SYSTEMAUSBAU JSC

2024-11-05 I THOMAS EICKERMANN, BENEDIKT VON ST. VIETH | JÜLICH SUPERCOMPUTING CENTRE

Member of the Helmholtz Association
EuroHPC Joint Undertaking
Bundesministerium für Bildung und Forschung
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen
GCS Gauss Centre for Supercomputing
JÜLICH Forschungszentrum Shaping Change

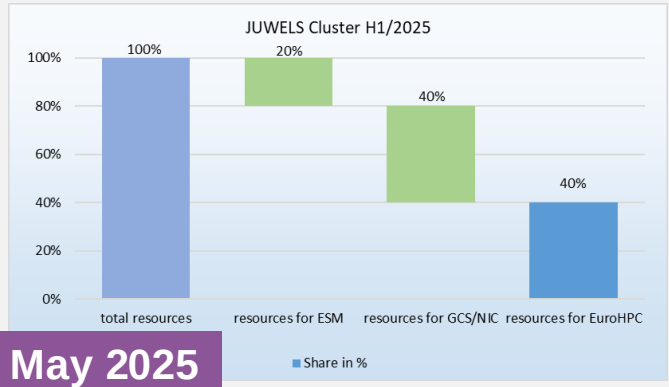
SYSTEMSTATUS

GCS SYSTEM @ JÜLICH

JUWELS (Jülich Wizard for European Leadership Science): Modular Supercomputer

JUWELS Cluster

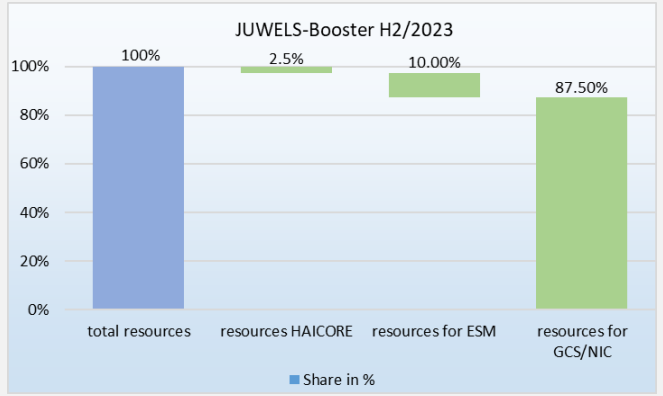
- Intel Skylake based system with **12 PF** peak (CPU:10.6 PF, GPU: 1.7 PF)
- 10 cells with altogether more than 2,500 nodes or 120,000 cores
- Mellanox InfiniBand EDR fat-tree network (2:1 pruning at leaf level)
- Entered #23 in Jun 2018 Top500



Starting from May 2025

JUWELS Booster

- Nvidia A100 based system with **75 PF** peak (CPU: 2 PF, GPU 73 PF)
- 936 nodes with 4 Nvidia A100 graphics cards each
- Mellanox InfiniBand HDR DragonFly+ topology with 20 cells - 5 TB/s connection to Cluster
- Entered #7 in Nov 2020 Top500, #1 in Europe, #1 in Green250



→ Connected to file server **JUST** with about **100 PB disk** capacity and more than **300 PB tape** capacity

HPC OPERATIONS

	JUWELS		JURECA		JUSUF
	Cluster	Booster	Data Centric	WestAI	Cluster
Installation	2017	2020	2020/2021	2024	2020
Peak Petaflops/s	12,3	75	18,5	3 (6)	1,4
Node Count	2567	936	768	16 (32)	180 (205)
Accelerators	224 GPUs	3744 GPUs	768 GPUs	64 (128)	61 GPUs
Resources Via	• GCS/NIC • ESM	• GCS/NIC • ESM/AI	• EU Fenix/PRACE • FZJ	• WestAI • Helmholtz	• EU Fenix/PRACE • FZJ
End of Life	2025/2026	2026	2026	2029	2025

STORAGE IN 2024FF

JUST6

- 154 PB raw
- 700 GB read, 500 GB write
- 22 TB Spinning Disk
- IBM SSS6000
- IBM Storage Scale
- 100 GE fabric, like JUST5

XCST (fading out)

- 45 PB raw (32 net)
- Lenovo DE6000

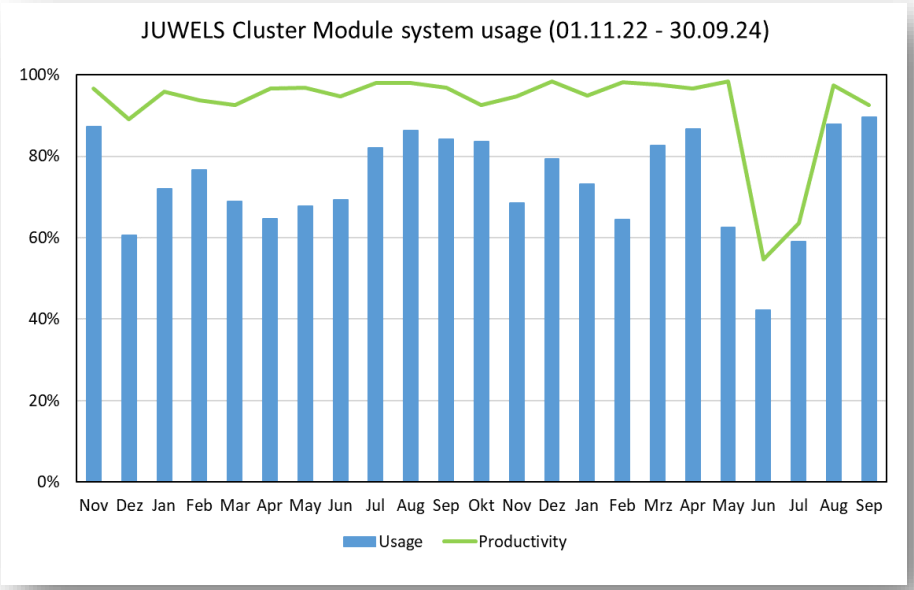
OBJECT (on XCST)

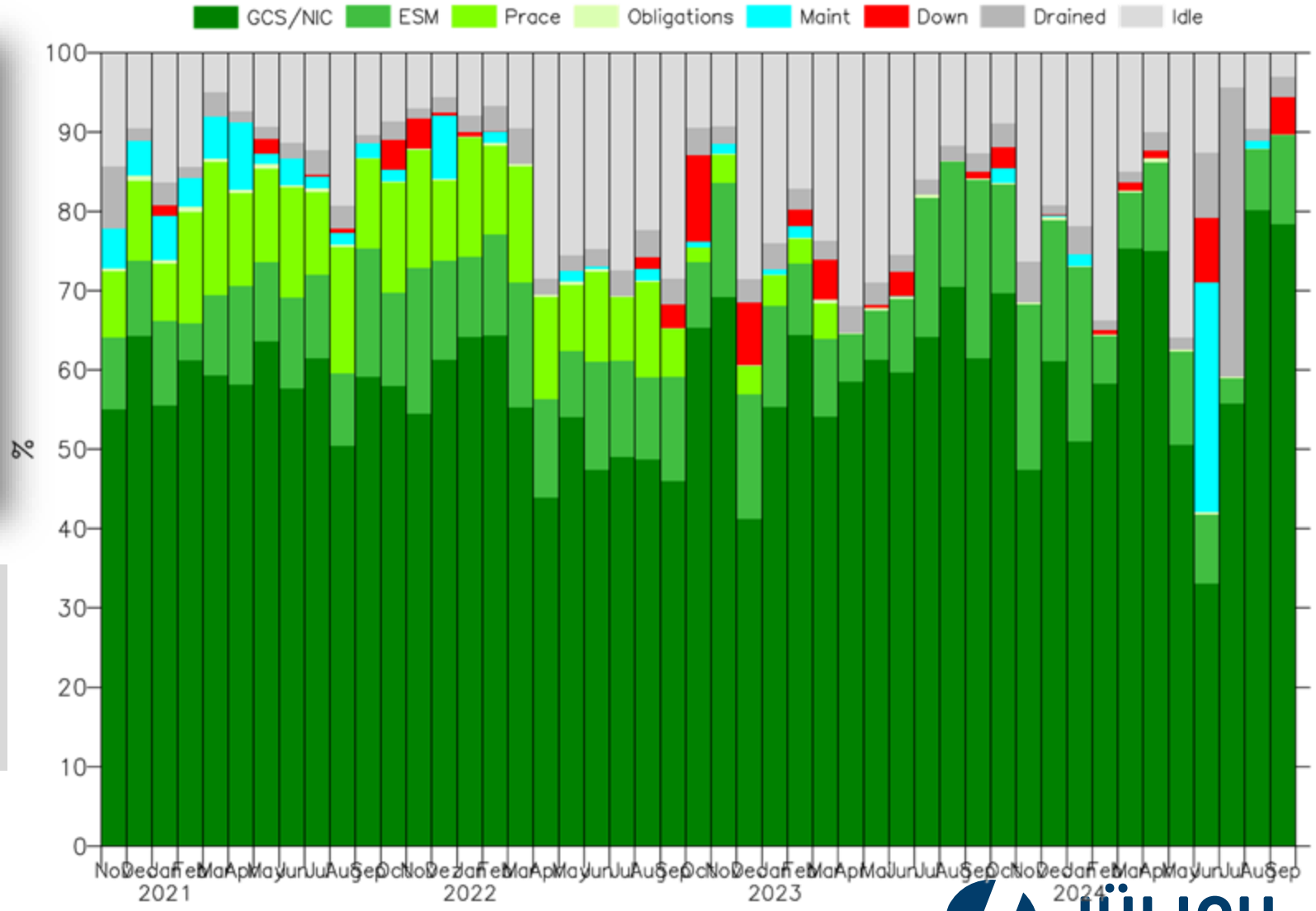
- 10 PB raw (3 net)
- OpenStack Swift/S3
- Deprecated

