Fusion Engineering and Design xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Fusion Engineering and Design

journal homepage: www.elsevier.com/locate/fusengdes

MARCONI-FUSION: The new high performance computing facility for European nuclear fusion modelling

F. Iannone[a,*], G. Bracco[a], C. Cavazzoni[b], R. Coelho[c], D. Coster[d], O. Hoenen[d], A. Maslennikov[b], S. Migliori[a], M. Owsiak[e], A. Quintiliani[a], B. Palak[e], V. Pais[f], F. Robin[g], E. Rossi[b], I. Voitsekhovitch[h]

[a] *ENEA, via Lungotevere Thaon di Ravel 76, Rome 00196, Italy*
[b] *CINECA, via Magnanelli 6/3, Casalecchio di Reno, Bologna 40033, Italy*
[c] *Instituto de Plasmas e Fusão Nuclear, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal*
[d] *Max-Planck-Institut für Plasmaphysik, Boltzmannstraße 2, 85748 Garching bei München, Germany*
[e] *Poznan Supercomputing and Networking Center, 61-704 Poznan, Noskowskiego 10, Poland*
[f] *National Institute for Laser, Plasma and Radiation Physics, Str. Atomistilor, Nr. 409, Bucharest, Romania*
[g] *CEA, Centre de Saclay, 91191 Gif-sur-Yvette, France*
[h] *EUROFusion, Boltzmannstraße 2, 85748 Garching bei München, Germany*

ARTICLE INFO

Keywords:
High performance computing

ABSTRACT

The new European HPC facility for Fusion is in operation since July 2016. It replaces, for European fusion researchers, the Helios supercomputer installed in Japan in the context of the Broader Approach agreement. The supercomputer is hosted at CINECA and it is a fraction of the MARCONI system. Thanks to a customized technical project done by ENEA, in a joint development agreement with CINECA, the European community of fusion modelling can exploit the latest available CPU technologies, following the CINECA HPC roadmap towards 50 PFlops planned for 2019. The MARCONI Fusion fraction is being delivered in two phases: the first one, 1 PFlops of CPU multi-core architecture based on the Intel Broadwell processors, is already in operation since July 2016, and the second one, 5 PFlops of the same architecture based on the INTEL Skylake processors, will be deployed in July 2017. Furthermore the project includes 1 PFlops of the third generation of Intel Xeon Phi many-core architecture (Knights Landing generation).

Within this framework, ENEA/CINECA provides, in addition, the operation support of the Gateway infrastructure of EUROFusion Work-Package Code Development. A new Gateway HPC system is in operation at CINECA since Jan. 2017 thanks to the data migration and software porting activities carried out by ENEA/CINECA team together with the Core Programming Team of the Infrastructure and Support Activity work package from EUROfusion. The new Gateway infrastructure is tightly coupled with the MARCONI Fusion fraction, sharing the same 100 Gbps low-latency network based on the Intel OmniPath technology.

The paper describes the technical details and the performances of MARCONI, one of the largest HPC OmniPath based infrastructure.

1. Introduction

The European Fusion Research roadmap [1] provides a long-term perspective to enable electricity production by mean of nuclear fusion well before 2050. ITER is the key facility in the roadmap and it is already on-going placing the EU in a leading position.

Theory and the modelling effort in plasma and materials relevant for nuclear fusion is a crucial key in the roadmap as capability to validate experimental physics results produced by ITER and to enable the design of the next fusion power plant DEMO.

Considerable progress has been made with detailed modelling and the simulation of plasma evolution in tokamak configurations, control, stability and its impact on materials is now at high level of exploitation in the fusion specialists community, developing integrated models able to simulate the whole scenario of a tokamak experiment.

Special provisions for High Performance Computing (HPC) are being recommended in the fusion roadmap as essential facilities to support basic research and the modelling effort for the various

* Corresponding author.
E-mail address: francesco.iannone@enea.it (F. Iannone).

https://doi.org/10.1016/j.fusengdes.2017.11.004
Received 30 May 2017; Received in revised form 3 November 2017; Accepted 3 November 2017
0920-3796/

objectives.

The European Fusion community has used dedicated HPC facilities since 2009, with HPC-FF [2]: a multiprocessor Intel® Xeon Nehalem EP™ cluster of 100 TFlops of peak power and Infiniband QDR low latency network, in operation until 2013 at Jülich Supercomputing Centre, followed by Helios: a supercomputer with a performance of 2 PFlops consisting in an Intel Xeon Sandy-bridge EP™ partition of 1.6 PFlops and in a Many Integrated Core (MIC) Intel Xeon Phi (KNC generation) partition of 0.4 PFlops with the same low latency network, in operation until the end of 2016 at IFERC-CSC in Rokkasho, Japan [3]. Both HPC facilities have been exploited by means specific allocation time project calls relevant in plasma and materials modelling and nuclear fusion technology simulations.

Furthermore the High Level Support Team (HLST), the developers team based in IPP Garching, has provided to optimize codes in order to improve the speedup on the new CPU technologies and low latency network as well, including also the new MIC architectures Intel Xeon Phi™.

Following the increasing need in the issues related to plasma physic simulations (turbulence, MHD, edge physics and integrated modelling) together with an important number of computing projects addressing technology issues, a growing number of fusion modelling developers are optimizing with success their own codes making able for scaling up to a large number of cores. Therefore the EU fusion community are extending its technical computational capabilities by establishing a new HPC facility for Fusion Applications under EUROfusion. The facility, named MARCONI-FUSION is a dedicated part of a larger HPC system hosted at CINECA [4] (Bologna) under an EUROfusion Project Implementing Agreement (PIA) with ENEA/CINECA. It consists in two parts: a conventional processor part (based on Intel Xeon processors) and a many-cores processor part (based on Intel Xeon Phi processors).

