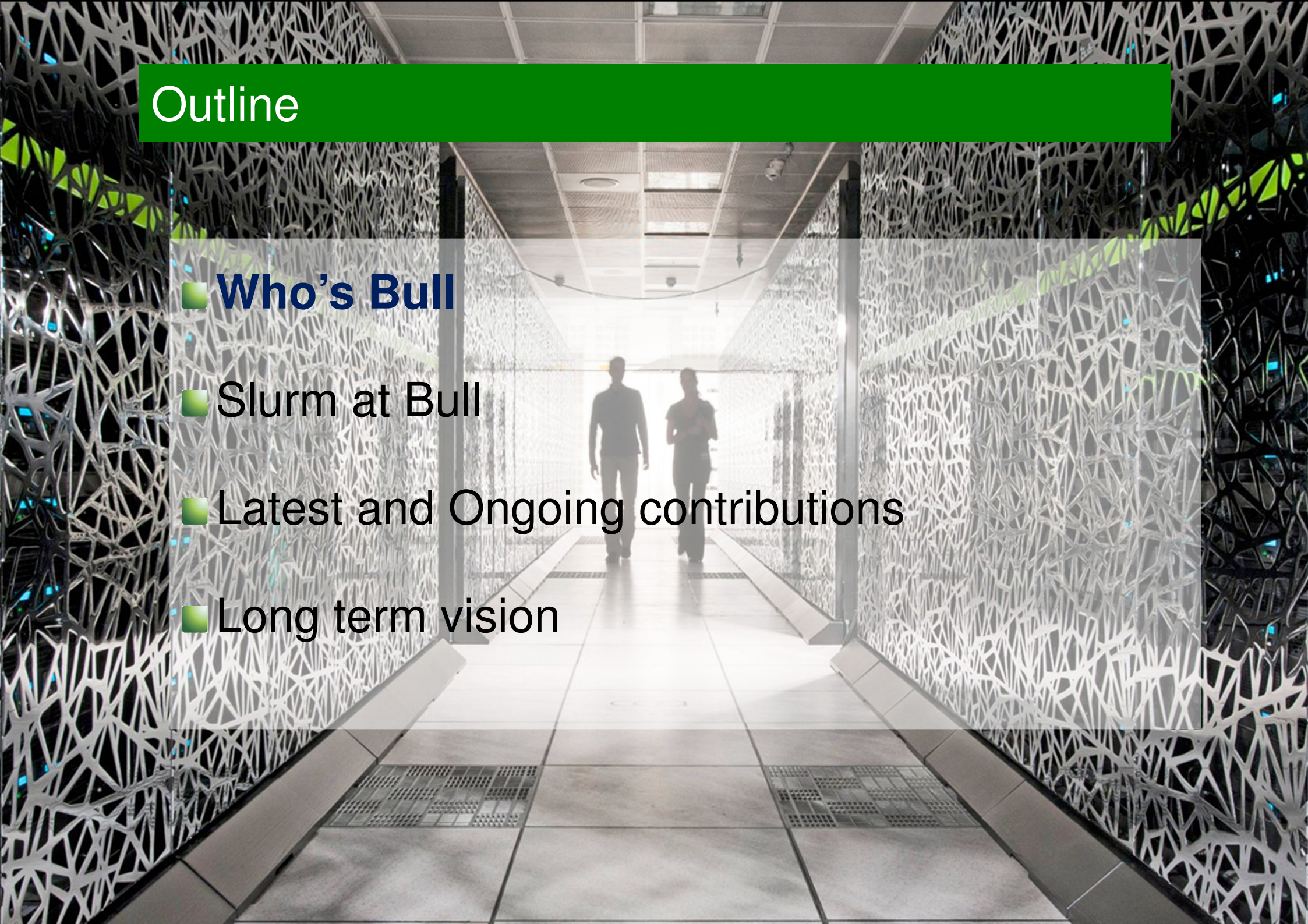
BULL's Slurm Roadmap

Eric Monchalin, Head of Extreme Computing R&D



Outline

- **Who's Bull**
- Slurm at Bull
- Latest and Ongoing contributions
- Long term vision

Who’s Bull?

Who's Bull?

