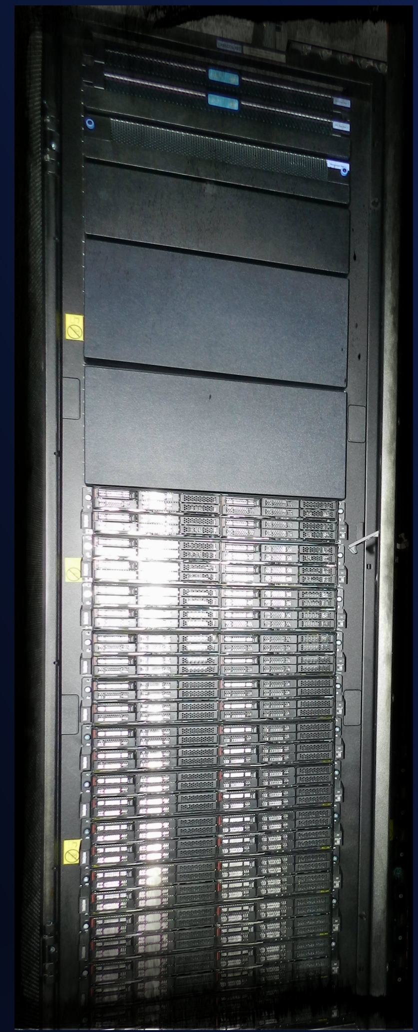
CRESCO6: tech specs & benchmarks

ENEA DTE-ICT-HPC: F. Ambrosino, G. Aprea, T. Bastianelli, I. Bellagamba, R. Bertini, G. Bracco, L. Bucci, F. Buonocore, M. Caporicci, M. Caiazzo, B. Calosso, M. Celino, M. Chinnici, A. Colavincenzo, A. Cucurullo, P. D'Angelo, D. De Chiara, M. De Rosa, D. Di Mattia, S. Ferriani, G. Ferro, C. Ferrelli, A. Funel, D. Giammattei, M. Galli, S. Giusepponi, G. Glorioso, R. Guadagni, G. Guarnieri, M. Gusso, F. Iannone, M. Marano, A. Mariano, G. Mencuccini, S. Migliori, M. Mongelli, P. Ornelli, S. Pagnutti, F. Palombi, S. Pecoraro, A. Perozziello, S. Pierattini, S. Podda, G. Ponti, A. Quintiliani, G. Santomauro, A. Scalise, F. Simoni, D. Visparelli

INTRODUCTION

In 2015 ENEA and CINECA signed a strategic partnership agreement aimed at providing supercomputing and data storage services to EUROfusion, the European consortium for the Development of Fusion Energy. CINECA is currently the main supplier of HPC services in Italy. The agreement signed with ENEA aimed at promoting the joint devlopment of research activities in the field of HPC and was the basis for the purchase and installation of CRESCO6, a TIER1 cluster at the Portici ENEA Research Centre.

CRESCO6, a machine with a nominal power of about 0.7 Pflop/s* succeeds CRESCO4 and CRESCO5, two older supercomputers already installed at the same centre and still operating, with a nominal computing power of 0.1 and 0.025 Pflop/s respectively. As such, CRESCO6 alone increases the total computing power currently available for research at ENEA by a factor of 7.



TECH SPECS

CRESCO6 is a high performance computing system (HPC) consisting of 216 nodes for a total of 10,368 cores. Each node is equipped with:

- **Two Intel Xeon Platinum 8160 CPUs, each with 24 cores and operating with a clock frequency of 2.1 GHz**
- **A RAM of 192 GB, corresponding to 4 GB/core**
- **A low-latency Intel Omni-Path 100 Series Single-port PCIe 3.0 x16 HFA network interface.**