ARCHIVE/Backup

- ~476 PB
- ~42.775 Tapes
- 3 Libraries
- 14PB HSM
- IBM Storage Protect

JUWELS CLUSTER USAGE



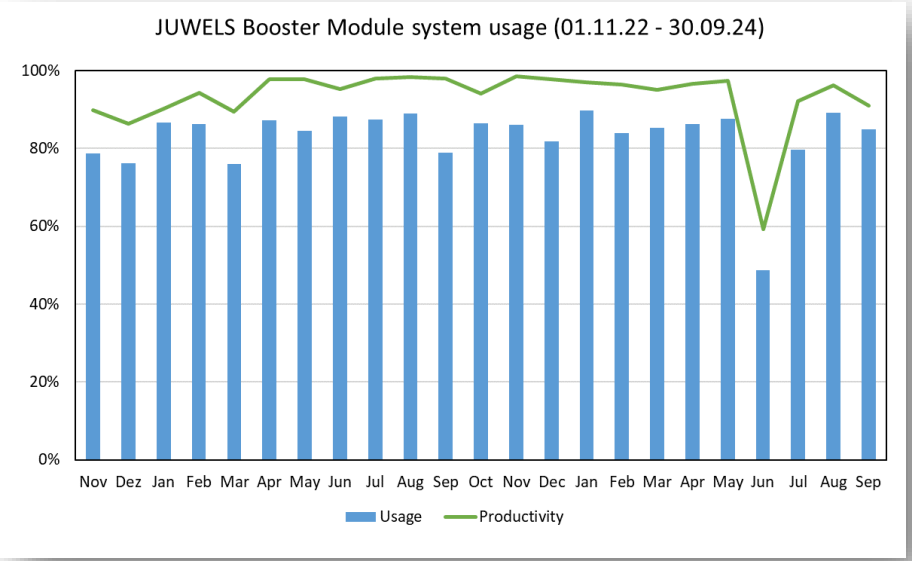


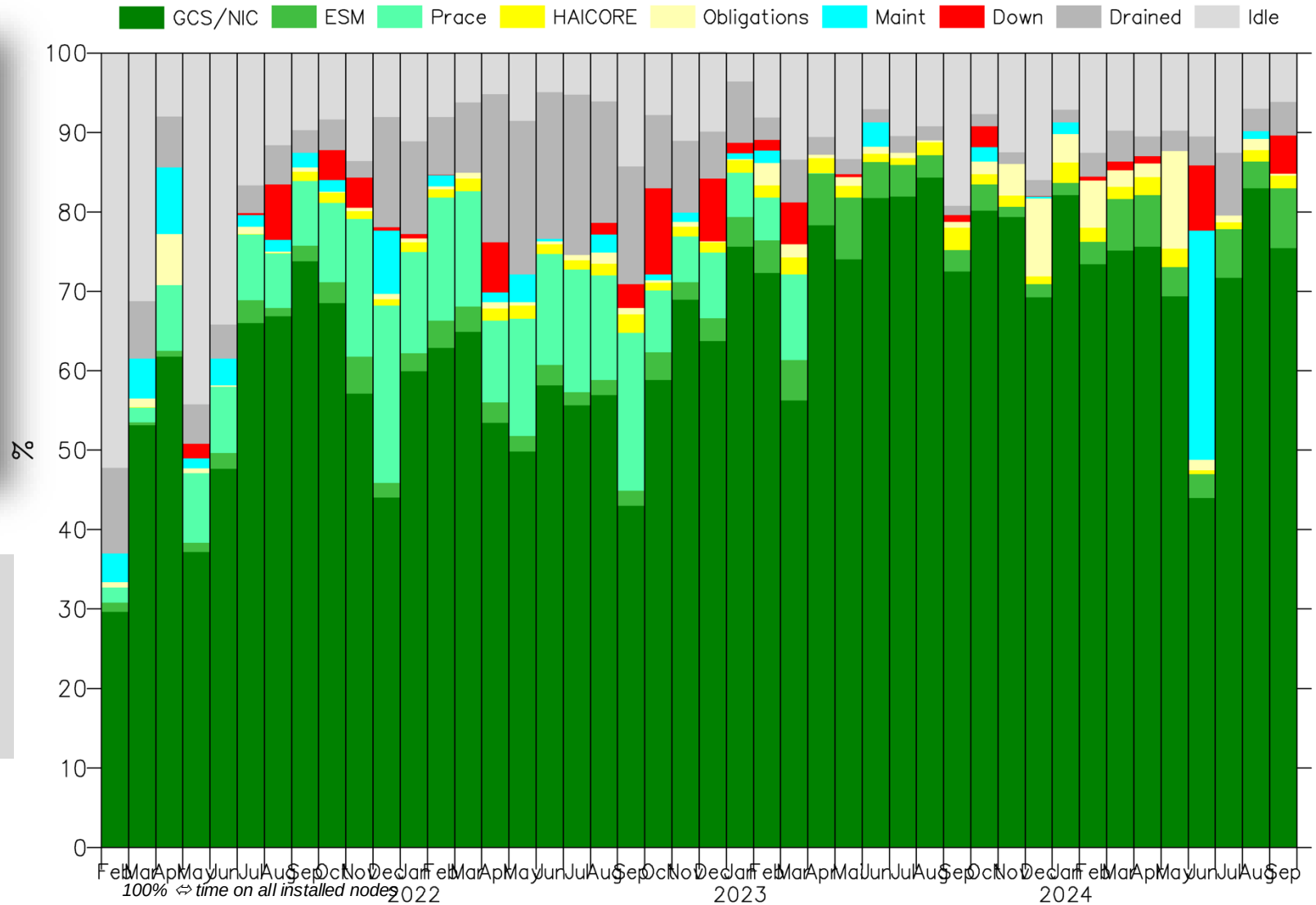
2024:

- JUST6 migration / Power outage datacenter
- 2x leakage in heat-exchanger (in JUWELS racks)

→ replacement of cooling components in racks of JURECA and JUWELS

JUWELS BOOSTER USAGE



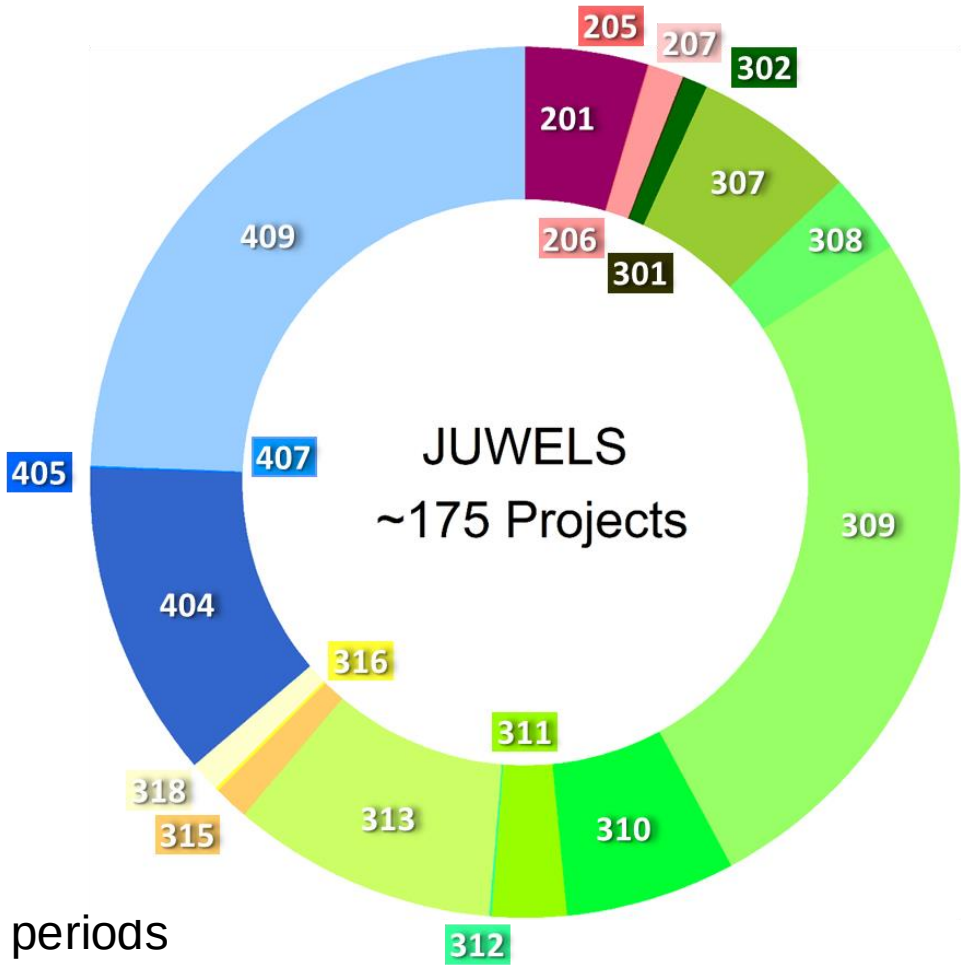


2024:

- JUST6 migration / Power outage datacenter
- 2x leakage in heat-exchanger (in JUWELS racks)

→ replacement of cooling components in racks of JURECA and JUWELS

JUWELS (CLUSTER + BOOSTER): GCS RESEARCH FIELDS



Reseach Fields

201	Basic Biological and Medical Research
205	Medicine
206	Neurosciences
207	Agriculture, Forestry and Veterinary Medicine
301	Molecular Chemistry
302	Chemical Solid State and Surface Research
307	Condensed Matter Physics
308	Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas
309	Particles, Nuclei and Fields
310	Statistical Physics, Soft Matter, Biological Physics, Nonlinear Dynamics
311	Astrophysics and Astronomy
312	Mathematics
313	Atmospheric Science, Oceanography and Climate Research
315	Geophysics and Geodesy
316	Geochemistry, Mineralogy and Crystallography
318	Water Research
404	Heat Energy Technology, Thermal Machines, Fluid Mechanics
405	Materials Engineering
407	Systems Engineering
409	Computer Science

Granting periods

11/2023 – 10/2024

05/2024 – 04/2025

OVERALL SYSTEM STATUS IN 2024

- Availability JUWELS Cluster, Booster, JURECA-DC and JUSUF good
  - >95-98%
- Migration JUST5 -> JUST6: One week in June 2024, heavily impacted by JUWELS leakage
- Power Saving (node shutdown): Enabled everywhere but not JUWELS (high utilization)
- Flipping Links (JWB, DC): Root cause found, component replacement ongoing
- Maintenance: From block maintenance to rolling, higher availability
- MFA implemented and tested 2023, rolled out to users in 2024
  - 05/2024: Promote opt-in for all users
  - 10/2024(?): ~~Change to opt-out~~
  - 05/2025(?): Change to opt-out

THE PRESENT

JUPITER CONTRACT ANNOUNCEMENT 3.10.2023

HPC wire

Since 1987 - Covering the Fastest Computers in the World and the People Who Run Them

- Home
- Topics
- Sectors
- Exascale
- Specials
- Resource Library
- Podcast
- Events
- Solution Channels
- Job Bank
- About
- Subscribe

EU Grabs ARM for First ExaFLOP Supercomputer, x86 Misses Out

By Agam Shah

October 4, 2023

The configuration of Europe's first exascale supercomputer, Jupiter, has been finalized, and it is a win for Nvidia and a disappointment for x86 chip vendors Intel and AMD. The Jupiter supercomputer, which will cost €273 million to build, will pair SiPearl's Rhea processor, which is based on ARM architecture, with accelerator technology from Nvidia.

