

JUSUF: Modular Tier-2 Supercomputing and Cloud Infrastructure at Jülich Supercomputing Centre

Forschungszentrum Jülich, Jülich Supercomputing Centre *

Instrument Scientists:

- Supercomputing Support, Jülich Supercomputing Centre, Forschungszentrum Jülich, phone: +49(0)2461 61 2828, sc@fz-juelich.de

Abstract: JUSUF is a petaflop supercomputer operated by Jülich Supercomputing Centre at Forschungszentrum Jülich as a European supercomputing and cloud resource. JUSUF was funded via the ICEI project (Interactive Computing E-Infrastructure for the Human Brain Project, 2021) and especially serves the Human Brain Project (Human Brain Project, 2020) and PRACE (Partnership for Advanced Computing in Europe, 2019) via ICEI and the Fenix Research Infrastructure (Fenix RI, 2019). The system consists of two parts, an HPC cluster partition and an Infrastructure-as-a-Service cloud partition. The system entered production phase in spring 2020. It is based on the Bull X400 product family with AMD Rome processors, partially accelerated by Nvidia V100 GPUs, and Nvidia Mellanox HDR InfiniBand.

1 Introduction

Jülich Supercomputing Centre (JSC) (Forschungszentrum Jülich, 2019b) at the Forschungszentrum Jülich (Forschungszentrum Jülich, 2019a) has been operating some of the largest supercomputers in Europe for three decades and is part of the larger Gauss Centre for Supercomputing (Gauss Centre for Supercomputing, 2019), offering three Tier-0 supercomputers to members of the European research community. As part of its involvements in many European research projects, JSC hosts and operates the Tier-2 supercomputer JUSUF (Jülich Support for Fenix), which was partially funded by the ICEI project and offers HPC and cloud resources to users via Fenix and PRACE (Partnership for Advanced Computing in Europe) calls. It entered production in spring 2020.

*Cite article as: Jülich Supercomputing Centre. (2021). JUSUF: Modular Tier-2 Supercomputing and Cloud Infrastructure at Jülich Supercomputing Centre. *Journal of large-scale research facilities*, 7, A179. <http://dx.doi.org/10.17815/jlsrf-7-179>

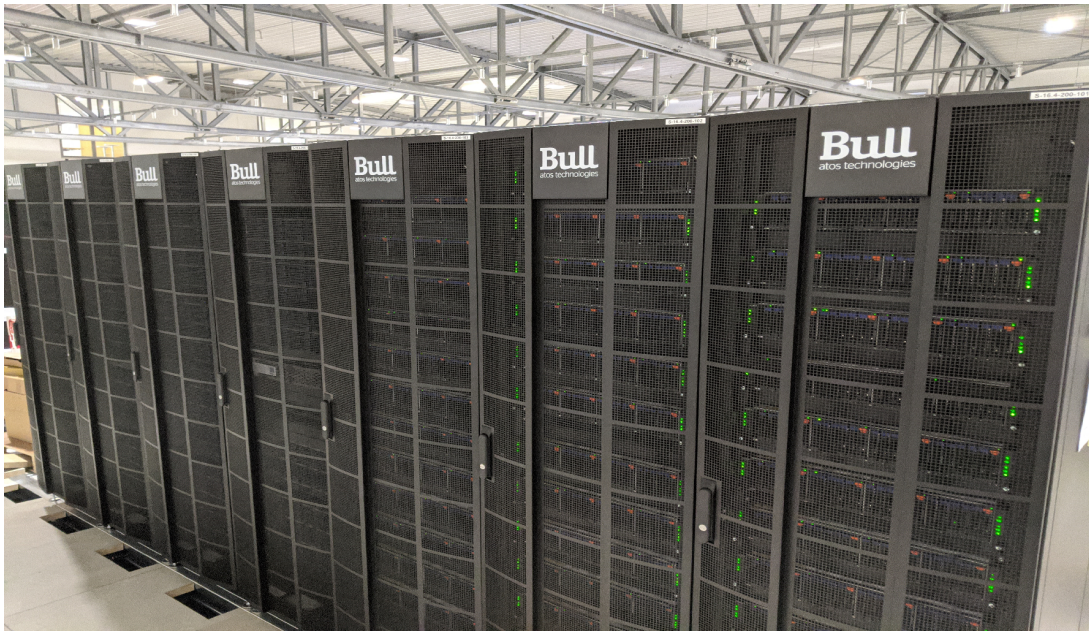


Figure 1: The supercomputer JUSUF in the facility of Jülich Supercomputing Centre. Copyright: Forschungszentrum Jülich GmbH.

1.1 ICEI and the Fenix RI

The ICEI project (Interactive Computing E-Infrastructure for the Human Brain Project) is an initiative of five supercomputing centres, namely BSC (Spain), CEA (France), CINECA (Italy), CSCS (Switzerland) and JSC (Germany), to support in particular but not only the neuroscience community via providing scalable compute and data services to the HBP (Human Brain Project). To achieve this goal, several clusters and services have been procured and brought into production. To guarantee a sustainable service offer, the participating sites and additional interested European supercomputing facilities implemented the Fenix Research Infrastructure (Fenix RI).

2 JUSUF System Details

The JUSUF system implements a hybrid approach for providing resources. It has a cluster module, providing HPC/batch system like resources to user communities for scalable and interactive computing. In addition, the system was designed to provide a cloud module for serving Infrastructure-as-a-Service demands. These two partitions use the same set of 205 compute nodes, based on Atos Bull X400 servers. All compute nodes are equipped with two AMD EPYC ROME 7742 64-Core processors, 256 GB of main memory and an 800GB NVMe. Out of 205 nodes, there are 144 CPU-only nodes (Sequana X440A Double Twin servers) and 61 GPU nodes (Sequana X430A servers) that are equipped with one Nvidia Volta V100 GPU each. All nodes are equipped with dual-port HDR100 Nvidia Mellanox Connect-X6 HCAs.

2.1 Network Design and Connections

Cluster and cloud modules share the same hardware configurations but are separated on a network level.

On the cluster side, the nodes are connected to a high-speed InfiniBand fabric with FatTree topology via one of the Connect-X6 ports. Another system, procured in the ICEI context, the High Performance Storage Tier at JSC (T, 2020), is available via InfiniBand as well. The second Connect-X6 port uses 40

Gigabit Ethernet for connection to the central storage system JUST and with this access to filesystems like HOME and SCRATCH.

On the cloud side, InfiniBand is not available. One Connect-X6 port uses 40 Gigabit Ethernet for connection to the JUSTCOM storage system, providing backend storage for virtual machines and block devices via NFS. In addition specific NFS exports of the XCST storage system at JSC are available. Service and login nodes use a 100 Gigabit Ethernet link for uplink and storage connection.

2.2 Cluster Software

The JUSUF software stack, as well as the software stack on all the systems operated by JSC, is largely based on open-source software. On the login and compute nodes the CentOS 7 Linux operating system is used.

All JSC-operated software is deployed and configured via Ansible (Red Hat Inc, 2021), a software provisioning and configuration management tool. The use of Ansible enables to reuse configuration across nodes and systems easily, and facilitates homogeneity, traceability and reproducibility.

JUSUF uses the open-source Slurm workload manager (SchedMD LLC, 2019) in combination with the ParaStation resource management which has a proven track record in scalability, reliability and performance on several clusters operated by JSC. The ParTec ParaStation ClusterSuite (ParTec Cluster Competence Center GmbH, 2019) is used for node imaging and health monitoring.

The management of the scientific software stack in JUSUF relies on EasyBuild (Hoste et al., 2012). JUSUF has GCC, Intel and NVHPC compilers available. Support for AOCC (AMD Optimizing C/C++ Compiler) is in development at the moment. The supported MPI (Message Passing Interface) implementations are mainly ParaStationMPI and OpenMPI. Different compilers, optimized mathematical libraries and pre-compiled community codes are available. We refer to the JUSUF webpage (Forschungszentrum Jülich, 2020) for more information. Monitoring of batch jobs is possible using the latest version of the graphical monitoring tool LLview (Forschungszentrum Jülich, 2019d) .

Scientists can also use UNICORE (UNICORE Forum e.V., 2019) to create, submit and monitor jobs on JUSUF. In addition, the system can be accessed via the Jupyter-JSC service (Forschungszentrum Jülich, 2019c).

2.3 Cloud Software

The JUSUF Cloud is the first of its kind that is operated at the High Performance Computing Systems department at JSC. The software stack is based on RedHat OpenStack Platform (RedHat OpenStack Platform, 2021) and started with version 16. In addition, parts of the JSC configuration management system for HPC systems is used.

Available OpenStack services and some key configuration decisions are listed below:

- Keystone: The OpenStack authorization service is connected to the JSC LDAP which is managed by JUDOOR. In addition, via OIDC a connection to the Fenix AAI is available to support researchers from different sites.
- Nova: The compute services use KVM. Live-migration is available for VMs because of a shared storage backend implemented via NFS. Virtual GPUs and NVMeS are made available via specific virtual machine flavors but are not available for live-migration.

- Cinder: The storage service has up to 300 TB of available capacity via the JUSTCOM system, which is based on GPFS and made available via NFS.
- Neutron: The network service uses Open Virtual Network as the Software Defined Networking backend. Beside of project-local networks, access to and from the internet is available via a pool of publicly available IPs. A direct connection to the XCST via NFS is available, strict requirements for security reasons are imposed.
- OpenStack Heat, Octavia and Glance are available as well with default configurations.

RedHat OpenStack Platform is using OpenStack Triple-O and every service mentioned above, in general every service, is containerized. The API services are deployed using a highly-available setup with three controller nodes.

3 Access to JUSUF

Scientists and engineers interested in using JUSUF for their research have to apply for JUSUF cluster computing time or cloud resources by submitting an adequate proposal in answer to corresponding computing time calls published in January/February and July/August every year. Those calls are managed by Fenix (Fenix RI, 2019) in collaboration with PRACE (Partnership for Advanced Computing in Europe, 2019) and provide 40% of JUSUF's resources. The remaining 60% are available for applicants from JSC/FZJ.

Submitted proposals are evaluated scientifically through a review process. Additionally, the review process includes a technical assessment of the project to make sure that it fits to the offered resources.

References

- Fenix RI. (2019). *Fenix Research Infrastructure webpage*. Retrieved from <https://fenix-ri.eu>
- Forschungszentrum Jülich. (2019a). *Forschungszentrum Jülich webpage*. Retrieved from <http://www.fz-juelich.de>
- Forschungszentrum Jülich. (2019b). *Jülich Supercomputing Centre webpage*. Retrieved from <http://www.fz-juelich.de/ias/jsc>
- Forschungszentrum Jülich. (2019c). *Jupyter@JSC webpage*. Retrieved from <http://jupyter-jsc.fz-juelich.de>
- Forschungszentrum Jülich. (2019d). *LLview webpage*. Retrieved from <http://www.fz-juelich.de/jsc/llview>
- Forschungszentrum Jülich. (2020). *JUSUF webpage*. Retrieved from <http://www.fz-juelich.de/ias/jsc/jusuf>
- Gauss Centre for Supercomputing. (2019). *Gauss Centre for Supercomputing webpage*. Retrieved from <http://www.gauss-centre.eu>
- Hoste, K., Timmerman, J., Georges, A., & De Weirde, S. (2012). EasyBuild: Building Software with Ease. In *2012 SC Companion: High Performance Computing, Networking Storage and Analysis* (p. 572-582). <http://dx.doi.org/10.1109/SC.Companion.2012.81>
- Human Brain Project. (2020). *HBP webpage*. Retrieved from <https://www.humanbrainproject.eu>
- Interactive Computing E-Infrastructure for the Human Brain Project. (2021). *ICEI project information at Horizon 2020*. Retrieved from <https://cordis.europa.eu/project/id/800858>

ParTec Cluster Competence Center GmbH. (2019). *ParTec webpage*. Retrieved from <http://www.par-tec.com>

Partnership for Advanced Computing in Europe. (2019). *Partnership for Advanced Computing in Europe webpage*. Retrieved from <http://www.prace-ri.eu>

Red Hat Inc. (2021). *Ansible Configuration Manager webpage*. Retrieved from <https://www.ansible.com>

RedHat OpenStack Platform. (2021). *RedHat OpenStack Platform webpage*. Retrieved from <https://www.redhat.com/de/technologies/linux-platforms/openstack-platform>

SchedMD LLC. (2019). *Slurm Workload Manager webpage*. Retrieved from <http://slurm.schedmd.com>

T, H. P. S. (2020). *HPST webpage*. Retrieved from <https://www.link-to-hpst?>

UNICORE Forum e.V. (2019). *Uniform Interface to Computing Resources (UNICORE) webpage*. Retrieved from <http://www.unicore.eu>