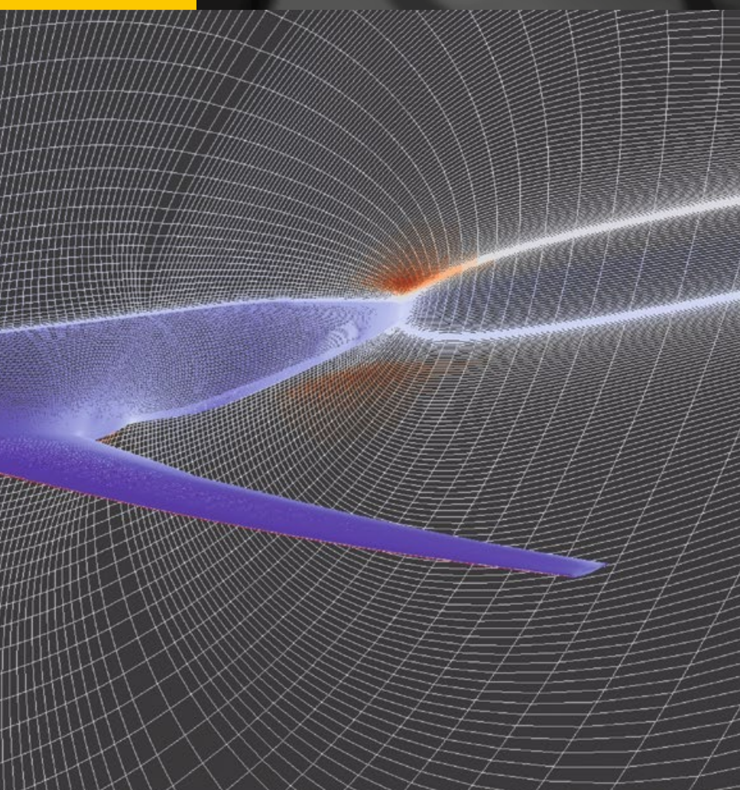
Michael M. Resch · Johannes Gebert · Hiroaki Kobayashi
Wolfgang Bez *Editors*

Sustained Simulation Performance 2021



HLRS

Springer

Sustained Simulation Performance 2021

Michael M. Resch • Johannes Gebert
Hiroaki Kobayashi • Wolfgang Bez
Editors

Sustained Simulation Performance 2021

Proceedings of the Joint Workshop on Sustained Simulation Performance, University of Stuttgart (HLRS) and Tohoku University, 2021

Springer

Editors
Michael M. Resch
High-Performance Computing Center
University of Stuttgart, HLRS
Stuttgart, Germany

Johannes Gebert
High-Performance Computing Center
University of Stuttgart
Stuttgart, Germany

Hiroaki Kobayashi
Graduate School of Information Sciences
Tohoku University
Aoba-ku, Japan

Wolfgang Bez
NEC High Performance Computing
Europe GmbH
Düsseldorf, Germany

ISBN 978-3-031-18045-3 ISBN 978-3-031-18046-0 (eBook)
https://doi.org/10.1007/978-3-031-18046-0

Mathematics Subject Classification (2020): 65-XX, 65Exx, 65Fxx, 65Kxx, 68-XX, 68Mxx, 68Uxx, 68Wxx, 70-XX, 70Fxx, 70Gxx, 76-XX, 76Fxx, 76Mxx, 92-XX, 92Cxx

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The Workshop on Sustained Simulation Performance was held online at HLRS in March 2021 and in a hybrid mode at the Cyberscience Center, Tohoku University in December 2021. The collaboration between the High-Performance Computing Center Stuttgart, Tohoku University and NEC has been marked by the Covid pandemic, in which we demonstrated our ability to adapt to new situations and continue our partnership. Ultimately, we are happy to continue the relationship that began in 2004 with the establishment of what we called the 'Teraflop Workshop'. While the homepage still remembers this name, the workshop evolved into the Workshop on Sustained Simulation Performance with more than 30 events on two continents.

Perhaps we were able to adapt so quickly to the pandemic because the field of high-performance computing has always evolved rapidly. While HPC systems were designed for many years as single processor vector machines, they now are large cluster systems with fast interconnects and rather typically with a combination of a variety of processors and accelerators – among them still vector processors. Climate and weather simulation is one of the scientific fields that has a particularly high demand for computing power, and research has shown that we want to use our resources more sustainably. This is at odds with the ever larger systems with ever higher energy consumption of modern HPC systems. At the same time, however, there has been a tremendous increase in efficiency. The contributions of this book and the upcoming workshops will help to continue and accelerate the development of fast and efficient high-performance computing.