The supercomputer is being built by the European High-Performance Computing Joint Undertaking (EuroHPC JU) and a consortium including Eviden and ParTec. Eviden is an Atos business focusing on advanced computing initiatives tha

The Jülich Supercompu
(*Correction: Jülich Supe
about 600 km or 375 m

Specifically, the superc
CPUs, and the initial co
supercomputers in the
on ARM.

That is a big disappoint
invest €33 billion to buil
development initiatives
leaders in a bid to get m

Jülich's fastest system,
November 2021 and is
Europe are the third-ran
performance of 309 pet

Off The Wire Industry Headlines

October 13, 2023

- Coherent File Format Accelerates Time-to-Solution with OpenFOAM
- HealthyCloud Project Unveils Roadmap to Maximize Impact of Health Data and Research Across Europe
- NCSA Welcomes 2023-24 Fellows
- Berkeley Lab CS Area to Share Computing Expertise at SC23

October 12, 2023

- Samsung Electronics to Host AI Forum 2023 Highlighting AI and Computer Engineering Innovation
- PacBio Announces Complete Computational Workflow for Human Whole Genome Sequencing Data Analysis
- SiFive Announces Differentiated Solutions for Generative AI and ML Applications
- EQTC 2023: Europe's Quantum Sector to Showcase Successes and Its Roadmap for Global Leadership
- EuroHPC JU Announces Procurement Call for Upgrading Discoverer Supercomputer
- Los Alamos Partners with AirMettle for Efficient In-Storage Data Analysis
- Caltech Researchers Demonstrate Quantum Eraser to Combat Erasure Errors in Quantum Systems
- Research Results, Computational Expertise Flourish at

SIGN IN / UP

The Register

HPC

Atos subsidiary Eviden scores contract win in Europe's first exascale system

$526M Jupiter set to rule EU's tech orbit by 2024

Dan Robinson

Wed 4 Oct 2023 | 16:45 UTC

The EU's supercomputing initiative, the European High Performance Computing Joint Undertaking (EuroHPC JU), has awarded a procurement contract for Europe's first exascale system, with installation due to start in early 2024.

Known as Jupiter (Joint Undertaking Pioneer for Innovative and Transformative Exascale Research), the system was announced last year followed by a call for tender in January of this year.

EuroHPC JU said the procurement contract for Jupiter was awarded to a consortium comprising of Eviden, the professional services side of French IT giant Atos, and ParTec, a German supercomputing hardware company.

The project is expected to have a total cost of €273 million ($287 million) covering the build, delivery, installation, and maintenance of Jupiter, according to the EuroHPC.

However, Eviden put the overall project cost at €500 million ($526 million), saying that this is the figure for the entire project, including the system manufacturing and its

THENEXTPLATFORM

HOME | COMPUTE | STORE | CONNECT | CONTROL | CODE | AI | HPC | ENTERPRISE | HYPERSCALE | CLOUD

LATEST > Intel To Set Its FPGA Unit Free To Pursue Its Own Path > COMPUTE

Search ...

HOME > HPC > Details Emerge On Europe's First Exascale Supercomputer

DETAILS EMERGE ON EUROPE'S FIRST EXASCALE SUPERCOMPUTER

October 5, 2023 Timothy Prickett Morgan

me details are emerging on Europe's first exascale system, codenamed "Jupiter" and to be installed at the ich Supercomputing Center in Germany in 2024. There has been a lot of speculation about what Jupiter ll include for its compute engines and networking and who will build and maintain the system. We now ow some of this and can infer some more from the statements that were made by the organizations rticipating in the Jupiter effort.

June 2022, the Forschungszentrum Jülich in Germany, which has played host to many supercomputers ce it was founded in 1987, was chosen to host the first of three European exascale-class percomputers to be funded through the EuroHPC Joint Undertaking and through the European national d state governments countries who are essentially paying to make sure these HPC and AI clusters are ere they want them. With Germany having the largest economy in Europe and being a heavy user of C thanks to its manufacturing focus, Jülich was the obvious place to park the first machine in Europe to eak the exaflops barrier.

at barrier is as much an economic one as it is a technical one. The six-year budget for Jupiter weighs in at 00 million, which is around $526.1 million at current exchange rates between the US dollar and the ropean euro. That is in the same ballpark price as what the "Frontier" exascale machine at Oak Ridge tional Laboratory and the "El Capitan" machine that is being installed right now at Lawrence Livermore tional Laboratory – both of which are based on a combination of AMD CPUs and GPUs and Hewlett ckard Enterprise's Slingshot variant of Ethernet with HPE as the prime contractor.

erybody knows that Jupiter was going to use SiPearl's first generation Arm processor based on the verse "Zeus" V1 core from Arm Ltd, which is codenamed "Rhea" by SiPearl and which is appropriate

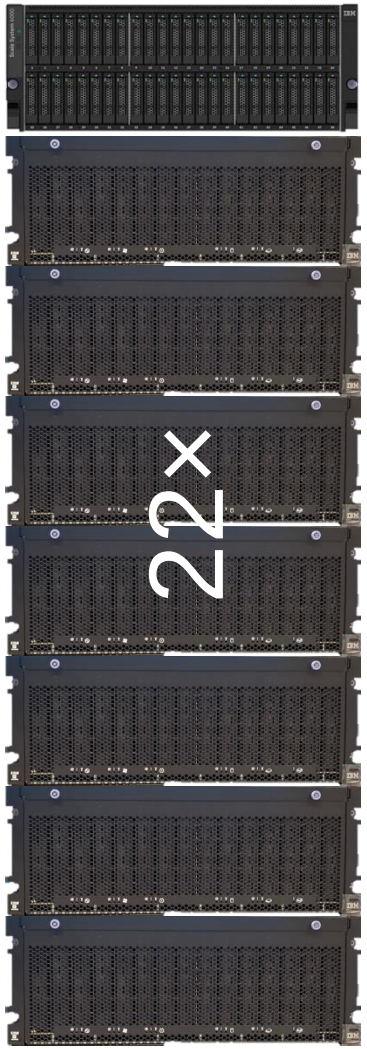
2023/2024 - THE PRESENT - SLAB, MDC, STORAGE, JEDI

… it is not only about waiting for JUPITER

ExaSTORE Procurement
300PB+ Spinning Disk Storage for JUPITER

JUPITER – EXASTORE

- >300 Petabyte Spinning Disk Raw Capacity
  - 14014 x 22 TB enterprise disks
  - Or e.g. 154154 x 2 TB consumer disks
- IBM Storage Scale System 6000
- HOME, PROJECT, DATA filesystems



2023/2024 - THE PRESENT - SLAB, MDC, STORAGE, JEDI

… it is not only about waiting for JUPITER

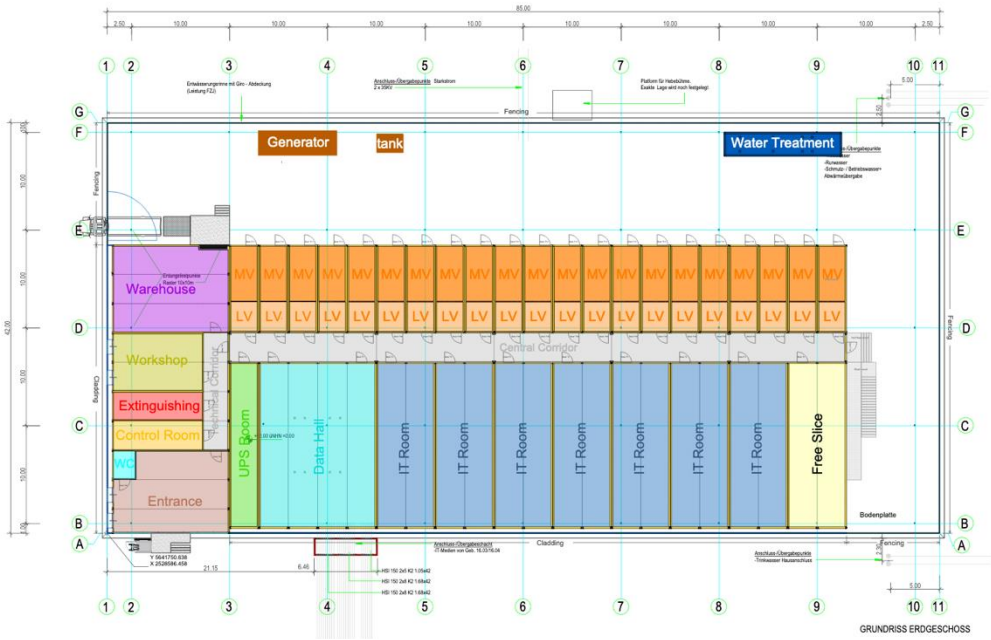
ExaSTORE Procurement
300PB+ Spinning Disk Storage for JUPITER

