

COST Full Proposal Reference oc-2010-1-7183

For the implementation of a European Concerted Research Action

Bingos

(Bioinformatics Next Generation Open Software)

Projects by Group

Proposing country Spain
Action proposer: Dr. Oswaldo Trelles
Institution: University of Malaga
Department of Computer Architecture
ETSI Informatica, Campus de Teatinos
Malaga
29071, Spain
Phone: +34 952 13 2823
Fax: +34 952 13 2790
Email: ots @ ac.uma.es, ortrelles @ uma.es

[SP1] University of Málaga, Spain (UMA)

- IBM cluster, 256 nodes JS20-IBM, with 512 processors and 1 Tb distributed memory, each node is 2 CPUs IBM PowerPC single-core 970FX de 64 bits and 2 GHz.
- Cluster superdome architecture HP of shared memory, with 64 nodes and 128 processors and 128 Gb of memory and each node is 2 CPUs Intel Itanium-2 Dual Core a 1,6 GHz.

[SP2] National Centre for Biotechnology (CNB)

The Scientific informatics service develops supporting tasks to the processing and analyzing of scientific data, providing access to:

- Local computational resources
- Wide computational resources in EGEE Grid
- Several data bases of scientific information updated daily.
- Software to analyze Biomedicine data.
- Continuous Bioinformatics training.
- Support to develop new applications and scientific software.
- Advice in the use of the service resources.

[SP4] Barcelona Supercomputing Center (BSC) Biochemistry and Molecular Biology. Institute for Research in Biomedicine. Life Sciences Dept.

SC-CNS hosts MareNostrum, one of the most powerful supercomputers in Europe and the number 60 in the world, according to the last Top500 list. In March 2004 the Spanish government and IBM signed an agreement to build one of the fastest computers in Europe. In November 2006 its capacity has been increased due to the large demand of scientific projects. MareNostrum has increased the calculation capacity of the supercomputer MareNostrum, until reaching 94.21 Teraflops (94.21 trillions of operations per second), doubling its previous capacity (42.35 Teraflops). It had 4.812 processors and has now 10.240 processors with a final calculation capacity of 94.21 Teraflops.

MareNostrum is a supercomputer based on processors PowerPC, the architecture BladeCenter, a Linux system and a Myrinet interconnection. These four technologies configure the base of an architecture and design that will have a big impact in the future of supercomputing.

See below a summary of the system:

- Peak Performance of 94,21 Teraflops
- 10240 IBM Power PC 970MP processors at 2.3 GHz (2560 JS21 blades)
- 20 TB of main memory
- 390 + 90 TB of disk storage
- Interconnection networks:
 - o Myrinet and Gigabit Ethernet
- Linux: SuSe Distribution

Other HPC resources:

SGI Altix 4700: SGI Altix is a shared memory machine, with a cc-NUMA architecture (Cache Coherent Non- Uniform Memory Access). Its hardware configuration is:

- 128 cpus Dual Core Montecito(IA-64). Each one of the 256 cores works at 1,6 GHz
- 8MB L3 cache and 533 MHz Bus.
- 2.5 TB RAM.
- Peak Performance : 819.2 Gflops

- 2 internal SAS disks of 146 GB at 15000 RPMs
- 12 external SAS disks of 300 GB at 10000 RPMs

Nord cluster JS20:

BSC Nord is a cluster with 98 JS20 blades with the following configuration:

- o 2 cpus Powerpc 970 at 2,2GHz.
- o 4 GB RAM per blade.
- o Peak Performance : 1.72 Tflops
- o Gigabit interconnection network.
- o SLES 10 SP1 Operating System.

[SP6] Dept. Comp. Architecture, Universidad Autonoma Barcelona

Cluster IBM:

- 32 Nodos IBM xSerie con doble procesador Intel(R) Xeon(R) Dual-Core.
- Front-End Node
- o IBM x3650
- o Dual-Core Intel(R) Xeon(R) CPU 5150 @2.66GHz 4MB L2 (2x2)
- o 8 GB Fully Buffered DIMM 667 MHz
- o RAID 5 1.8 TB hot-swap SAS
- o Integrated dual Gigabit Ethernet
- o
- 32 IBM x3550 Nodes
- o 2 x Dual-Core Intel(R) Xeon(R) CPU 5160 @ 3.00GHz 4MB L2 (2x2)
- o 12 GB Fully Buffered DIMM 667 MHz
- o Hot-swap SAS Controller 160GB SATA Disk
- o Integrated dual Gigabit Ethernet

Cluster DELL:

- o 5 Nodos Blade Dell PowerEdge M600.
- o 2 x Quad-Core Intel(R) Xeon(R) E5430 @2.66GHz Processors 2x6MB cache L2
- o 12 GB RAM Fully Buffered DIMMs (FBD) 667MHz
- o SAS6ir (H/W based) RAID 1 120GB 2.5" SAS / Plain 2.5" SATA (7.2k rpm): 80GB (Ethernet Pass-Through) Dual Gigabit Ethernet

[CH18] SIB- Swiss Institute of Bioinformatics

SIB manages an important HPC infrastructure from: Vital.IT. Vital-IT is a collaboration between the core SIB member institutions and industrial partners such as Hewlett-Packard and Oracle.

