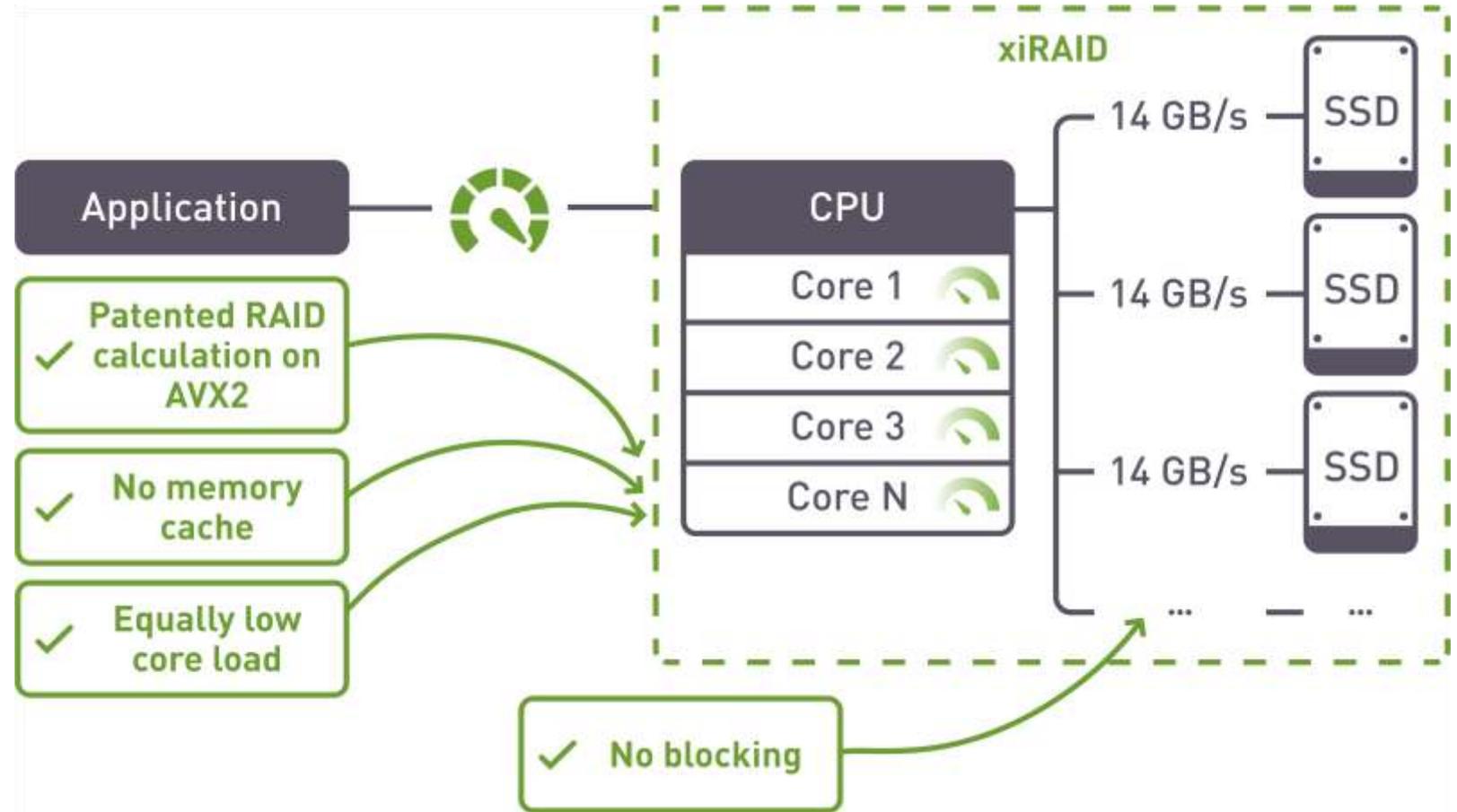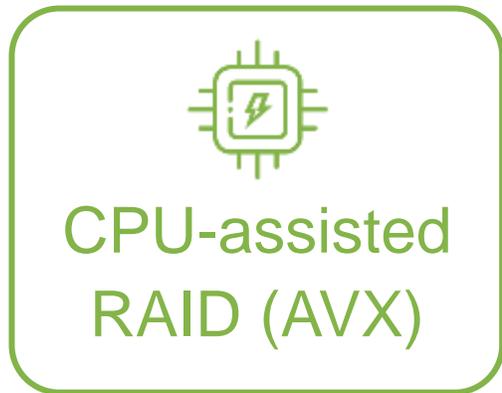# xiRAID Case Study at Aalen University

