The EuroHPC pre-exascale supercomputer of the North

Dr. Kimmo Koski, 30 June 2020

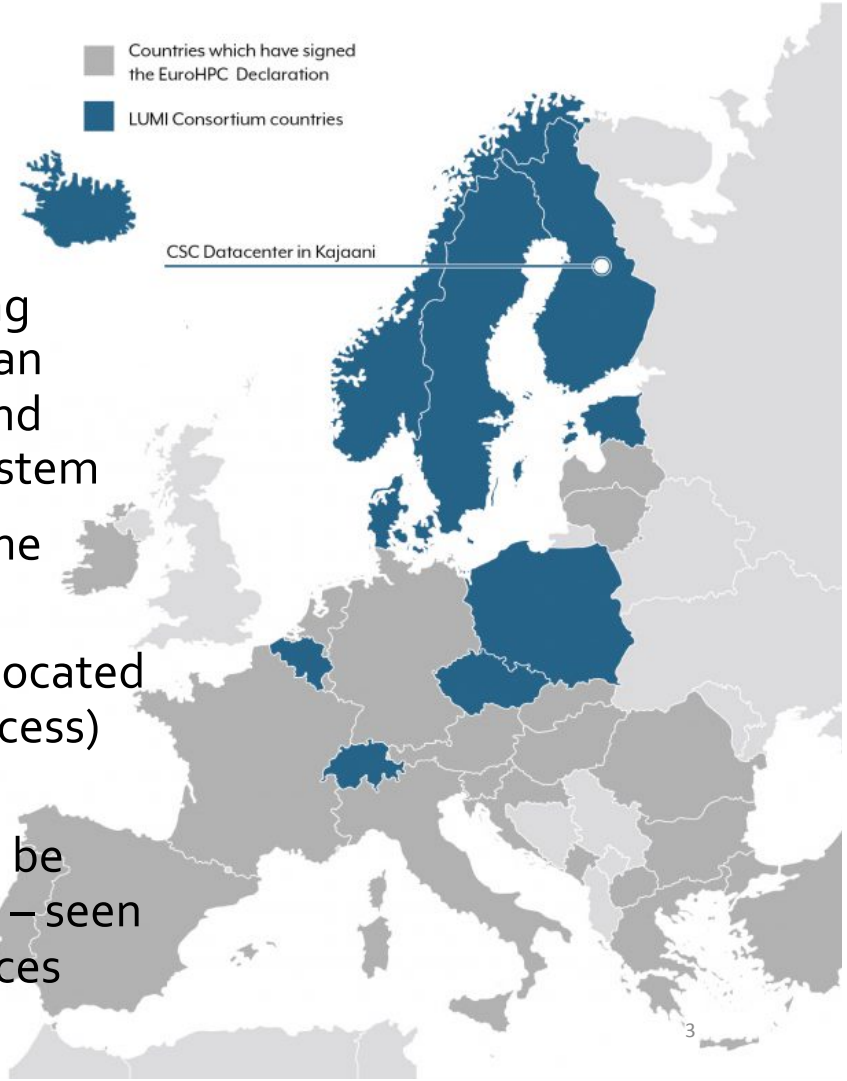
LUMI

The EuroHPC initiative

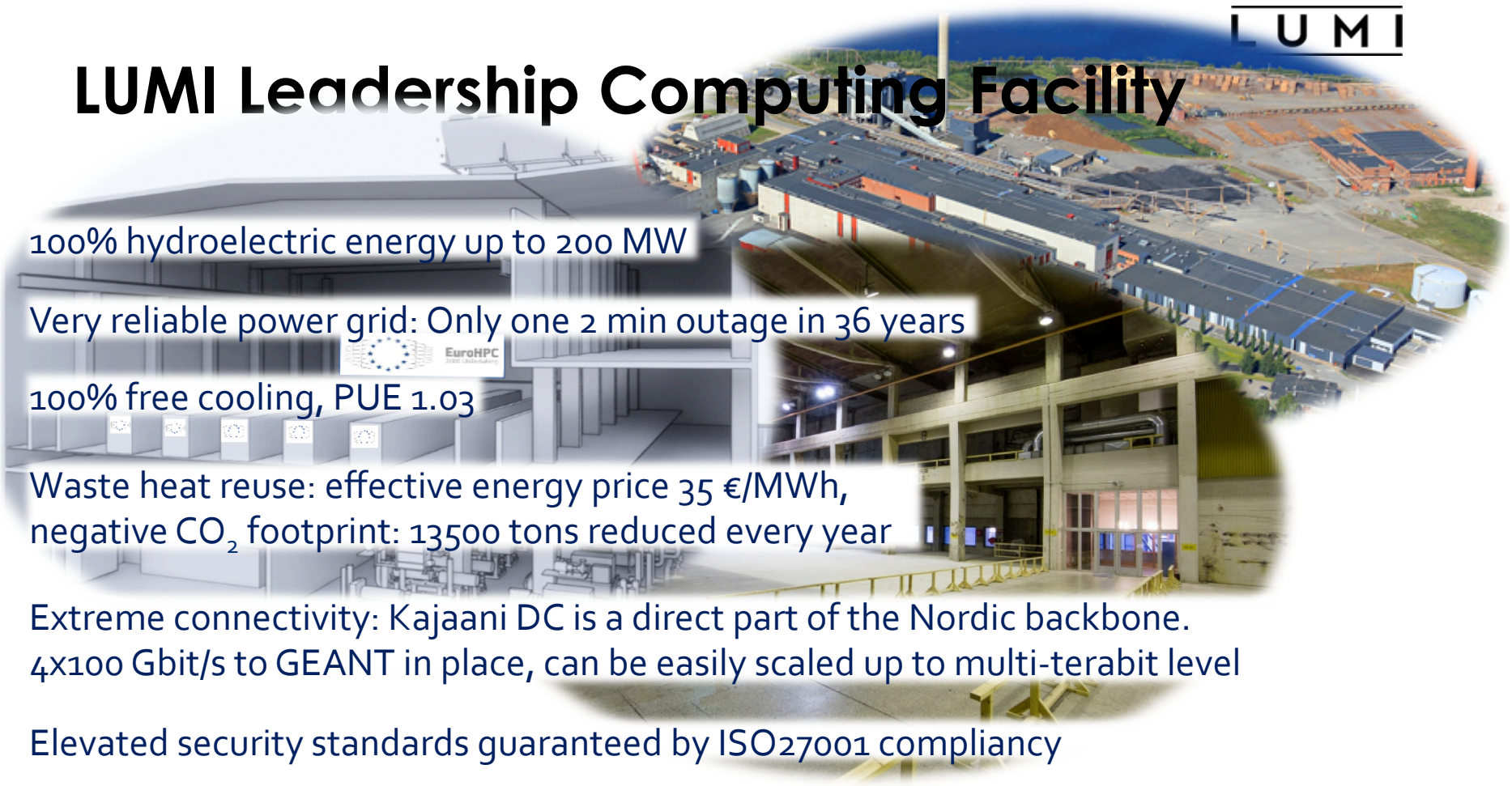
- The **EuroHPC Joint Undertaking** will pool EU and national resources in high-performance computing (HPC)
  - **acquiring and providing a world-class supercomputing and data infrastructure** for Europe's scientific, industrial and public users
  - supporting an ambitious **research and innovation agenda**
- The EuroHPC declaration has been signed by **32 European countries**
- The first generation of EuroHPC systems announced in June 2019
  - 3 pre-exascale systems (150+ Pflop/s) to Finland, Italy and Spain
  - 5 petascale systems (4+ Pflop/s) to Czech Republic, Bulgaria, Luxembourg, Portugal and Slovenia

LUMI consortium

- Unique consortium of 10 countries with strong national HPC centers and competence gives an unique opportunity for knowledge transfer and sharing and providing user support for the system
- The resources of LUMI will be allocated per the investments
- The share of the EuroHPC JU (50%) will be allocated by a peer-review process (cf. PRACE Tier-0 access) and available for all European researchers
- The shares of the LUMI partner countries will be allocated by local considerations and policies – seen and handled as extensions to national resources