Thanks to CINECA HPC Development Roadmap of the HPC infrastructure for the period 2016–2020, the EU fusion community could deploy HPC resources based on the newest technology generation of processors.

Furthermore as consequence of the Gateway shutdown at IPP Garching, additional resources are allocated to the new Gateway closely connected to MARCONI-FUSION hosting the IM (Integrated Modelling) tool developed by EUROfusion. For this purpose, a cluster of 24 compute nodes, has been configured in a flexible environment allowing graphical interactive remote sessions and batch jobs submissions as well. Thanks to a common hardware/software infrastructure, the Gateway users can exploit the HPC computing resources of MARCONI.

The paper describe hardware/software configuration of the CINECA MARCONI HPC facility, the Fusion partition and the new Gateway as well. It includes performance benchmarks, issues in the new high performance interconnectivity network for configurations with a large number of compute nodes and performance indexes during the first phase of operations.

2. Marconi

The CINECA is currently one of the Large Scale Facilities in Europe and it is a PRACE [5] hosting site. Aimed at supporting scientific research in HPC, CINECA approved a development roadmap in two provisioning phases, from 2016 to 2020. The first provisioning phase in 2016–2017 is being provided a computing peak of 18 PFlops, named MARCONI. The complete MARCONI deployment is being expected on Jul.2017 and it is based on the following three steps:

- A1: 2 PFlops Intel Xeon conventional “Broadwell (BDW)” based, in full production on Jul. 2016
- A2: 11 PFlops Intel Xeon Phi “Knigths Landing (KNL)” based, in full production on Jan.2017
- A3: 5+ PFlops Intel Xeon “Skylake (SKL)” based, it is being expected on Jul. 2017.

Table 1
MARCONI A1/FUSION main characteristics.

	MARCONI A1	MARCONI-FUSION
Model:	Lenovo **NextScale**	
Architecture:	**Intel OmniPath Cluster**	
Nodes:	**1512**	**806**
Processors:	2 × 18 cores Intel Xeon E5-2697 (BDW) at 2.3 GHz	
Cores:	36 cores/node – Total = 54432	Total = **29016**
RAM:	**128 GB/node**, 3.5 GB/core	
Racks:	**21** + **1** (mgmt.shared)	~**11** + **1** (mgmt. shared)
Peak Perfomance:	= 2 cpu × 18 cores × 2.3 Hz × 16 fpo × 1512nodes = **2.003 PFlops**	= 2 cpu × 18 cores × 2.3 Hz × 16 fpo × 806nodes = **1.068 PFlops**
Disk Space:	**17 PB** (raw)	**5 PB** (raw)
Power:	**700 kW**	**350 kW**

In total MARCONI is 18 PFlops peak performance and 17 PB raw of high performance storage.

2.1. Marconi A1

MARCONI *A1* is the first part of the new HPC system, co-designed by CINECA and based on the LENOVO NeXtScale platform. It takes advantage of the new Intel® Omni-Path Architecture (OPA), which provides the high performance interconnectivity (100 Gb/s) required to efficiently scale the system's thousands of servers. A high-performance Lenovo GSS storage subsystem, that integrates the IBM Spectrum Scale™ (GPFS) file system, is connected to the Intel Omni-Path Fabric and provides data storage capacity. This system is air-cooled.

A dedicated fraction of MARCONI *A1*, (named MARCONI-FUSION) is reserved for the EUROfusion HPC activities, next replaced by 5 PFlops of MARCONI *A3* shall be in production until Dec.2018, end of EUROFusion HPC Phase 1 commitment. The Table 1 shows the main characteristics of the MARCONI *A1* and the dedicated fraction of MARCONI-FUSION as well.

The compute node of MARCONI *A1* is based on server LENOVO NeXtScale nx360M5. It's a ½U two-socket server supporting Intel Xeon V4 up to 22 cores per processor, DDR4 memory at up 2400 MHz and DIMM capacities up to 64 GB. The NeXtScale architecture provides a dense compute nodes assembly thanks to the n1200 chassis that can host up to 12 nx360M5 nodes in a 6U enclosure.

The MARCONI *A1* compute node configuration is the following:

- 2 × CPU Intel Xeon E5-2697 v4/18 core at 2.3 GHz
- 8 × 16GB DIMM of RAM DDR4 2400 MHz
- 1 × 120 GB SATA MLC S3500 Enterprise Value SSD
- 1 x link OPA (Omni-Path) 100Gb/s, 2 link 1GbE

The 1512 compute nodes nx360M5 are packaged in 126 nx1200 chassis assembled in 21 racks 42U.

The Front-End system of MARCONI *A1*, in sharing with the FUSION fraction, is composed of 8 login nodes LENOVO x3550M5 1U (rackmount). The Management System is composed of: *i*) the Management server of MARCONI *A1*, in sharing with the FUSION fraction, is composed of 6 managment servers LENOVO x3550M5 1U (rackmount) configured as above except for cpu having an INTEL Xeon E5-2600 v3 (Haswell); *ii*) the Management network is composed of 48 LENOVO RackSwitches mod. G8052: 48 ports 1 GbE, 4 x uplink 10 GbE. The layout is 2 RackSwitch for each rack. The RackSwitch G8052 are linked to 2 LENOVO RackSwitch mod G8296 main switch: layer 3, 86 port 10 GbE and 10 ports 40 GbE.

The low-latency interconnection network is based on the new Intel Omni-Path fabric based on the product line Intel OPA 100 series running at 100 Gbps, the same than Infiniband Enhanced Data Rate (EDR),.