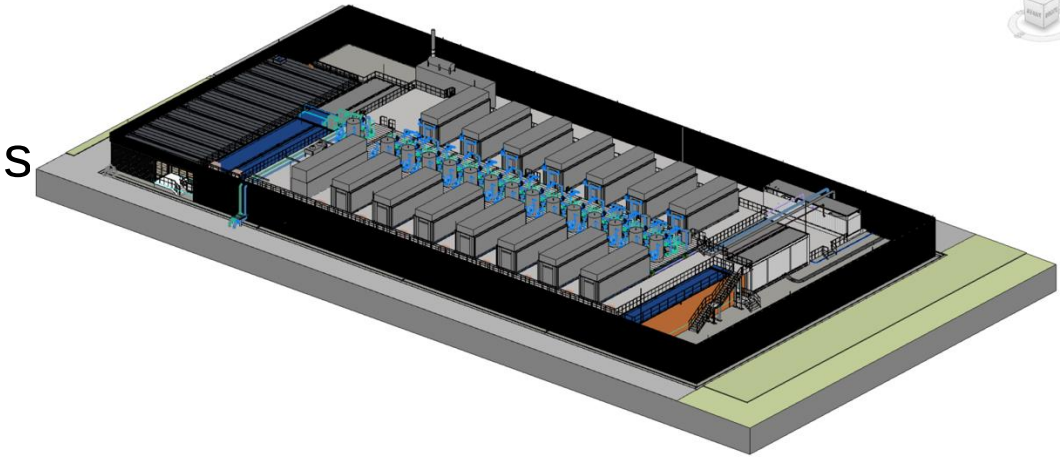
MDC Procurement

MODULAR DATA CENTER FOR JUPITER

EVIDEN

- Vendor: Eviden
- Area: ~2300m^2
- 1x Datahall (Storage, Management)
- 7x IT Modules (20 Racks per module)
- UPS, Generator
- Entrance area
- Workshop, Warehouse
- 15x 2,5 Megawatt Power Stations





MODULAR DATA CENTER FOR JUPITER

EVIDEN

2023/2024 - THE PRESENT - SLAB, MDC, STORAGE, JEDI

… it is not only about waiting for JUPITER

ExaSTORE Procurement
300PB+ Spinning Disk Storage for JUPITER

MDC Procurement

Concrete Foundation (Slab)

CONCRETE FOUNDATION

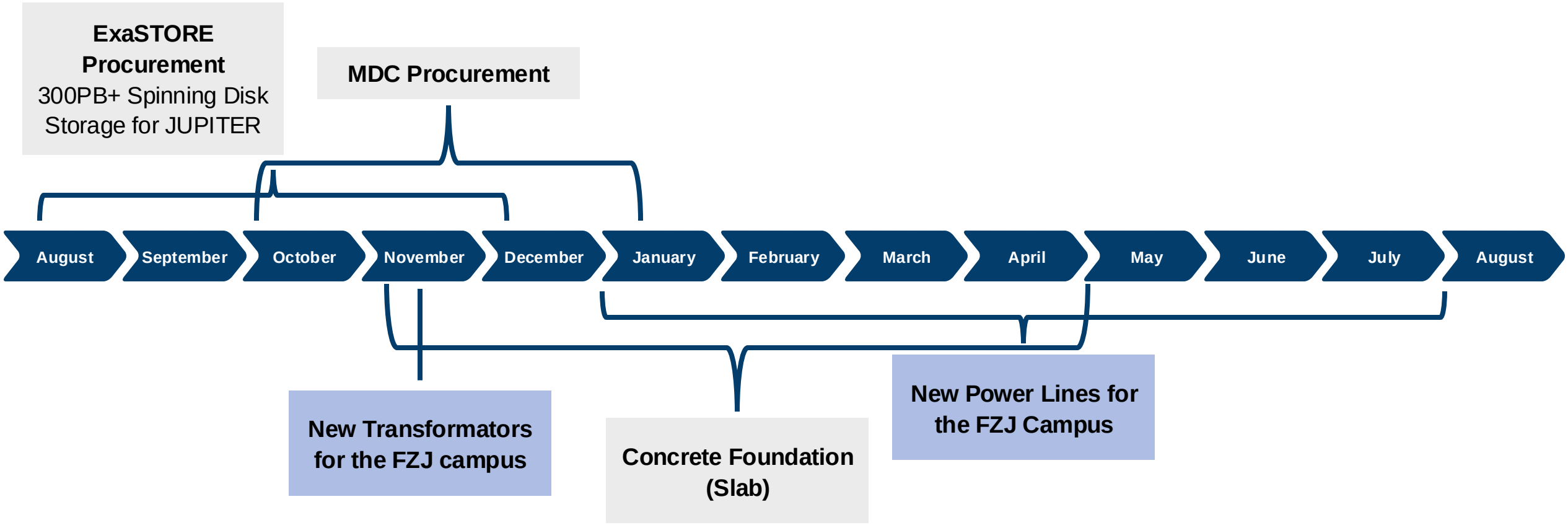
CONCRETE FOUNDATION

Construction of concrete slab 85 m x 42 m x 0.5 m


2024-08-15 17:29:56

2023/2024 - THE PRESENT - SLAB, MDC, STORAGE, JEDI

… it is not only about waiting for JUPITER

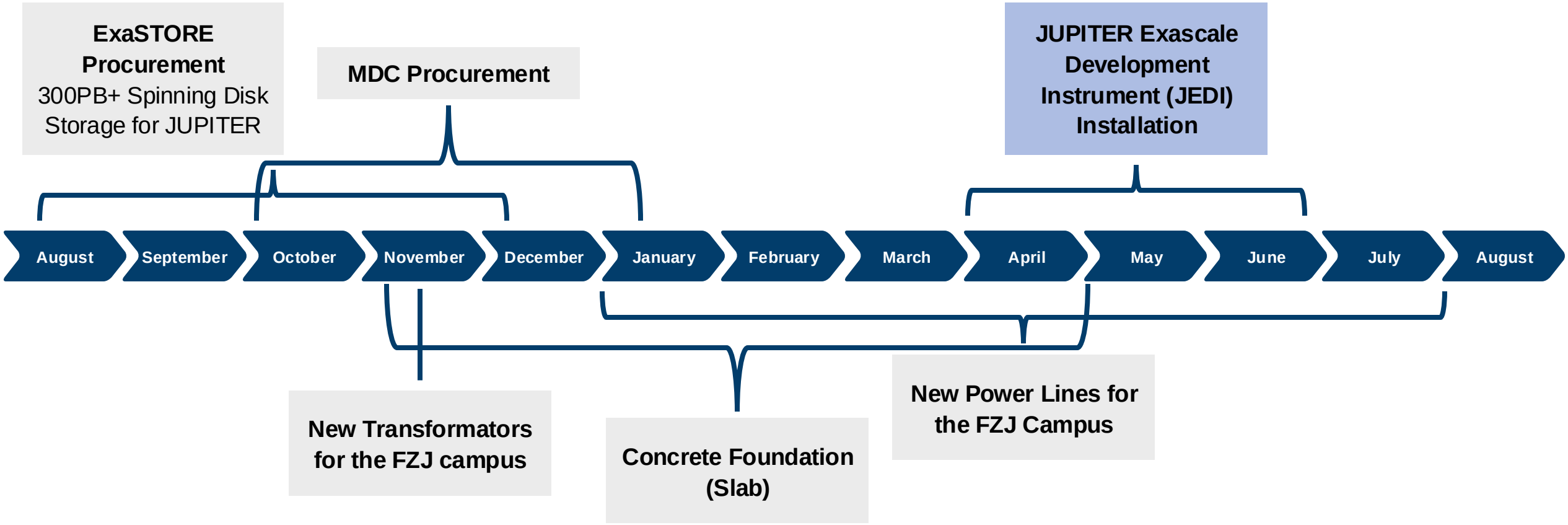


POWER TRANSFORMER SUBSTATION AND LINES

Upgrade of transformers 110 kV / 35 kV from 2 x 40 MVA to 2 x 60-80 MVA and upgrade 110kV power line

2023/2024 - THE PRESENT - SLAB, MDC, STORAGE, JEDI

… it is not only about waiting for JUPITER



OPMENT INSTRUMENT (JEDI)

Public Relations

JEDI

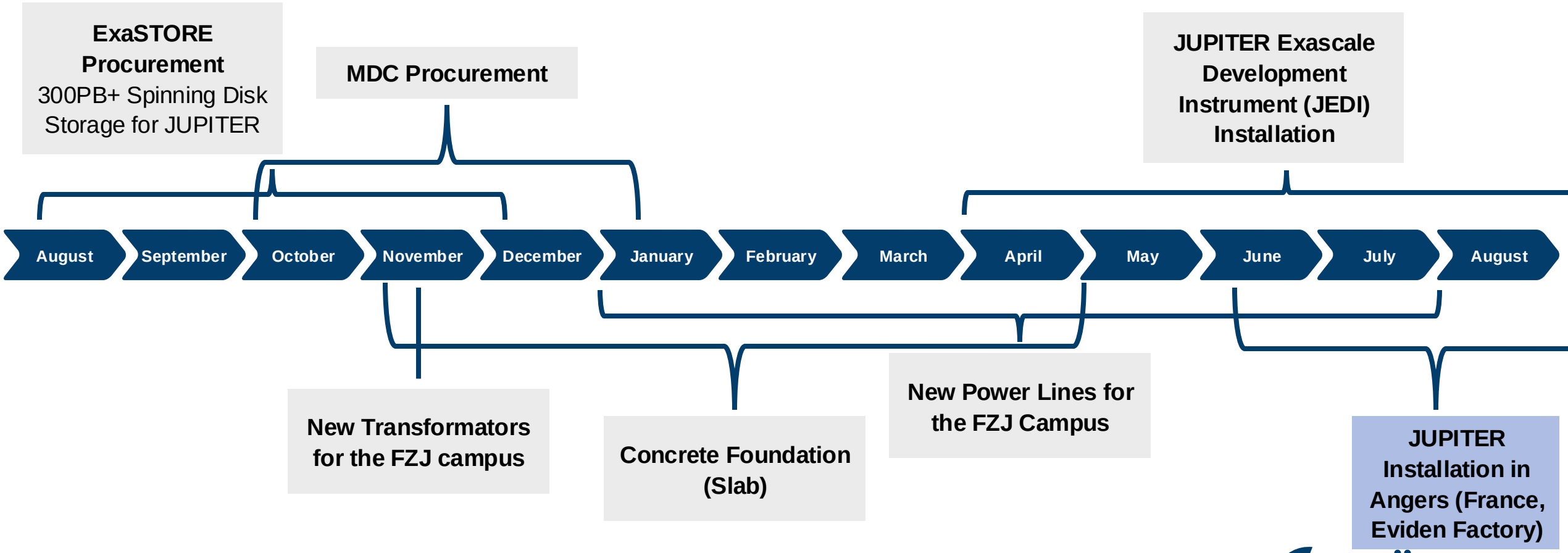
#1 in Green500 (05/2024)