We would like to thank all the contributors and organizers of this book and the Sustained Simulation Performance Workshops. We especially thank Prof. Hiroaki Kobayashi for his close collaboration over the past years and look forward to intensifying our cooperation in the future.

Stuttgart, Germany — *Michael M. Resch*
December 2021 — *Johannes Gebert*

Contents

Supercomputer for Quest to Unsolved Interdisciplinary Datascience (SQUID) and its Five Challenges

Susumu Date, Yoshiyuki Kido, Yuki Katsuura, Yuki Teramae and Shinichiro Kigoshi

Abstract The Cybermedia Center at Osaka University started the operation of a supercomputing system named Supercomputer for Quest to Unsolved Interdisciplinary Datascience (SQUID) in May 2021. SQUID is a hybrid supercomputing system composed of three kinds of heterogeneous compute nodes and delivers 16.591 PFlops as the theoretical performance. This paper overviews the architecture and structure of SQUID and then explains the five challenges which we have set in designing SQUID: Tailor-made computing, HPC and HPDA integration, Cloud-interlinked and -synergized, Secure computing environment, and Data aggregation environment. After that, the future issues to be tackled through the actual operation of SQUID are described.

1 Introduction

Recently, the globalization of academic research has been accelerating. It requires the aggregation and integration of computing, data and even human resources. Accompanied with the globalization of academic research, it would become more common and general that researchers and scientists who are with different organizations work together as a team for solving a common scientific problem [4]. This trend is not exceptional in Osaka University but observed worldwide. For the higher productivity in globalized academic research, the information and communication technologies (ICT) would take a role of greater importance. For the reason, the Cybermedia Center at Osaka University (CMC) which is a supercomputing center and in charge of the administration and management of ICT infrastructures including supercomputing

Susumu Date and Yoshiyuki Kido,
Cybermedia Center, Osaka University, Japan, e-mail: date@cmc.osaka-u.ac.jp

Yuki Katsuura, Yuki Teramae and Shinichiro Kigoshi
Department of Information and Communication Technology Services, Osaka University, Japan

M. M. Resch et al. (eds.), *Sustained Simulation Performance 2021*,
https://doi.org/10.1007/978-3-031-18046-0_1

system for research and education [2], is expected to implement the supercomputing systems well prepared for the rapid expansion and globalization of academic researches.

Furthermore, high performance data analysis (HPDA) has been increasing its importance. Today, many researchers and scientists are enthusiastic about applying data analysis techniques, characterized with keywords such as artificial intelligence (AI), machine learning (ML) and deep learning (DL), to a large amount of data set to solve their scientific problems. Such enthusiasm, expectation and concern to HPDA have triggered the utilization of supercomputing systems by researchers who have never used any supercomputing system so far. As a result, it is expected that newly developed supercomputing systems should accommodate the new computing needs and requirements derived from HPDA as well as traditional high performance computing (HPC).

In the background above, the CMC has developed and installed a new supercomputing system named Supercomputer for Quest to Unsolved Interdisciplinary Datascience (SQUID) [13] in May 2021, in a hope that the new supercomputing system facilitates researchers and scientists who work on researches for the advancement of academia and industries to explore unsolved data scientific problems. For realizing SQUID, we have set the five challenges toward our envisaged next-generation supercomputing systems. In this paper, we briefly introduce SQUID and then explain the five challenges.

This paper is structured as follows. Section 2 briefly introduces the hardware configuration of SQUID. In Section 3 the five challenges set in realizing SQUID are explained. After that, Section 4 describes the issues to be tackled. Section 5 summarizes this paper.

2 Hardware configuration of SQUID

Figure 1 shows the exterior view of SQUID installed at the CMC. This SQUID is a hybrid supercomputing system composed of three different architectures; general-purpose CPU, GPU and vector nodes. All of processors and accelerators deployed on the compute nodes of SQUID are cooled with DLC (direct liquid cooling) for stable operation and high performance delivery purpose. For the parallel filesystem, Lustre-based DDN EXAScaler was adopted to provide users with a single and fast disk image of 20 PB HDD and 1.2 PB NVMe SSD. Mellanox InfiniBand HDR (200 Gbps) was adopted to connect all of three types of compute nodes and the Lustre parallel filesystem (Fig. 2). As the topology, the combinational use of the Dragonfly+ and Fat-tree was adopted. As to the Dragonfly+ topology for CPU nodes, 1520 CPU nodes are divided to three groups (513 nodes, 513 nodes and 494 nodes) and a group is connected to each of other two groups with 95 IB HDR links (19 Tbps). The CPU nodes in each group are connected as the Fat-tree topology to take advantage of full-bisectional bandwidth. On the other hand, the GPU nodes, the vector nodes, the file servers for Lustre filesystem, and other management servers for SQUID are

Fig. 1: Exterior view of SQUID.