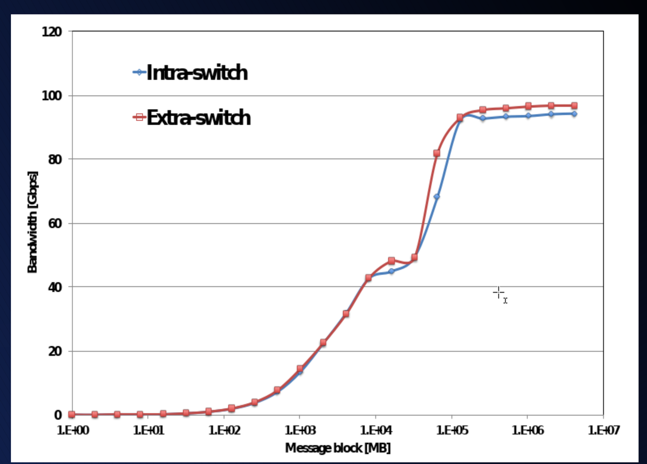
The nodes are interconnected by an Intel Omni-Path network with 15 switches of 48 ports each, bandwidth equal to 100 GB/s, latency equal to 100ns. The connections between the nodes have 2 tier 1:1 no-blocking fat-tree topology.

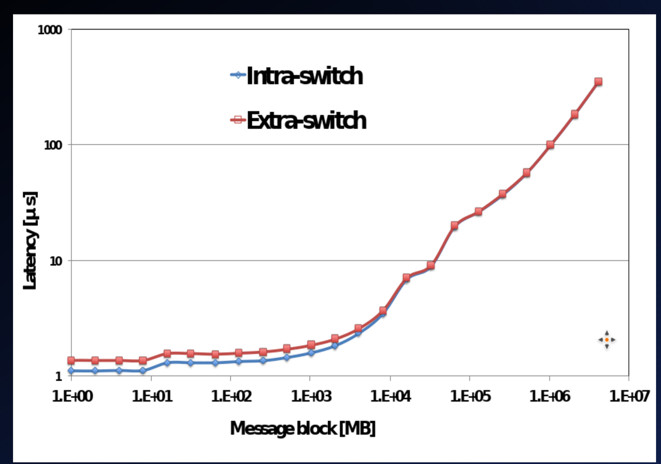
The consumption of electrical power measured during the HPL tests amounts to 95 kW.

OSU BENCHMARKS ON BANDWIDTH & LATENCY

Bandwidth tests were carried out by having the sender sending out a fixed number (equal to the window size) of back-to-back messages to the receiver and then waiting for a reply from the receiver. The receiver sends the reply only after receiving all these messages. This process is repeated for several iterations and the bandwidth is then calculated.

Latency tests are carried out in a ping-pong fashion. The sender sends a message with a certain data size to the receiver and waits for a reply from the receiver. The receiver receives the message from the sender and sends back a reply with the same data size. Many iterations of this ping-pong test are carried out and average one-way latency numbers are thus obtained.





CRESCO6 @ ENEA - PORTICI




HPL BENCHMARK

HPL (High Performance Linkpack) solves a (random) dense linear system in double precision (64 bits) arithmetic on distributed memory computers.