#189 in TOP500

JÜLICH
Forschungszentrum

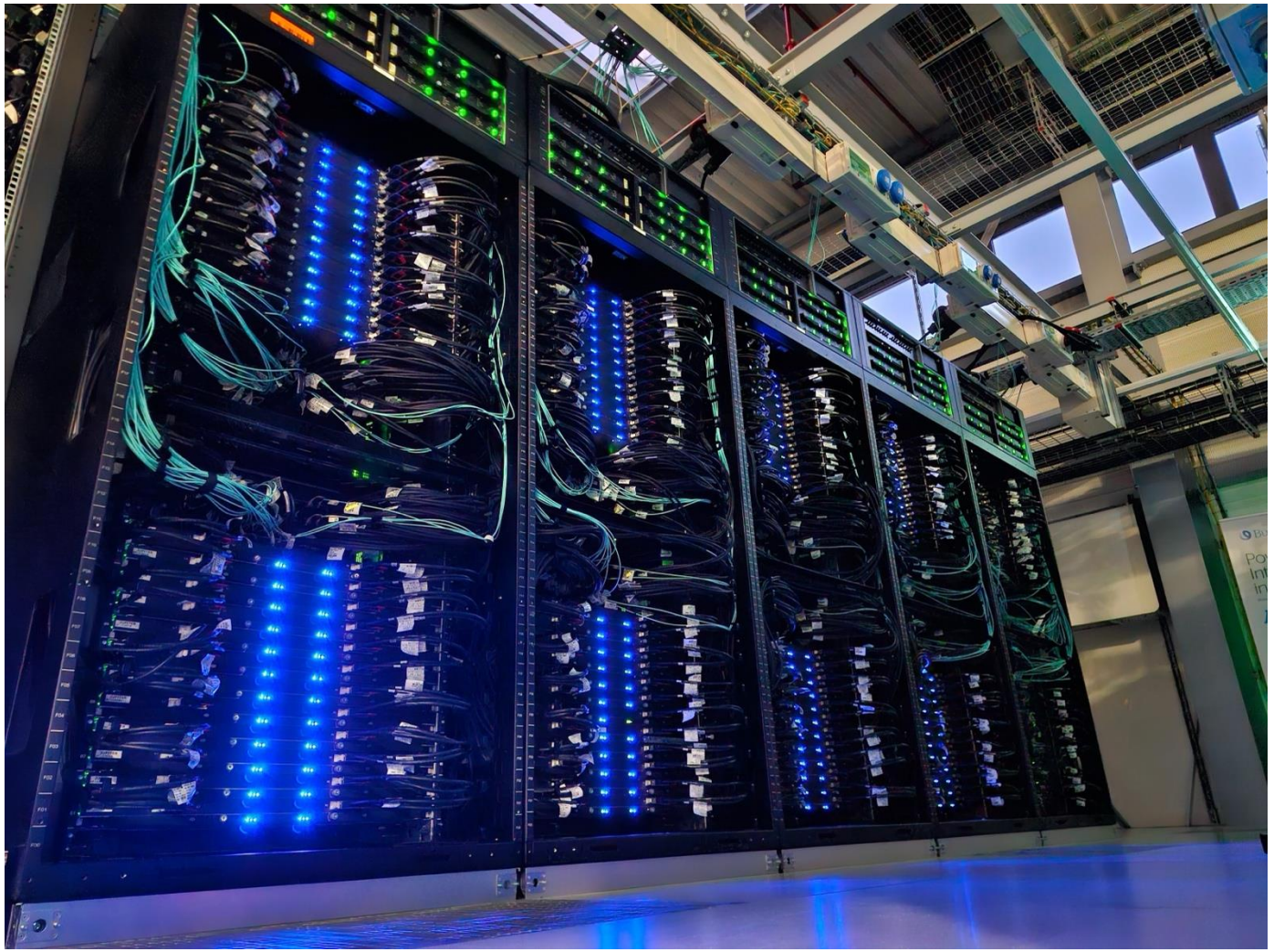
2023/2024 - THE PRESENT - SLAB, MDC, STORAGE, JEDI

… it is not only about waiting for JUPITER



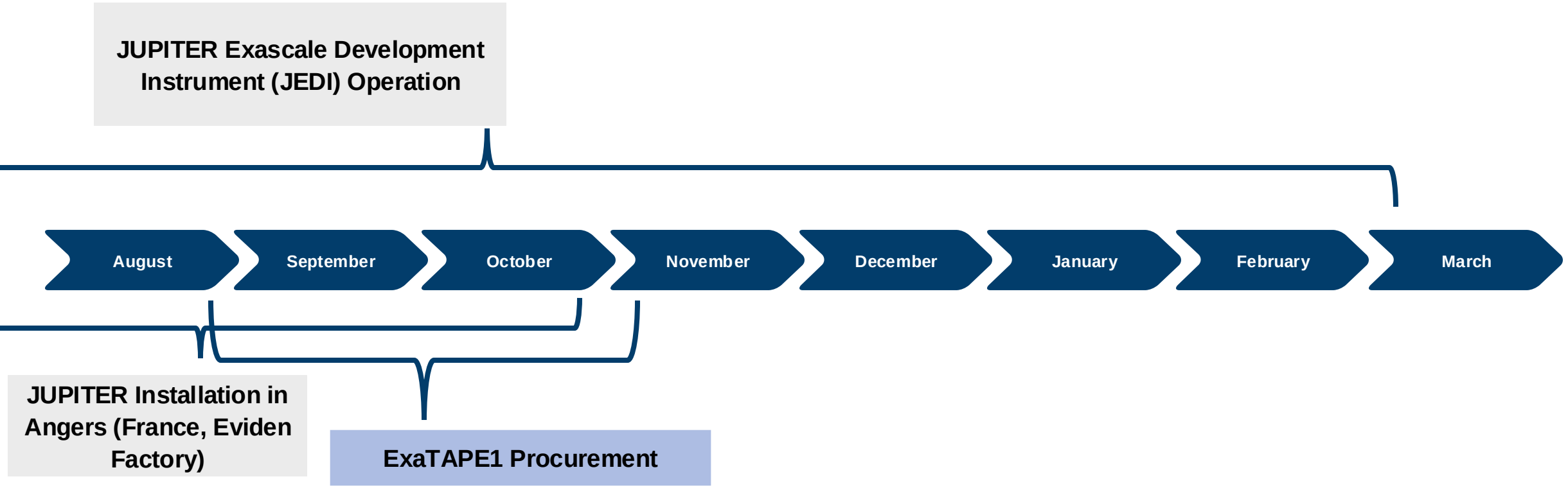
JUPITER INSTALLATION IN ANGERS (EVIDEN FACTORY)

- 10 XH3000 racks, 480 nodes
- Hardware tests and benchmark preparation
- JUPITER Management Stack installation
  - Currently driven by JSC
- **Afterwards**
  - Shipment to Jülich
  - “Plug in and run”