LUMI Leadership Computing Facility



100% hydroelectric energy up to 200 MW

Very reliable power grid: Only one 2 min outage in 36 years

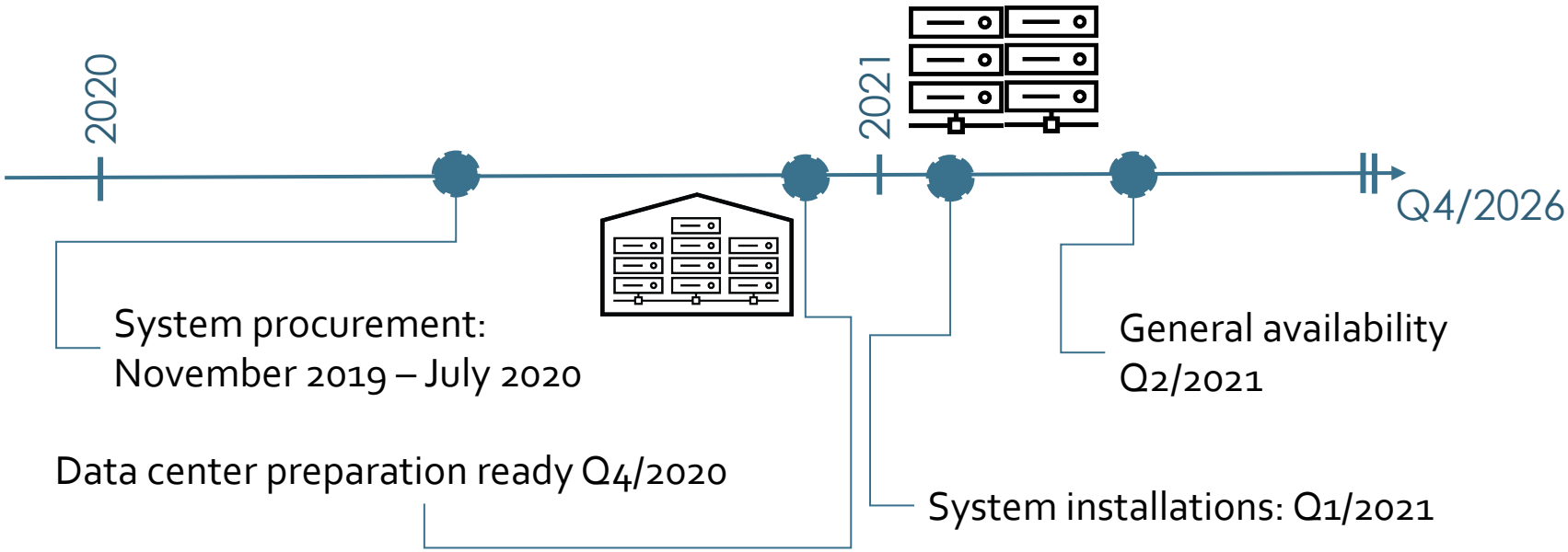
100% free cooling, PUE 1.03

Waste heat reuse: effective energy price 35 €/MWh,
negative CO_2 footprint: 13500 tons reduced every year

Extreme connectivity: Kajaani DC is a direct part of the Nordic backbone.
4x100 Gbit/s to GEANT in place, can be easily scaled up to multi-terabit level