Visit us

Booth #2643 on floor 4

Bull Extreme Computing: a complete offering

What if unlimited innovation were as simple as this?

bullx supercomputers

extreme factory HPC Cloud

mobull container

Meeting customer requirements

Integrated

- Installed, deployed
- and operated as a
- single software

Complete

All functionalities
- Cluster management
- Development factory
- Execution environment
- Data storage and access

All sizes
- From department to Top 5

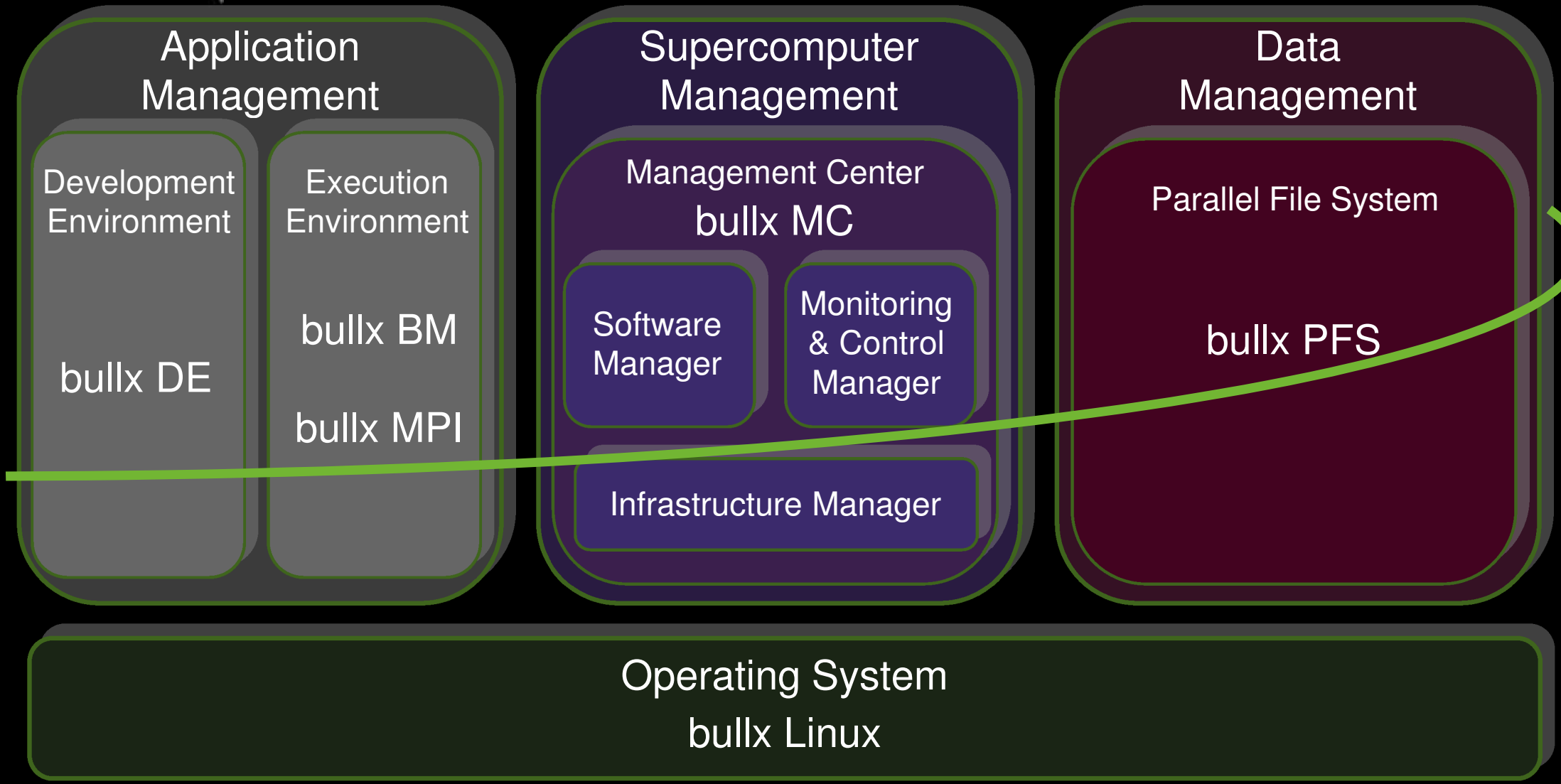
Open

- Best of breed
- Linux, OpenMPI, HPC Toolkit, Nagios, OFED, Slurm, Lustre, Shine, ...
- Bull added value