IBM Scale User Group SC24

Davide Villa – CRO – davide.villa@xinnor.io

# xiRAID Classic

CPU-assisted RAID (AVX)

Lockless data path

Application

Patented RAID calculation on AVX2 ✓

No memory cache ✓

Equally low core load ✓

No blocking ✓

xiRAID

CPU
- Core 1
- Core 2
- Core 3
- Core N

14 GB/s — SSD

14 GB/s — SSD

14 GB/s — SSD

...

# What makes xiRAID unique

- High performance both in **normal** and **degraded** mode

- Flexibility: any drive, any capacity, RAID level 0,1,5,6,7.3,50,60,70 and N+M supported, up to 64 drives in a single RAID, unlimited number of RAID

- Features: RAID Migration, Restriping, Hot Spare, Notification, Merge, Write-hole protection, Variable strip-size, CPU affinity, High Availability

- No PCIe "slot tax", No extra hardware, No extra power consumption

- Leverage existing system resources with minimal additional load

- Fault-tracking system to handle network issues on disaggregated infrastructure

- It works with block storage, regardless of transport – local and networked

*Hochschule Aalen*

# Aalen University Case Study

High-performance storage for ML based on

- IBM Scale Data Access Edition

- xiRAID

- commodity hardware

deployed by

**ABC SYSTEMS AG**

APPLIED BRAINWARE & COMPUTER SYSTEMS

# Storage Requirements:

- High-Performance:
  - to saturate the network bandwidth to the client (Nvidia DGX and ML workstations)
- Data Availability and Integrity:
  - capable of surviving hardware failures
- Geographical Distribution:
  - to synchronize the data between the 2 sites of the campus for DR
- Scalability:
  - to accommodate future growth in data and compute resources.
- Cost efficiency

