasion to review our strategy, establish contacts and look for new opportunities.
Many thanks for your attention