Flexible

- Modular:
  Get what you need
- when you need

bullx Supercomputer suite modularity

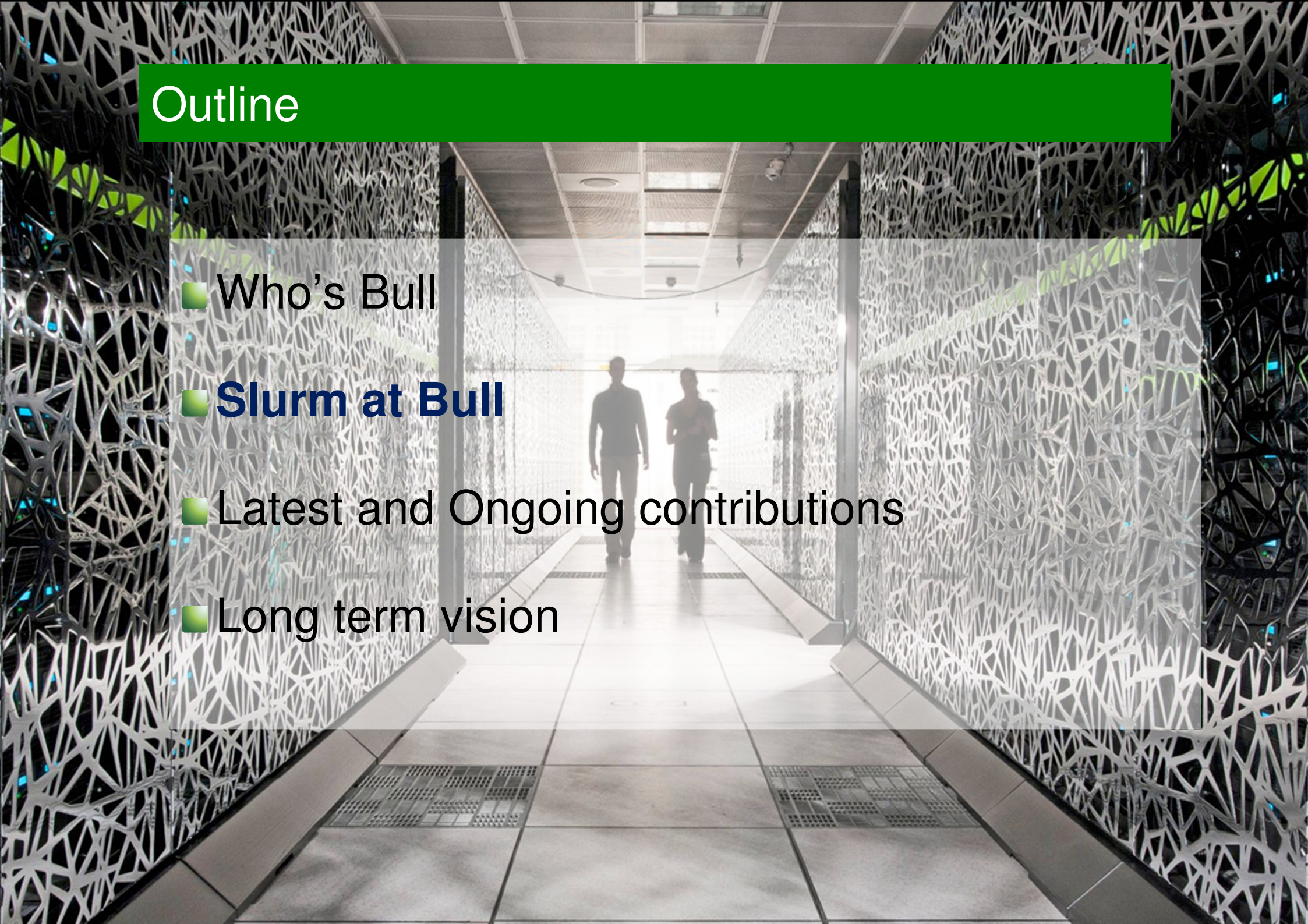


Slurm delivered with bullx Batch Manager

- SLURM integrated into Bullx Super Computer Suite offer since 2006

bullx BM
&
Extended
Offer

- ✓ Enhanced support and active development of SLURM
- ✓ Integration and support of commercial products: LSF & PBSPro