# Deployed Architecture
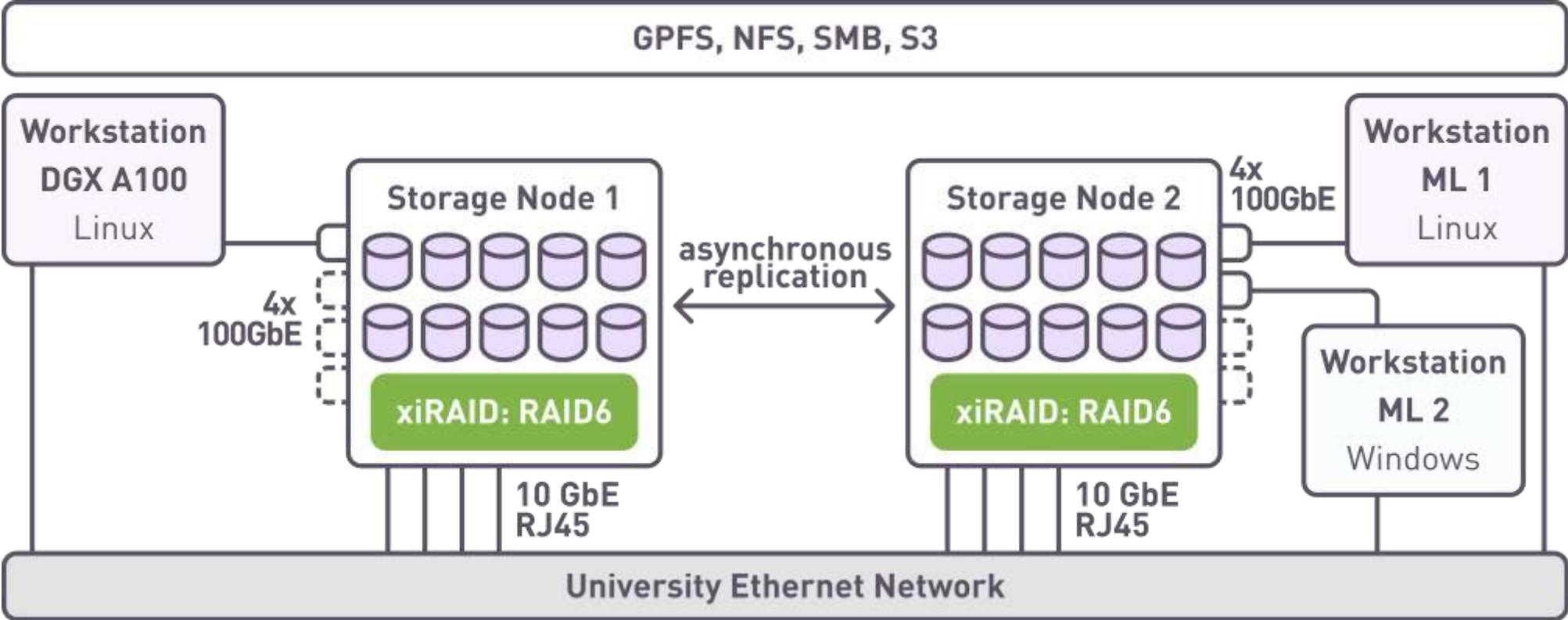
# RAID Options for Data Access Edition

☐ **IBM GPFS Native RAID (GNR).** Only available on proprietary hardware.

☐ **IBM Erasure Code Edition (ECE).** Minimum 4 NVMe storage nodes, with 6 being preferrable.

Third-party RAID engine:

☐ **Hardware RAID** would have limited performance and wasted 16 precious PCIe lanes, forcing to add a switch (latency)

☑ **xiRAID Software RAID** was selected thanks to its superior performance

# Single Node Performance

| | Sequential Read | Sequential Write |
|---|---|---|
| Raw drive performance | 92 GB/s | 55 GB/s |
| Storage Scale + xiRAID | 80.2 GB/s | 42.8 GB/s |
| Overall efficiency | 87 % | 78 %* |

\* All 10 drives are considered when calculating the efficiency

- xiRAID delivers throughput speeds very close to the raw performance of the drives.
- The solution saturates the network.

```
GPFS FS with gpfsperf:
```
/usr/lpp/mmfs/samples/perf/gpfsperf write  seq 120Gfile -n
120g -r 4m -th 10 -fsync -dio
  recSize 4M nBytes 120G fileSize 120G
  nProcesses 1 nThreadsPerProcess 10
  file cache flushed before test
  using direct I/O
  offsets accessed will cycle through the same file segment
  not using shared memory buffer
  not releasing byte-range token after open
  fsync at end of test
    Data rate was 42836998.24 Kbytes/sec, Op Rate was
10213.14 Ops/sec, Avg Latency was 0.977 milliseconds, thread
utilization 0.997, bytesTransferred 128849018880

/usr/lpp/mmfs/samples/perf/gpfsperf read   seq 120Gfile -n
120g -r 4m -th 10 -fsync -dio
  recSize 4M nBytes 120G fileSize 120G
  nProcesses 1 nThreadsPerProcess 10
  file cache flushed before test
  using direct I/O
  offsets accessed will cycle through the same file segment
  not using shared memory buffer
  not releasing byte-range token after open
    Data rate was 80189171.64 Kbytes/sec, Op Rate was
19118.59 Ops/sec, Avg Latency was 0.513 milliseconds, thread
utilization 0.980, bytesTransferred 128849018880
```

# Conclusions

- **Performance:**
  - fully saturating the Client's network bandwidths
- **Data Integrity and Availability:**
  - Xinnor's xiRAID RAID6 configuration provides protection for up to 2 drives failures per server,
  - IBM Storage Scale's asynchronous replication ensures high availability and disaster recovery.
- **Scalability:**
  - The solution is provisioned to handle more compute clients
  - Capacity and performance can scale easily by adding more drives per node or more storage servers.
- **Cost-Effectiveness:**
  - Xinnor software-based RAID allows to use the Data Access Edition of IBM Storage Scale on flexible hardware choices, minimizing the cost.



*Hochschule Aalen*

Prove it yourself:

https://xinnor.io/