Elevated security standards guaranteed by ISO27001 compliancy

LUMI Timeline



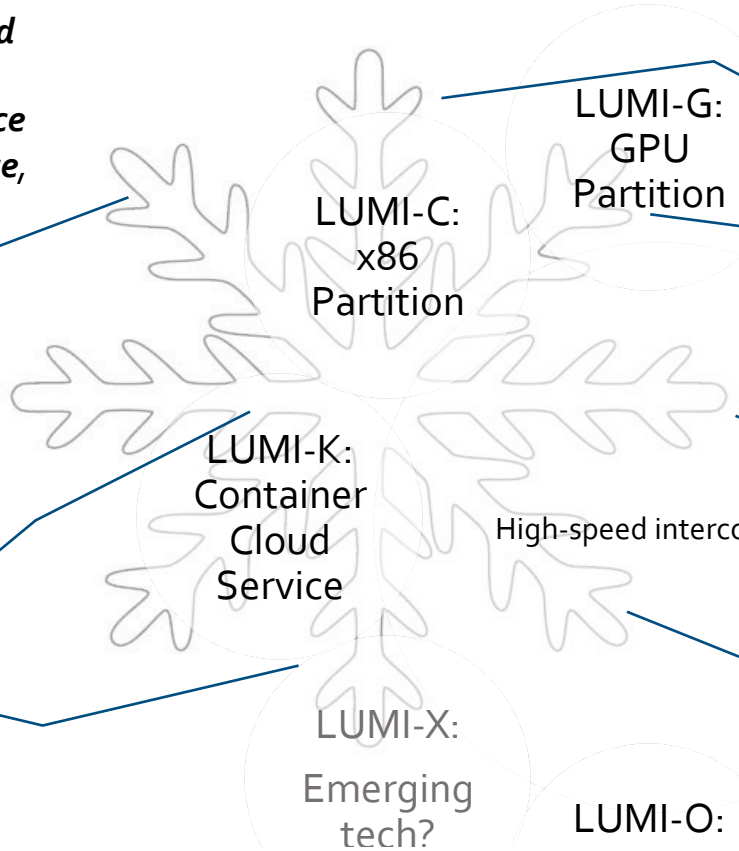
LUMI system architecture

LUMI is a Tier-0 ***GPU-accelerated supercomputer*** *that enables the convergence of* ***high-performance computing***, ***artificial intelligence***, *and* ***high-performance data analytics.***

- Supplementary "Tier-1" CPU partition
- M, L and XL memory nodes

Possibility for combining different resources within a single run

Encrypted object storage (Ceph) for storing, sharing and staging data



Tier-0 GPU partition

LUMI-D: Data Analytics Partition

Interactive partition with very large shared memory and graphics GPUs for data analytics and visualization

LUMI-F: Accelerated Storage

Flash-based storage layer with extreme I/O bandwidth and IOPS capability

LUMI-P: Lustre Storage

Large parallel storage

LUMI

Enhanced user experience

- In addition to traditional command-line interface, we wish to support high-level interfaces on LUMI, i.e. seamlessly integrate Jupyter Notebooks, Rstudio and similar to back-end to LUMI
- A rich stack of pre-installed software, community and commercial both
- Datasets as a Service: curated large reference datasets available and maintained

LUMI user support

- LUMI user support and a centralized help-desk by the distributed LUMI User Support Team
  - The model is based on a network of dedicated LUMI experts: each LUMI partner will provide one full-time person for the task
  - User Support Team will also provide end-user training, maintain the software portfolio and user documentation of the system
- ”Level 3” support (e.g. application enbling, methodology support) via national efforts

Resource allocation (tentative)

- LUMI resources are allocated in terms of GPU-hours, CPU-core-hours and storage hours
  - Each project applies and gets a combination of this
  - No dedicated hardware - all users can access the whole system within the batch job policies
  - All countries receive shares of these pools per their share of the TCO
- Access provided by the LUMI Consortium countries according to local policies, for the EC's share PRACE or a similar mechanism will be used
- Resources brokered in terms of
  - Preparatory access projects (XS) – single-PI
  - Development access projects (S) – single-PI
  - General access (Tier-1) projects (M) – single-PI
  - Extreme scale (Tier-0) projects (L) – single-PI, should be mostly GPU hours
  - Community access projects (XL) – multi-PI, multi-year

Plans for batch job queue policies

Queue	Max Wall-Time (hh:mm)	Max Resources	Remarks
test	00:30	2% GPUs / 2% CPU cores	Testing and debugging queue. Max 1 job per user, high priority
short	06:00	2% GPUs / 2% cores	Small jobs for backfilling
default	48:00	50% GPUs / 50% cores	Standard queue for production work
longrun	168:00	2% GPUs / 2% cores	Maximum one long job per user, lower priority
large	24:00	99% GPUs / 99% cores	By special arrangement only. One job per user at a time. Draining of the system for large queue jobs during weekends.
interactive	06:00	4 GPUs / 128 cores	for interactive batch jobs, max 1 job per user
serial	06:00	1 GPU / 64 cores	Number of jobs per user limited

How to prepare for LUMI?

- Thinking projects and use cases for LUMI
  - Cases for Tier-0 grand challenges
  - Combining simulation and AI methods within the same workflow
- There is a vast pool of GPU-enabled community codes
  - See if your favorite software suite already has been enabled, and if not, consider moving to a competing package that is
- Modernizing applications and GPU-enabling them
  - "even if it works, fix it"
  - CUDA/HIP, OpenACC or OpenMP5, or high-level libraries and frameworks

Concluding remarks

- **Unprecendent amount of computational resources and capabilities** available for universities and research institutions in the LUMI consortium countries
  - Complemented by competence building and user support activities
- **LUMI, the Queen of the North**: leadership-class resource designed for a broad range of user communities and workloads, with an enhanced user experience
  - **LUMI will be a GPU system,** which needs some preparatory work – but it will be a robust production system, and not experimental or esoteric in any manner
- **Modernizing HPC applications** for harnessing the largest systems is not trivial, and needs a lot of focused effort – but it will pay off
  - It is time already to start preparing for the LUMI era