Outline

- Who's Bull
- **Slurm at Bull**
- Latest and Ongoing contributions
- Long term vision

BULL involvement in Slurm community

- BULL initially started to work with SLURM in 2005
  - Development of new features
  - Bugs fixing
  - Tutorials and Trainings
- Collaborations with INRIA, CEA, SchedMD, …
- BULL sponsored and organized with SchedMD the 2nd SLURM User Group Conference
  - User, Admin Tutorials
  - Technical presentation for developers

Largest BULL clusters powered by Slurm

powered by Bull

cea

PRACE

FUSION FOR ENERGY

PetaFlops	1.25	1.7	1.3
# cores	140,000	90,000	80,000
Memory (TB)	256	360	300
Storage (PB)	300	10	60

Outline

- Who's Bull
- Slurm at Bull
- **Latest and Ongoing contributions**
- Long term vision

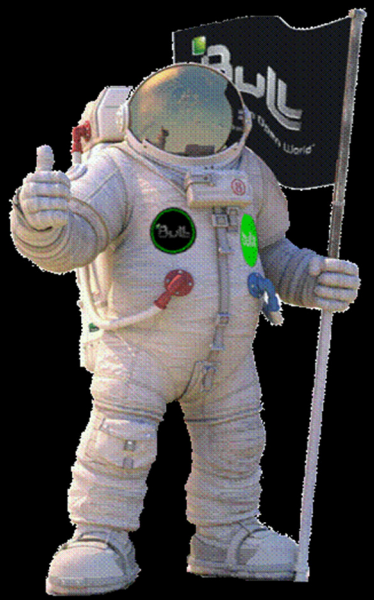
Latest contributions

- Fine grain ressource Management
  - cgroups support (CEA-BULL)
  - CPU Management enhancements and documentation
- Performance
  - Scalability / high throughput optimizations (CEA-BULL)
  - Preemption improvements (grace time delay)
- Cluster Integration / Utilization
  - Sview graphical tool enhancements
  - High Availability and event handling

Directions

From CPUs to Many Cores
(Scalability, Robustness, Resource Mngt)

Power Management



HPC On Demand Cloud Computing

From CPUs to Many cores infrastructure

- Fine grain resource Management
  - Extension of cgroups support (BULL-CEA-LLNL)
  - Multi-parameter / Multi-objective Scheduling (BULL-INRIA)
- Performance Optimizations (BULL-CEA-INRIA)
  - Whatever the cluster size
  - Whatever the number of jobs
- Resources Selection/Allocation Improvements
  - Extension of CPU selection and allocation algorithms to support NUMA hierarchy