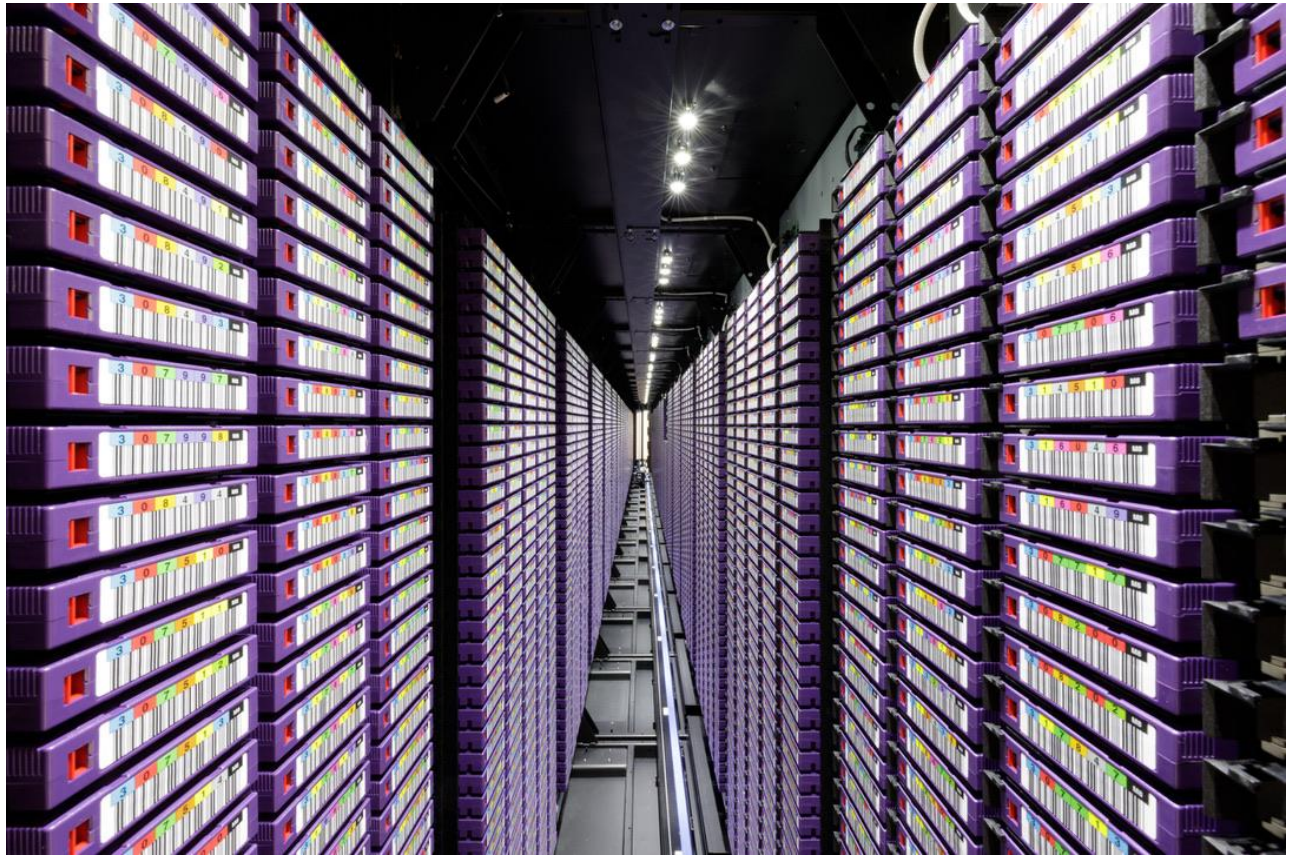
THE FUTURE

Difficult to see; always in motion is the future…



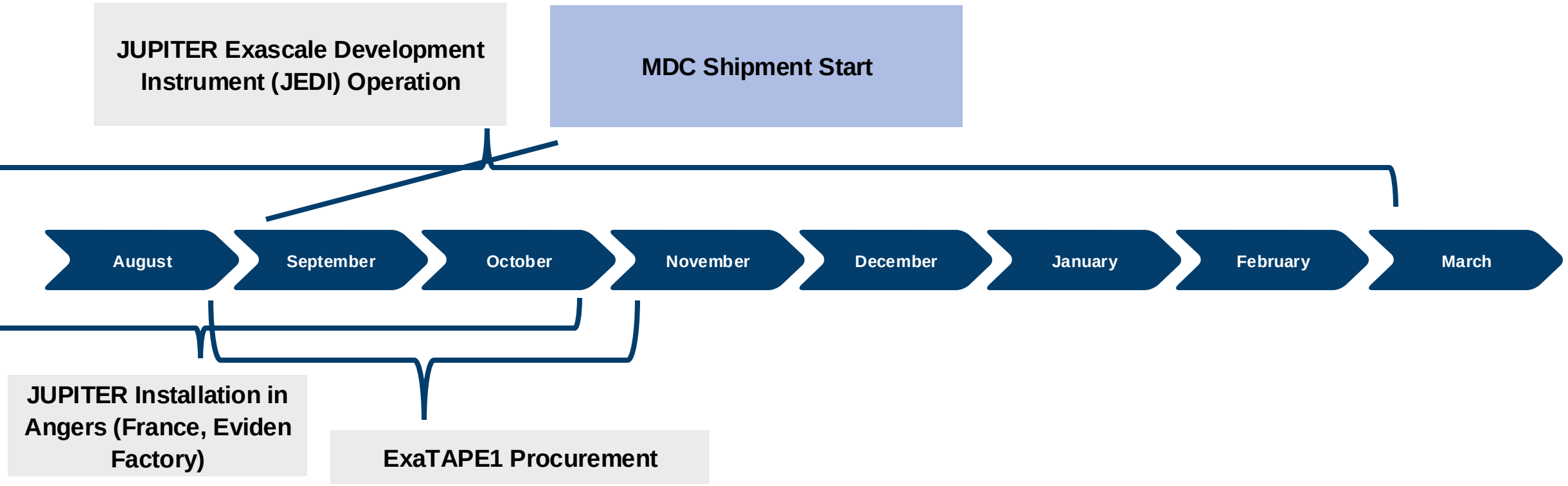
JUPITER – EXATAPE1

- >240 Petabyte Tape Capacity
- LTO9
- 1 or 2 Libraries
- Integration into IBM Storage Protect Environment
- Bidders have been informed about elected winner



THE FUTURE

Difficult to see; always in motion is the future…



MDC SHIPMENT START

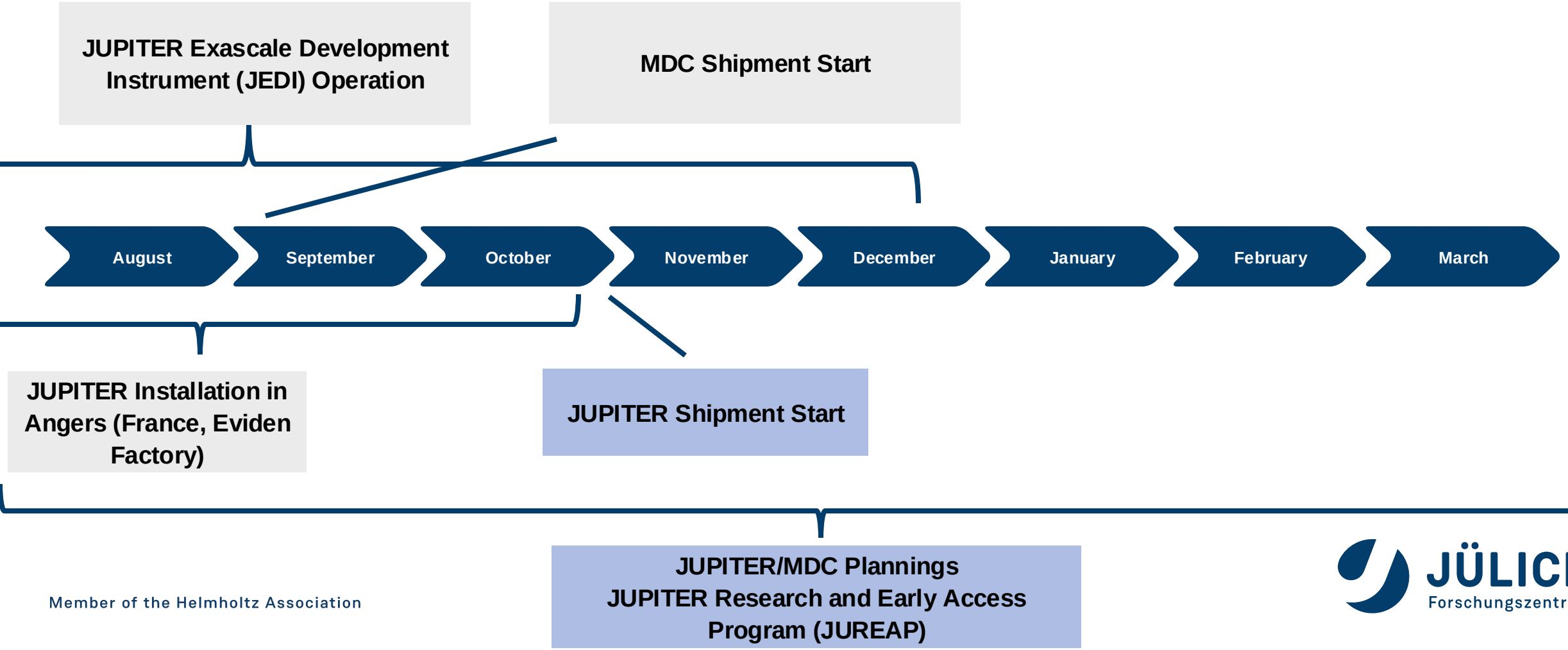
10./11.9.2024

THE DATAHALL ARRIVED

28.10.2024

THE FUTURE

Difficult to see; always in motion is the future…



JUREAP

Seeding Exascale in Europe!

jureap@fz-juelich.de • https://www.fz-juelich.de/en/ias/jsc/jupiter/jureap

JUPITER Research and Early Access Program

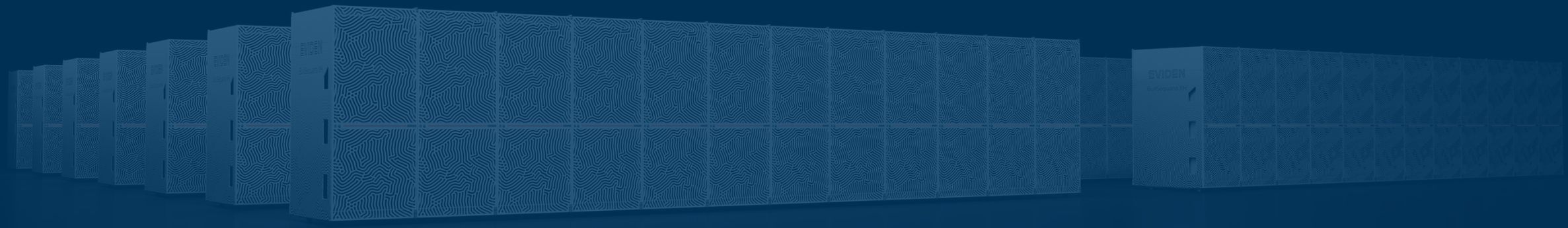
CURRENT STATE

Timeline and plans

IMPORTANT: Timeline not final yet!

Phase 1: Scalability and Performance Evaluation Phase (SPEP)
Phase 2: Porting and Optimizing Phase (POP)
Phase 3: Outstanding Research Phase (ORP)