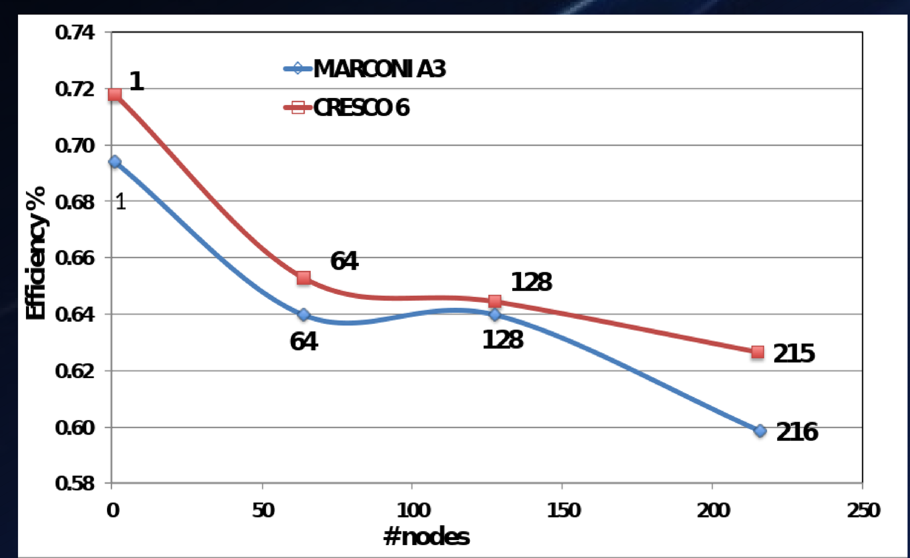
The algorithm used by HPL can be summarized by the following keywords: Two-dimensional block-cyclic data distribution - Right-looking variant of the LU factorization with row partial pivoting featuring multiple look-ahead depths - Recursive panel factorization with pivot search and column broadcast combined - Various virtual panel broadcast topologies - bandwidth reducing swap-broadcast algorithm - backward substitution with look-ahead of depth 1.

The HPL was run on CRESCO6 and the MARCONI A3 systems. The MARCONI A3 system is the Tier-0 system, co-designed by CINECA. A Marconi A3 node is based on dual socket 24-cores Intel Xeon 8160 (SkyLake) at 2.10 GHz with Intel OmniPath.

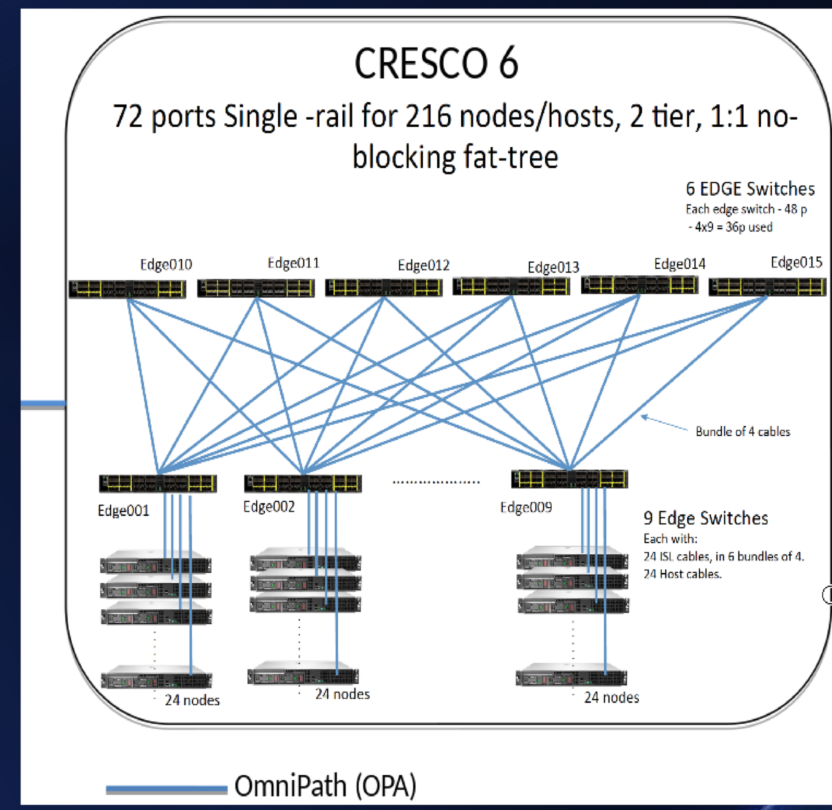
The graph below summarizes the performance as run on 1, 64, 128 and 216 nodes. The benchmark shows good performance around 63% efficiency and up to 434 Tflop/s for the CRESCO6 system on 216 nodes.