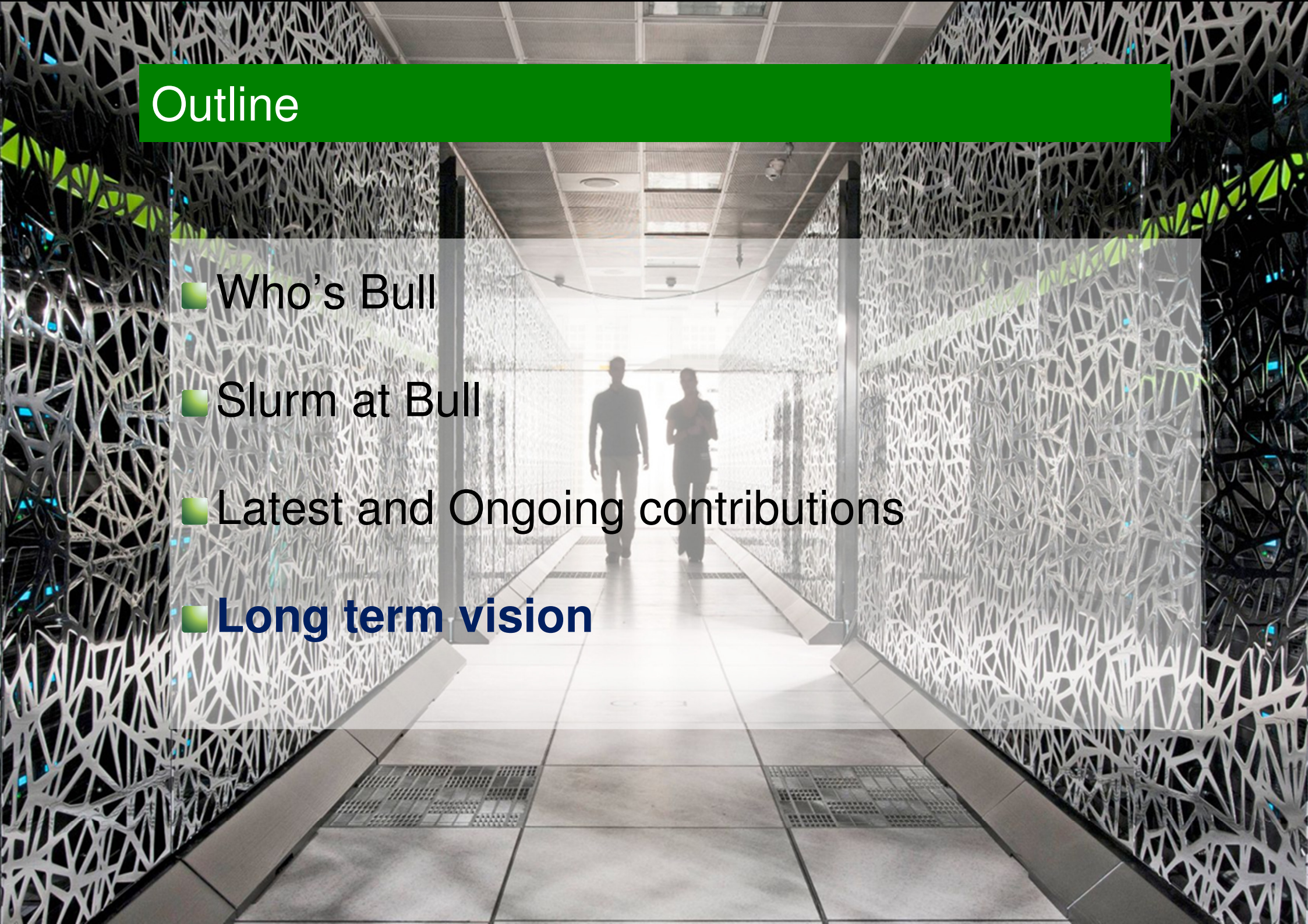
Power management

- Power Management Integration (BULL-SchedMD)
  - Calculation of power consumption per job
    - Either based on power sensors (node, switch, rack,etc)
    - Or according to CPU utilization (cycles and frequencies)
- Energy Efficient Scheduling
  - Scheduling according to jobs' energy consumption needs and clusters' power states and thresholds

