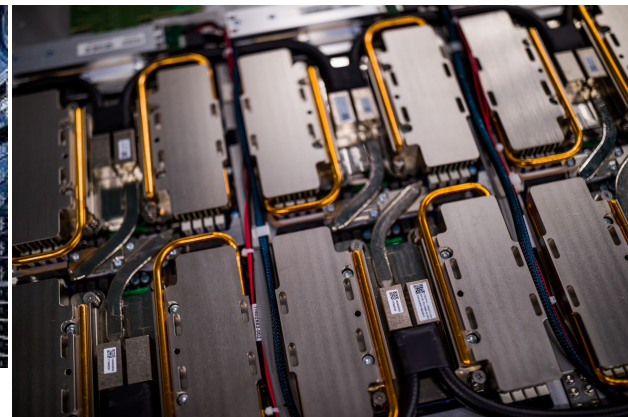


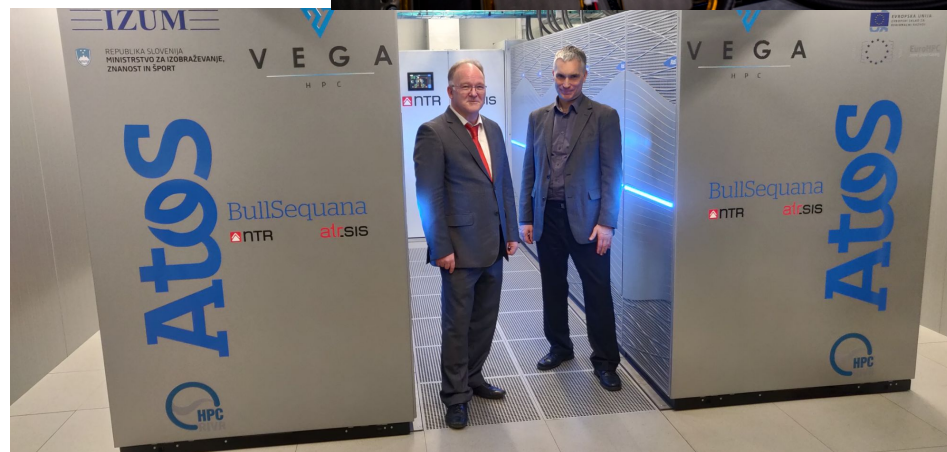
Vega

Andrej Filipcic

Vega/EuroHPC - IZUM, Maribor, Slovenia



Number of Nodes	1020
Number of login nodes	4 CPU login nodes and 4 GPU login nodes
Compute partitions	CPU partition: 960 (768 standard CPU nodes and 192 high memory CPU nodes * GPU partition: 60 nodes
Storage capacity	1 PB high-performance NVMe storage, 23 PB raw large-capacity storage (18 PB usable)
Sustained Performance	6,9 petaflops
Peak Performance	10.1 petaflops



Storage pools

- 3 servers, 4x250TB RBD each on ceph 16+3 EC
 - Read is fine, >6GB/s per server
 - Write is not good due to XFS fragmentation, high IO wait - 32MB ceph objects with 4k xfs chunks...
 - Write back not working on kRBDs
- Current activity:
 - Proxmox 3 servers, 12 VMs (8-core, 64GB), one pool/VM
 - dCache pool on CephFS
 - >2GB/s write/read per VM
 - Min_free and dirty adjustments needed to prevent out-of-memory for writeback cache (crash)
 - ~4 days to migrate, 1PB copied

Further testing

- VMs with current RBDs
 - librbd + write back cache
 - Xfs optimizations
- VMs with real ceph pools (file==rbd)
 - Worth trying? NO
- S3?

ARC & Computing

- Improvements in ARC-CE handling
 - >200Hz file transfer rate
 - Arex throughput increased by factor of 200 in processing rate (fork -> vfork)
- Cache moved to lustre (nvme), ceph hdd not fast enough
 - Will try cephfs on nvmes in near future
- 1/3 of ATLAS production with all workloads
 - 1st tuned workloads - 16-core event generation , 4.2B events, grid 10%, Vega 90% - too demanding for average grid sites
- Input staging from cern (added as storage element)
 - ~60Gb/s download rate on average during 2 days testing
 - Suitable for T0 data processing