Vital-IT provides the hardware, software and expertise to support the Swiss bioinformatics community. It is equipped with several large Linux clusters for computationally intensive applications, with peak performance in the Teraflops range and disk storage capacity in the hundreds of TB range. The infrastructure also includes several database servers and large shared-memory machines. As services to the community, the facility mirrors a large number of important biological knowledge bases, with powerful cross-indexing capability, and provides a strong expertise in bioinformatics software development and optimization.

The core team consists of about 10 people responsible for system management, software development and software engineering for HPC applications, business development and administration.

[RO24] Bioinformatics & Structural Biochemistry Department, Institute of Biochemistry of the Romanian Academy

PC Linux cluster for parallel computing, Octane 2x600 SGI Workstation.
Human Genome Center, Institute of Medical Science, University of Tokyo

The HGC super computer system was upgraded in January 2003 in response to the explosive growth of genome research data. The system is composed of the HITACHI HA8000 distributed-memory server, and of SunFire 15K, SunFire 6800 and SGI Origin3900T shared-memory servers. In total, 1,428 CPUs, 3.3 Tb RAM and 145 Tb disk space are available.

[IT28] Istituto Nazionale per la Ricerca sul Cancro (INRC)

- 100 PA-Risc CPUs HP Hyperplex cluster (64bits technology),
- 4 CPUs NEC vector computer, 16 GByte RAM and 250 GBytes disk storage,
- 128 nodes Eurotech SpA cluster computer using 256 INTEL XEON at 3,06 GHz. 512 GByte RAM and 11 TByte disk storage
- 3 HP computers Intel Itanium processor based (64 bits).

[IT30] Politecnico di Milano, Italy

- SUN and Intel machines, PCs.
- FPGA boards with the Virtex, Virtex-II PRO, Virtex 4, Virtex 5 and Spartan 3 from Xilinx, FPSLIC from Atmel, Cyclon-II from Altera.
- DSP boards from Atmel, i.e. Diopsis740 and Diopsis940HF
- GPGPU boards from NVIDIA: Tesla C1060, GeForce 8600 GT

[CL37] Centre for Bioinformatics (CBUC); Chile

Among the extensive and state-of-the-art computer-based infrastructure available at the CBUC, is interesting to note:

- 1 Axon GenePix 4100A microarray scanner.
- A 4 CPU's Silicon Graphics Origin 200 64 bits NumaLink cluster with 4 GB of total RAM.
- 3 computer clusters composed of 8, 8 and 12 Pentium IVHT processors with 1GB RAM per node.
- 4 XEON CPU's dedicated database server with 250 GB of hard drive capacity.
- 1 hard drive network server with 8 TB of data storage capacity.
- HP IntegrityServer Supercomputer Cluster with 64 Itanium II Montecito DualCore processors with 128 GB of total RAM equipped with Voltaire Infiniband 10Gbps interconnect.

[CA44] The Sun Center of Excellence for Visual Genomics, University of Calgary

The COE features the following main machines:

- SunFire[tm] 6800 Midframe server with twenty 1250MHz UltraSPARC IV CPUs, four 1.5GHz UltraSPARC IV+ CPUs, and 96GB of main memory (shared).
- SunFire[tm] V880z Midframe server with four 1.2GHz UltraSPARC III+ CPUs, 32GB of main memory, six 72GB internal disk drives, and dual XVR-4000 graphics accelerators.
- Two SunFire[tm] V440 Entry level servers each with four 1.6GHz UltraSPARC III+ CPUs, 16GB of main memory, four 73GB internal disk drives, and two XVR-1200 graphics cards.
- Twelve SunFire[tm] V210 Entry level servers each with two 1.06 UltraSPARC III+ CPUs, 2GB of main memory, and two 36GB internal hard drives.
- Sun StorEdge[tm] 6320 Disk array with a capacity of 12TB.
- Sun StorEdge[tm] L180 Digital Linear Tape (DLT) library with 3TB native capacity (6TB compressed).
- Sun StorEdge[tm] L700 Super Digital Linear Tape (SDLT) library with 35TB native capacity (70TB compressed).
- The operating system currently is mixture of the Solaris[tm] 9 and 10 releases.
- Sun Ray[tm] 1 and 1g Thin Clients as server outlets throughout the laboratory.
- Two Tadpole Comet 12 Mobile SunRay clients.
- Dell PowerEdge 6850 Server with four 3.0GHz Intel(R) Xeon(TM) MP CPUs and Windows 2003 Server, used as a terminal server.

Genome Canada funded machines:

- SunFire[tm] V880 with four 900 MHz CPUs, 8GB of main memory and six 73GB internal hard drives.
- Two SunFire[tm] V890 Midrange servers each with four 1.35GHz UltraSPARC IV CPUs, 16GB of main memory, and six 146GB internal disk drives.
- Two SunFire[tm] V240 Entry level servers each with two 1.5GHz UltraSPARC IIIi CPUs, 2GB of main memory, and two 73GB internal disk drives.
- Twelve SunFire[tm] V210 Entry level servers each with two 1.4 UltraSPARC IV CPUs, 2GB of main memory, and two 36GB internal hard drives.
- Tadpole SPARCbook 6500 with 650 MHz CPU, 2GB of main memory and two 60GB internal hard drives.
- Eight TimeLogic Decypher boards.
- Sun StorEdge[tm] 6130 Serial Advanced Technology Attachment (SATA) disk array with a capacity of 40TB.
- Sun StorEdge[tm] L700 SDLT library with 107TB native capacity (215TB compressed).