HPC on Demand / Cloud computing

extreme factory

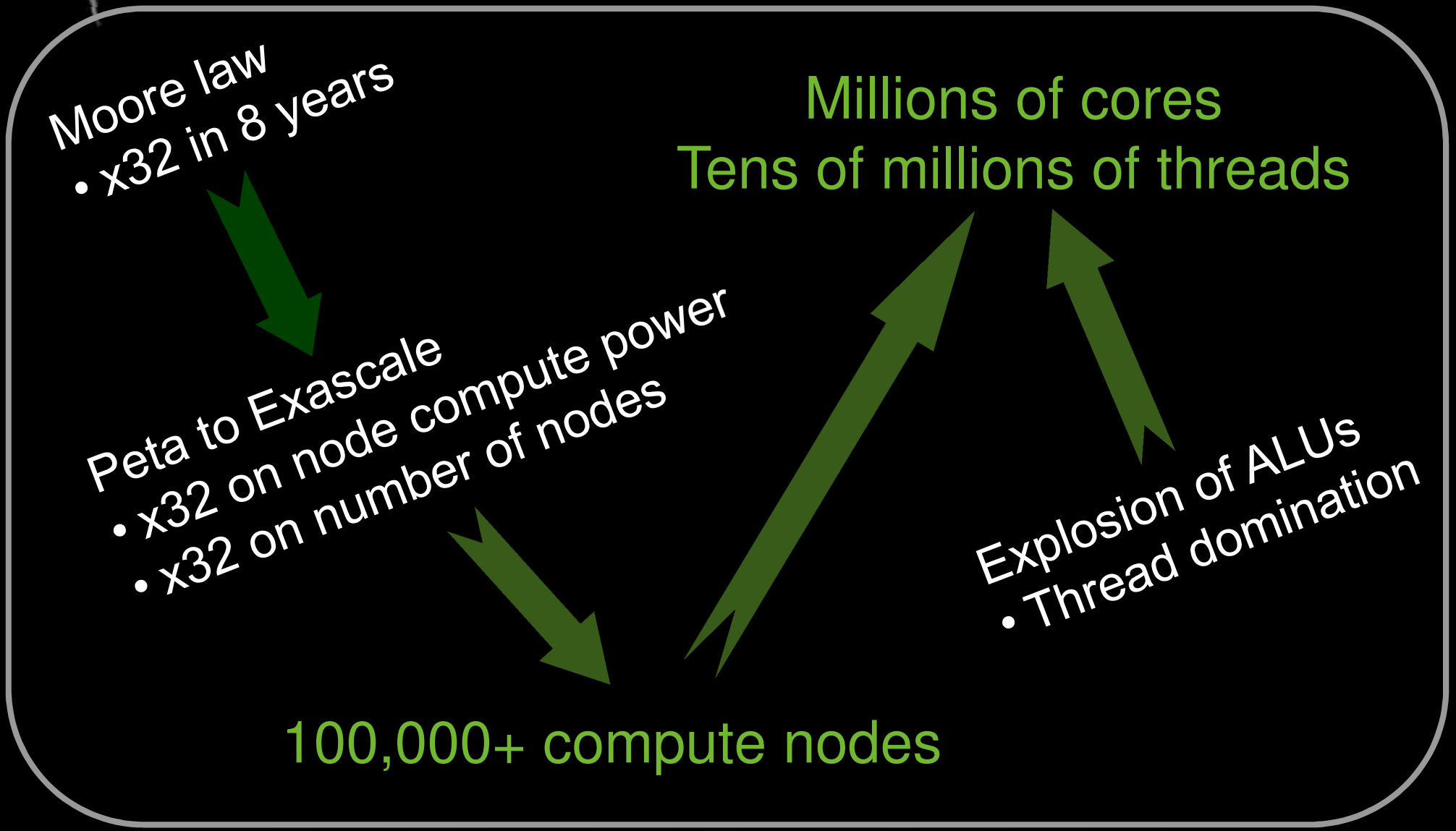
- HPC on Demand
  - SLURM Integration upon BULL's extreme Factory HPC on Demand solution
  - DRMAA API v2 upon SLURM



Outline

- Who's Bull
- Slurm at Bull
- Latest and Ongoing contributions
- **Long term vision**

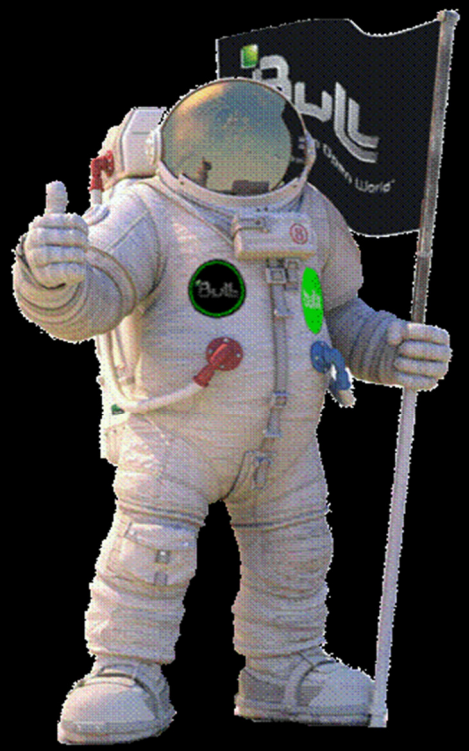
Exaflop era: explosion of resources



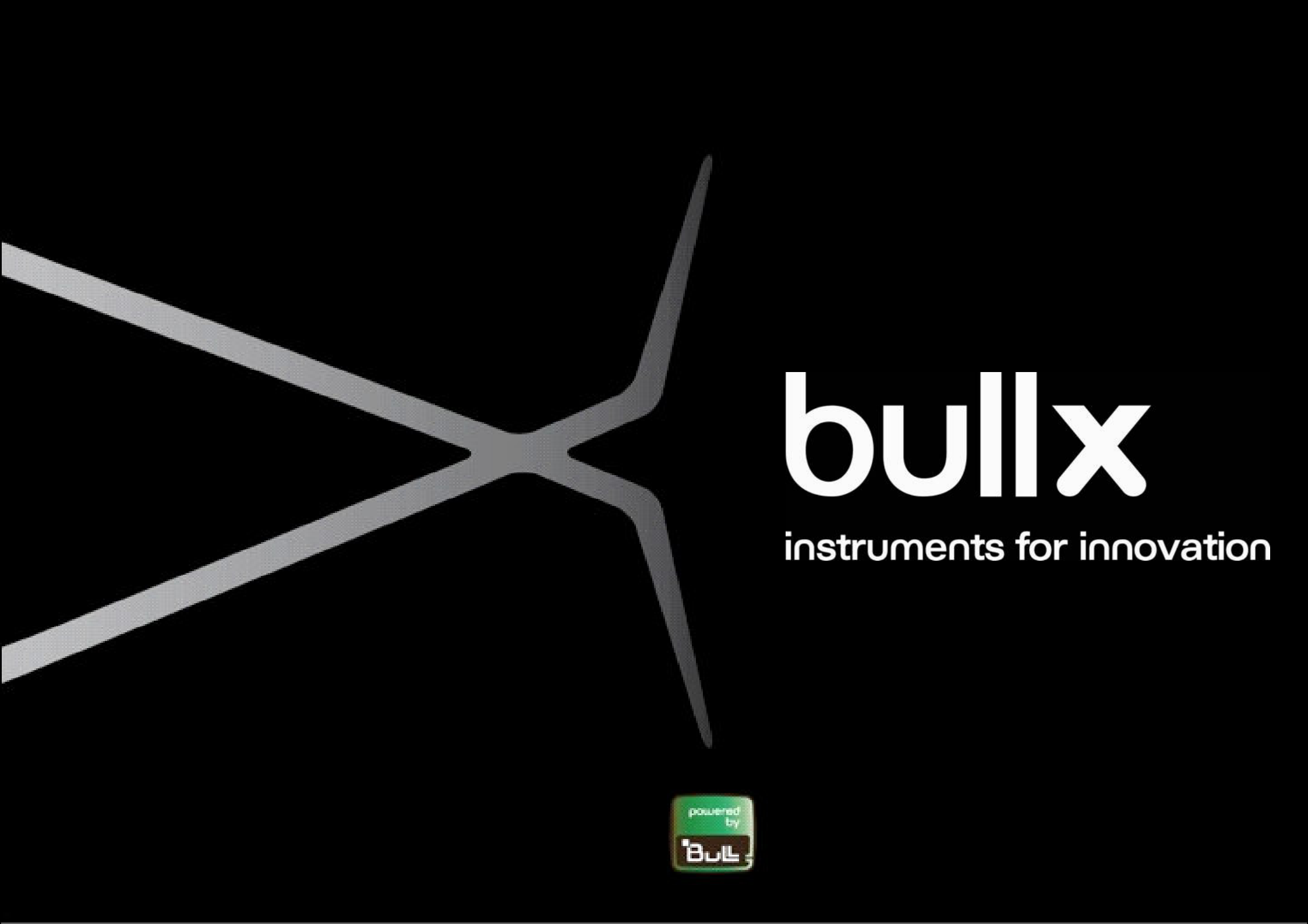
Offer new services to applications

- Optimize compute environment
  - Describe key characteristics of applications
  - Elect the most appropriate set of nodes
  - Manage resources with heuristics predicting future workload
- Migrate Processes
  - To reduce resource fragmentation
  - To isolate nodes with predicted hardware failures
- Allow dynamic application frameworks
  - To balance the load of the application
  - To optimize refinement of meshes
  - To restart lost processes in case of failure

First steps on the Exaflop the road



- Much more numerous
  - Scalability improvements
  - Elastic jobs to a better efficiency
- Much more heterogeneous
  - Re-design of SLURM's core algorithms for resources selection and allocation
  - Management of network and I/O bandwidth resources


bullx
instruments for innovation
powered by
Bull