The benchmarks were made with the Intel/IntelMPI compiler v.18 and MKL library, 2 MPI processes per node (one for each socket). The CPU clock frequency was set on max performance. N.B.: the benchmarks on Marconi A3 were performed while the machine was in production mode (jobs were running on the other nodes).

SKL Intel Xeon 8160 @ 2.1 GHz 24 cores 440 W/node MARCONI A3			
# nodes	Rmax Gflop/s	Rpeak Gflop/s	Eff %
1	2237.8	3225.6	0.694
64	131983.3	206438.4	0.639
128	264078.0	412876.8	0.640
216	416911.0	696729.6	0.598
SKL 2 x Intel Xeon 8160 @ 2.1 GHz 24 cores 440 W/node CRESCO6			
# nodes	Rmax Gflop/s	Rpeak Gflop/s	Eff %
1	2315.2	3225.6	0.718
64	134710.0	206438.4	0.653
128	265980.0	412876.8	0.644
216	434418.0	693504.0	0.626



CRESCO6 INTEL OMNIPATH NETWORK



N-BODY GRAVITATIONAL BENCHMARK

The aim of this case study is to compare the performances of MARCONI A1 (BDW), MARCONI A2 (KNL), MARCONI A3 (SKL) and CRESCO6 (SKL) using a classic *n*-body simulation.

A Particle-Particle (PP) method based on a brute-force approach [1] is used to simulate the *n*-body problem, where the force acting on the *i*-th particle is the result of its interactions with all other particles.

A case study of the *n*-body simulation is used to compare an hybrid MPI+OpenMP algorithm running on different HPC systems in operation at CINECA and on the new ENEA HPC system CRESCO6.

An hybrid master-only model is used with one MPI rank per node and OpenMP threads scattered on the cores of the node, with no MPI calls inside MP parallel regions.

As the *n*-body PP method has a complexity of O(n^2), tests were carried out getting the execution time for one time-step adopting two cases for the number of particles: 10^6, 10^7.

[1] S. Oikawa, "A force evaluation free method to N-body problems: Binary interaction approximation". Communications in Nonlinear Science and Numerical Simulation (2016), Vol. 32, pp. 273-284.

Newton's Law $m_i\ddot{\vec{r}}_i = \vec{F}_i$

Numerical Integration

$$\dot{\vec{r}}_i(t+\Delta t) = \dot{\vec{r}}_i(t) + \Delta t \cdot \frac{\vec{F}_i}{m_i}$$

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \Delta t \cdot \dot{\vec{r}}_i(t)$$

Gravitational Force

$$\vec{F}_i = \sum_{j\neq i} \frac{G m_i m_j (\vec{r}_j - \vec{r}_i)}{\left|\vec{r}_j - \vec{r}_i\right|^3}$$

	BDW Intel Xeon E5-2697 v4 @ 2.3 GHz 18 cores 475 W/node MARCONI A1			KNL 1 x Intel Phi7250 @ 1.4 GHz 68 cores 360W/node MARCONI A2		
n-body	08 nodes (sec)	16 nodes (sec)	32 nodes (sec)	08 nodes (sec)	16 nodes (sec)	32 nodes (sec)
10^6	42.81	21.58	10.98	110.34	57.20	30.19
10^7	4296	2145	1072	10805	5405	2705

	SKL Intel Xeon 8160 @ 2.1 GHz 24 cores 440 W/node MARCONI A3			SKL 2 x Intel Xeon 8160 @ 2.1 GHz 24 cores 440 W/node CRESCO6				
n-body	08 nodes (sec)	16 nodes (sec)	32 nodes (sec)	08 nodes (sec)	16 nodes (sec)	32 nodes (sec)	64 nodes (sec)	128 nodes (sec)
10^6	22.95	11.72	6.03	23.85	12.5	6.62	3.3	1.93
10^7	2478	1208	617	2361	1181	590	296	150

Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile