

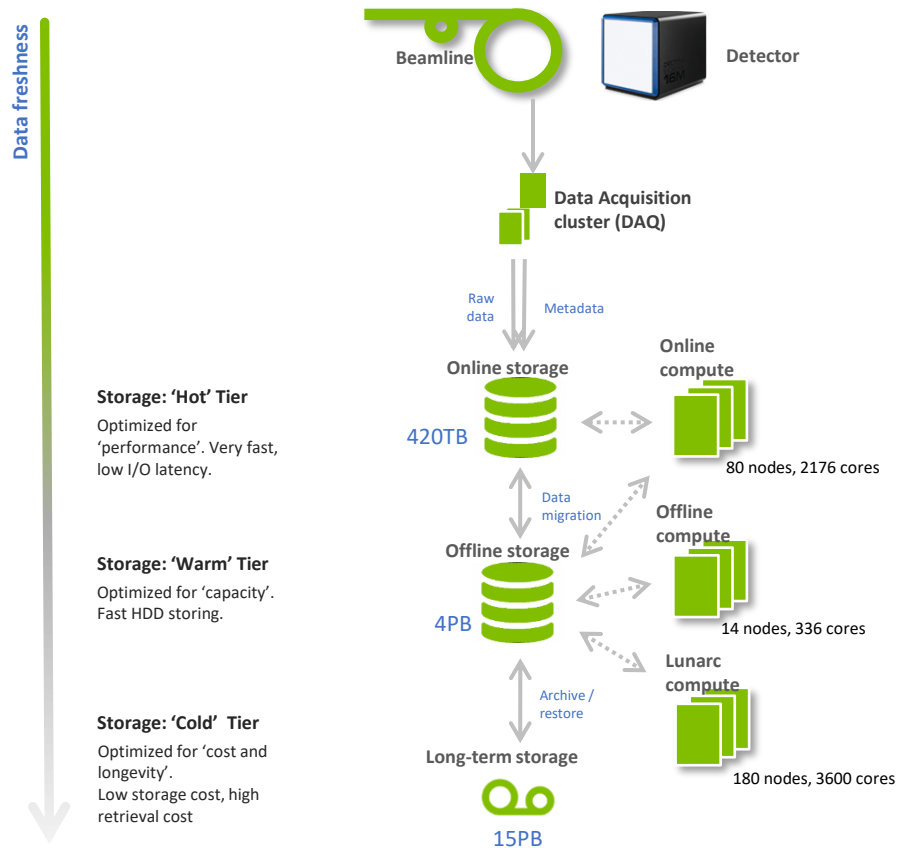


Infrastructure for data flow

Thomas Eriksson

IT Infrastructure Team

Data Flow - Overview



Storage

- IBM Elastic Storage Server GS4S and GL6S
- Spectrum Scale (GPFS)
- Hierarchical Storage Management using 3 tiers
 - Flash 420GB
 - Spinning disk 4PB
 - Tape 15PB
- Better than 33GB/s read, 20GB/s write
- Infiniband connectivity to DAQ and HPC via 16 HBAs

Network

- Two data centres with top level core switches forming a redundant cluster with access switches configured in a spine-leaf architecture.
- Dedicated networks for machine and experiment control, data acquisition and office network.
- Single mode fibres from beamline to data centres (12 pairs per BL and data centre)
- Data acquisition network can use up to 100GE for individual connections
- Infiniband network for all storage, HPC and DAQ with up to EDR 100 speed

Data Acquisition and Detectors

- Wide range of detectors
 - 128k pixels to 16M pixels
 - 1Hz to 2-3kHz
 - Burst readout to continuous readout
 - Normally not the worst case for all three at the same time
- Cluster of DAQ machines for detector readout
 - Optionally apply some inline data manipulation
 - Write to storage
 - Stream data to HPC cluster for further analysis
- Up to 100GE link on ingest and up to 100GB EDR link for storage
- Customized to each detector and experimental technique
- Choice of deployment; bare metal, docker and Kubernetes.
- FPGA and GPU can optionally be utilized for inline analysis/reduction

MAXIN

The image shows the word "MAXIN" in a bold, grey, sans-serif font. A bright yellow, curved swoosh or underline is positioned over the letters, starting under the 'M', looping under the 'A' and 'X', and ending under the 'I'.