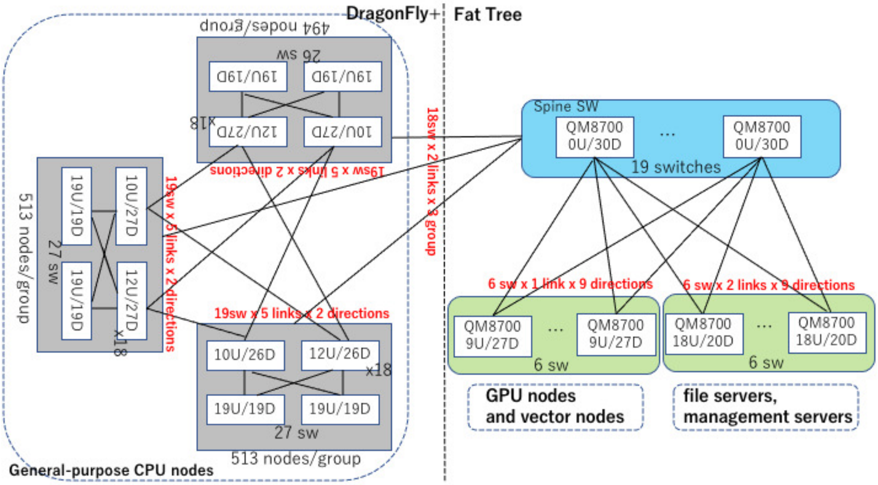


Fig. 2: Overview of the interconnect on SQUID.

connected as the Fat-tree topology to utilize full-bisectional bandwidth. The spine switches of the Fat-tree interconnect for the GPU node, the vector nodes, the file servers and other management servers are connected to each group of the CPU nodes with 36 IB HDR links (7.2 Tbps).

Table 1 shows the system performance and configuration of SQUID. The theoretical performance of SQUID reaches 16.591 PFlops. The major portion of SQUID, as the table indicates, is the cluster of general-purpose CPU nodes. SQUID has 1520

Table 1: System performance and configuration of SQUID.

compute node (16.591 PFlops)	general-purpose CPU nodes 1,520 nodes (8.871 PFlops)	CPU: Intel Xeon Platinum 8368 (Ice Lake / 2.4 GHz 38C) x 2 Memory : 256 GB
	GPU nodes 42 nodes (6.797 PFlops)	CPU: Intel Xeon Platinum 8368 (Ice Lake / 2.4 GHz 38C) x 2 Memory : 512 GB GPU: NVIDIA HGX A100 8-GPU board (Delta)
	vector nodes 36 nodes (0.922 PFlops)	CPU: AMD EPYC 7402P (ROME / 2.8 GHz 24C) x 1 Memory : 128 GB vector: NEC SX-Aurora TSUBASA Type20A x 8
storage	DDN EXAScaler(Lustre)	HDD: 20.0 PB NVMe 1.2 PB
interconnect	Mellanox InfiniBand HDR (200 Gbps)	
front-end node	front-end node for HPC 4 nodes	CPU: Intel Xeon Platinum 8368 (Ice Lake / 2.4 GHz 38C) x 2 Memory : 256 GB
	front-end node for HPDA 4 nodes	CPU: Intel Xeon Platinum 8368 (Ice Lake / 2.4 GHz 38C) x 2 Memory : 512 GB
	secure front-end node 1 node	CPU: Intel Xeon Platinum 8368 (Ice Lake / 2.4 GHz 38C) x 2 Memory : 256 GB

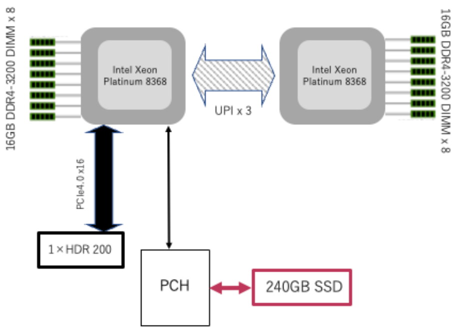


Fig. 3: The internal architecture of a SQUID CPU node.

CPU nodes in total. Figure 3 shows the block diagram of the CPU node. Each CPU node has 2 Intel Xeon Platinum 8368 (Ice Lake/ 2.4 GHz, 38 Core) processors and 256 GB memory deployed. The two processors are connected on 3 UPI (Ultra Path Interconnect) links and 8 channels of 16 GB DDR4-3200 DIMMs are connected to