JUPITER



EuroHPC Joint Undertaking

EVIDEN an atos business

JÜLICH Forschungszentrum

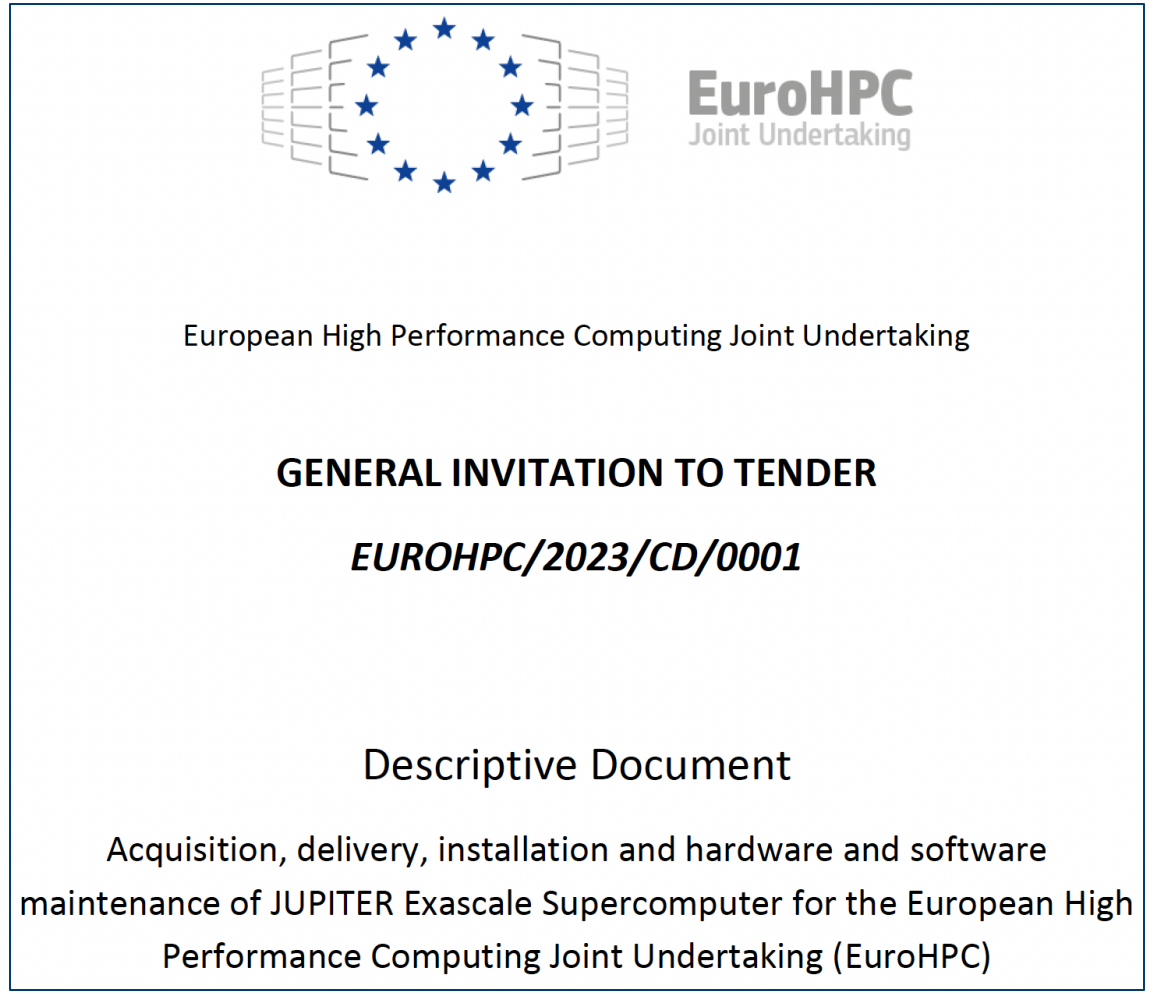
NVIDIA

ParTec

READY FOR TAKE OFF

Competitive Dialogue - Descriptive Document

- Description of procurement procedure
- Overall budget, **273 M€**
- High-level description of targeted system
  - Implementing the **MSA**
  - Booster to achieve **1 EF FP64**
  - Cluster, preferably based on **European IP**
  - Flash storage module
  - Interconnect expectations
  - Login system sizing
  - System management



EuroHPC Joint Undertaking

European High Performance Computing Joint Undertaking

GENERAL INVITATION TO TENDER

EUROHPC/2023/CD/0001

Descriptive Document

Acquisition, delivery, installation and hardware and software maintenance of JUPITER Exascale Supercomputer for the European High Performance Computing Joint Undertaking (EuroHPC)

DISCOVERING JUPITER

- ParTec/Eviden Supercomputer Consortium
- Implementing Modular Supercomputing Architecture
- JUPITER Booster: High scalability; 1 EFLOP/s HPL, >70 EFLOP/s FP8
- JUPITER Cluster: High versatility; 0.5 B/FLOP balance
- Network: 200/400 Gigabit NVIDIA Mellanox InfiniBand NDR
- Storage: 29 PB Flash IBM Storage Scale 6000
- 17 Megawatt Linpack Power Consumption
- Direct Liquid Cooled (36 -> 4x degree) to enable heat-reuse

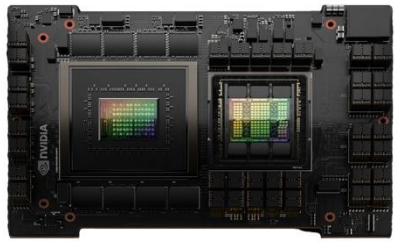
BullSequana XH3000, DLC

ParTec MODULAR SUPERCOMPUTING EVIDEN NVIDIA SIPEARL JÜLICH Forschungszentrum

JUPITER MODULES

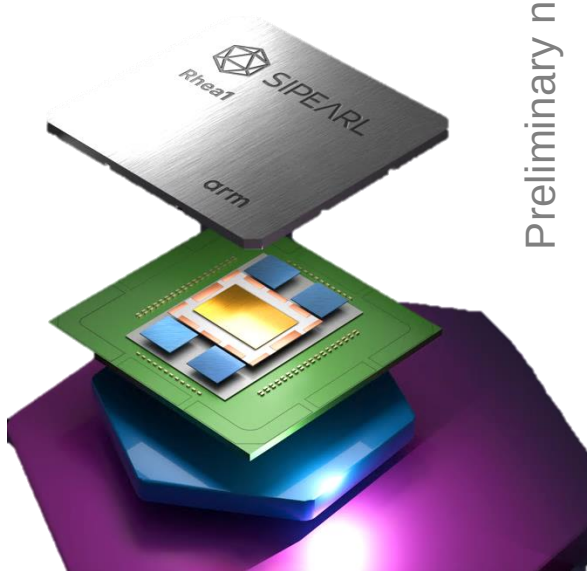
JUPITER Booster

- ~125 Racks BullSequana XH3000
- Node design
  - ~6000 nodes
  - 4× NVIDIA CG1 per node
- CG1: NVIDIA Grace-Hopper
  - 72 Arm Neoverse V2 cores (4×128b SVE2); 120 GB LPDDR5
  - H100 (132 SMs); 96 GB HBM3
  - NVLink C2C (900 GB/s)



JUPITER Cluster

- ~14 Racks BullSequana XH3000
- Node design
  - ~1300 nodes
  - 2× SiPearl Rhea1 per node
- Rhea1
  - 80 Arm Neoverse V1 cores (2×256b SVE)
  - 256 GB DDR5, 64 GB HBM2e



Preliminary numbers, might change during installation

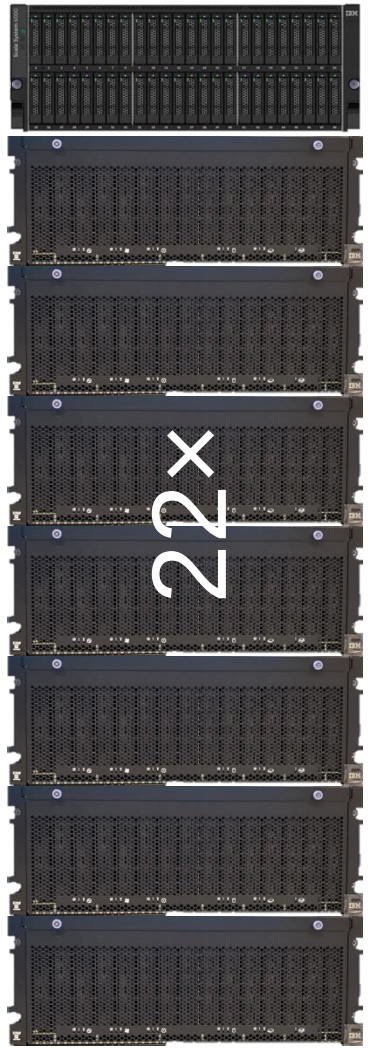
JUPITER – STORAGE (SCRATCH)

- Gross Capacity: 29 PB; Net Capacity: 21 PB
- Bandwidth: 2.1 TB/s Write, 3.1 TB/s Read
- 20× IBM SSS6000 Building Blocks (40 servers)
  - 2× NDR400 per server
  - 48× 30 TB NVMe drives per block
  - IBM Storage Scale (aka Spectrum Scale/GPFS)
- Manager and Datamover Nodes
- Exclusive for JUPITER
  - Integrated into InfiniBand fabric

JUPITER – STORAGE (EXASTORE)

In kind contribution from JSC, not part of the JUPITER procurement

- Gross Capacity: 308 PB; Net Capacity: 210 PB
- Bandwidth: 1.1 TB/s Write, 1.4 TB/s Read
- 22× IBM SSS6000 Building Blocks (44 servers)
  - 2× NDR200 per server
  - 7× JBOD enclosures, each with 91x 22 TB Spinning Disks per block
  - IBM Storage Scale (aka Spectrum Scale/GPFS)
- Manager and Datamover Nodes
- Exclusive for JUPITER
  - Integrated into InfiniBand fabric



JUPITER Management Stack

3 main pillars/actors

SMC xScale	ParaStation	xOPS
Core part of the stack. Vast majority of components come from here.	Enhancement of the core	Enhancement of the core
Developed by Eviden	Developed by ParTec	Developed by JSC
Heavily based on open source and cloud technologies	Integrates ParTec tools in SMCx to streamline their support workflows	Extensive set of Ansible roles for HPC, targeting JSC's requirements and needs

JUPITER MANAGEMENT STACK

A view from the vendor consortium

Ticketing Servers

SEMS
Smart Energy Management Suite

SDMS
Smart Data Management Suite

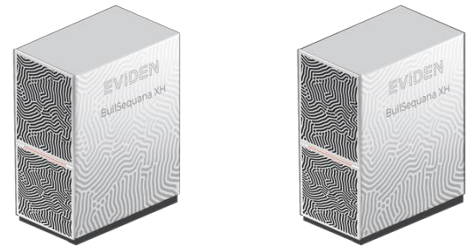
SMMS
Smart Maintenance Management Suite

IBMS
Interconnect faBric Management Suite

Red Hat Enterprise Linux 9

NVIDIA

SIPEARL



Management Servers

Red Hat Enterprise Linux 8

Reference
netbox

ParaStation
ParaStation CLUSTERTOOLS

Health Check
ParaStation HEALTHCHECKER

Deployment
Image Builder
Boot Manager

ceph

Resource Manager
slurm workload manager

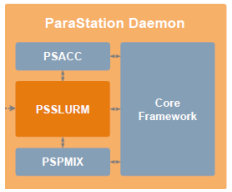
Automation Configuration
ANSIBLE
cloud-init

Monitoring Logging
Prometheus
Thanos
Grafana
fluentd
elastic

Hardware Interface
HAL
Hardware Abstraction Layer

SMC xScale

ParaStation



ParaStation MPI

OpenMPI

NVIDIA HPC

NVIDIA CUDA

arm COMPILER

Slide courtesy of EVIDEN

JÜLICH Forschungszentrum

BOOSTER NODE DESIGN

Node Specs

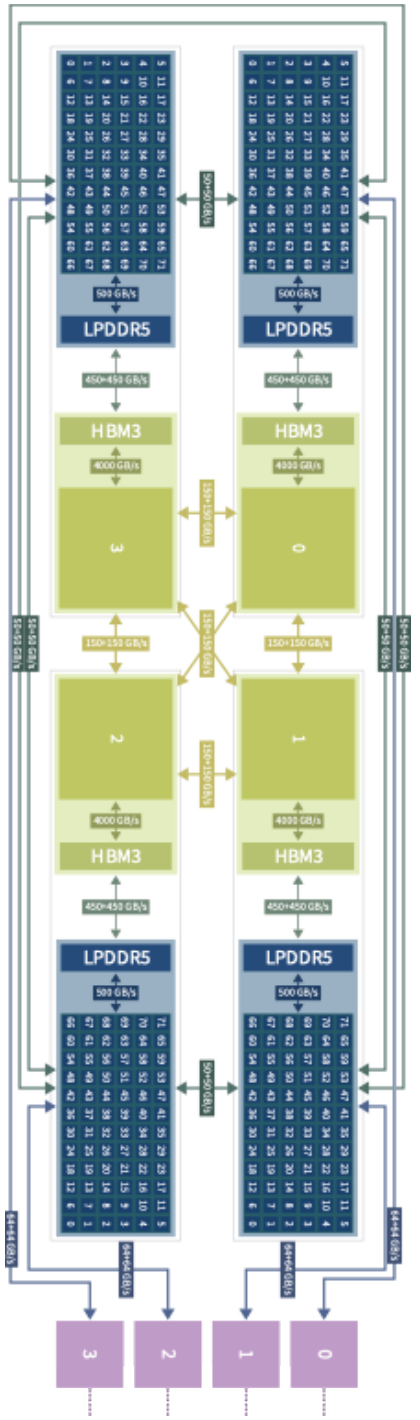
- 4× NVIDIA **Grace-Hopper** in SXM5 Board (4× 680W)
- 4× NVIDIA InfiniBand NDR200
- 480 GB LPDDR5X / 360 GB HBM3 (usable)
- NVLink 4
  - GPU-GPU 150 GB/s per dir
- Links: CPU-GPU 450 GB/s per dir, CPU-CPU 100 GB/s per dir

CPU Specs

- ARM Neoverse V2
  - SVE2/NEON (4x 128 bit vector op)
- 72 cores @ ~2.4GHz (~3.2 GHz turbo)
- 120 GB LPDDR5X (8 channels)
  - *≥450 GB/s*
  - *~150 ns latency*

GPU Specs

- H100
  - ~50 TFLOP/s (HPL single GPU)
- 90 GB HBM3
  - ≥3600 GB/s
  - ~450 ns latency



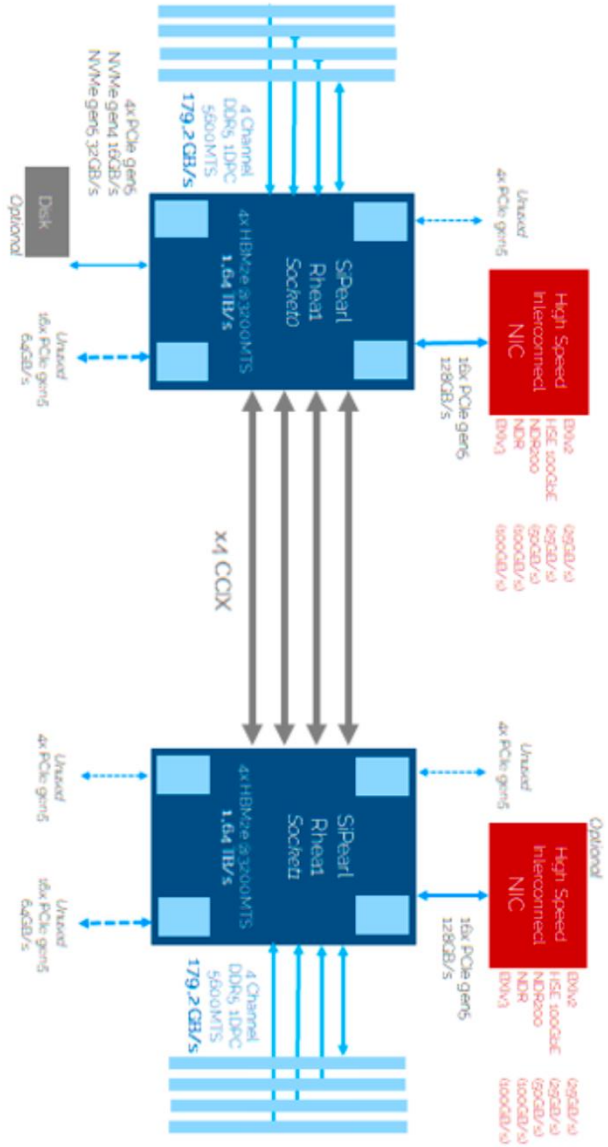
JUPITER – CLUSTER COMPUTE NODE ARCHITECTURE

Node Specs

- 2× SiPearl Rhea1
- 1× NVIDIA InfiniBand NDR200
- 512 GB DDR5 (36 nodes with 1024 GB)
- CCIX

CPU Specs

- ARM Neoverse Zeus
  - 2 x 256 SVE per core
- 2.5 GHz (~3.0 GHz Turbo)
- 64 GB HBM2e per Socket
  - 1.64 TB/s
- 256 GB DDR5
- PCIe Gen5
- 400W (estimation)



JUPITER – INTERCONNECT

One Network to Rule Them All

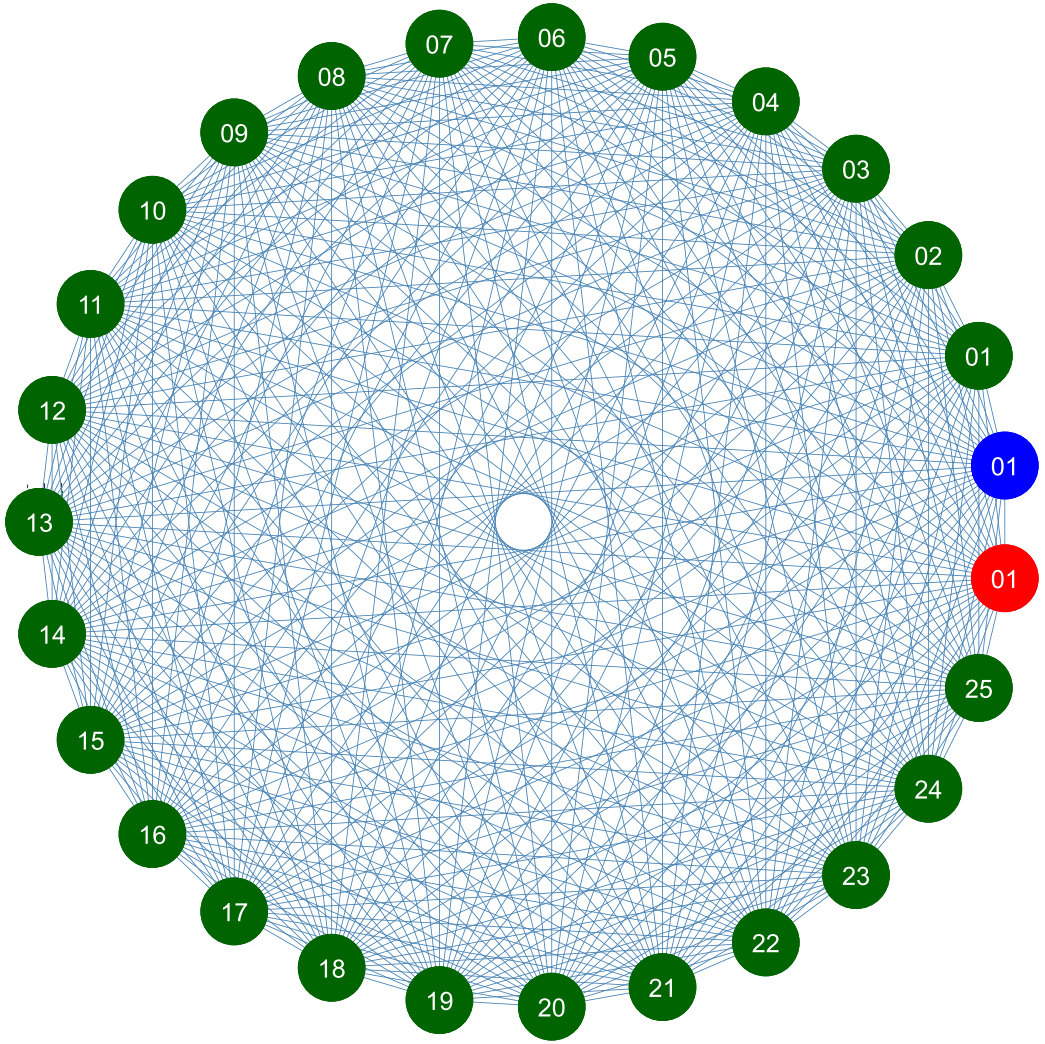
- NVIDIA Mellanox InfiniBand NDR/NDR200
  - NVIDIA Quantum-2 switches
  - NVIDIA Connect-X7 HCAs
- Dragonfly+ topology
  - 27 Dragonfly groups
  - Within each group: full fat tree
- 51000 links, 102000 logical ports, 25400 endpoints, 867 switches
- Adaptive Routing
- In-network processing on switch level (SHARPv3), tentatively

JUPITER – INTERCONNECT

One Network to Rule Them All

EVIDEN
an atos business

NVIDIA



JÜLICH
Forschungszentrum

JOINING FORCES

Ministerium für
Kultur und Wissenschaft
des Landes Nordrhein-Westfalen

fz-juelich.de/jupiter