DECTRIS

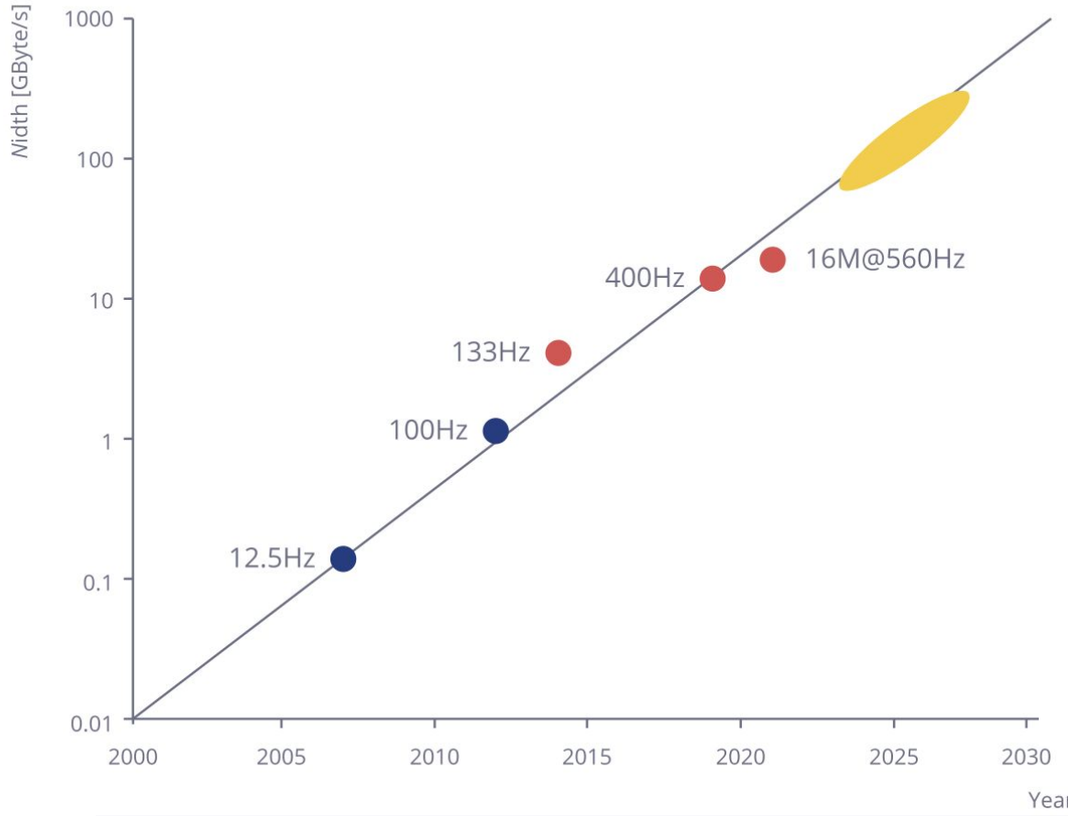
Addressing Big Data Challenges

Diego Gämperle, Product Owner
Camilla Larsen, Scientific Solution Architect
2024-03-19

Evolution of Data Rates

PILATUS

2007: 0.14 Gbyte/s of Raw Data



Prototype detectors in development (Jungfrau, Citius,)

EIGER2/PILATUS4

2022: 18 GByte/s of Raw Data

Moore's Law for X-Ray detectors

Detectors and Pipelines

Detector Head
e.g. EIGER2 XE 16M

160 Gbit/s

- Readout of Pixel Data
  (8 or 16 Bit UINT)
- Send/Route data to DCU

Detector Control Unit (DCU)

- Assembling images
- Applying corrections
- Adding image Metadata
- Compression of images
- Packaging into .h5

Data Interfaces

STREAM2
ZMQ: high BW, low latency

FILEWRITER
h5: high BW, medium latency

MONITOR
TIFF: low BW, medium latency

- Application specific processing

Data flow

What data reduction already exists?

Detector Control Unit (DCU)

Auto-Summation

- High count rates require summing up of images
- e.g.: counting 1E7 ph/s/pix with 16 Bit requires >150 Hz

e.g.: EIGER2: internal summation & correction of images

Threshold Difference Mode

- Output subtracted threshold images
- 1 instead of 2 images per exposure

Compression

- BSLZ4 compression on assembled images per default
- Compression ratios: 5-100

EIGER2: only achieves full specifications if images are compressed

Adaptive output

- Bit depth of output image depends on image
- Varies: 8, 16 or 32 Bit

EIGER2: Adaptive output image Bit-depth

BSLZ4 Compression

BSLZ4 compression established as default

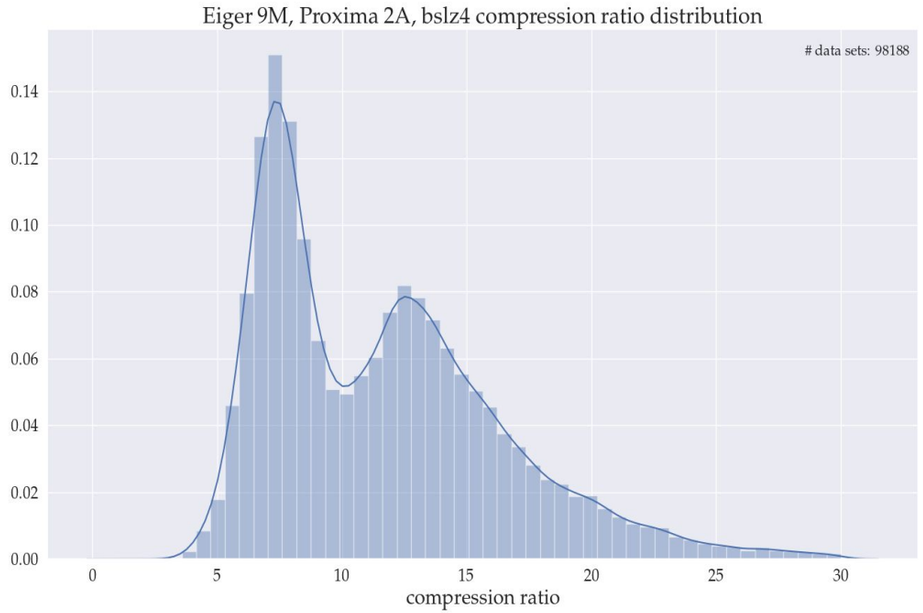
BitShuffle

- Originally implemented almost 10 years ago from Kiyoshi Masui

LZ4

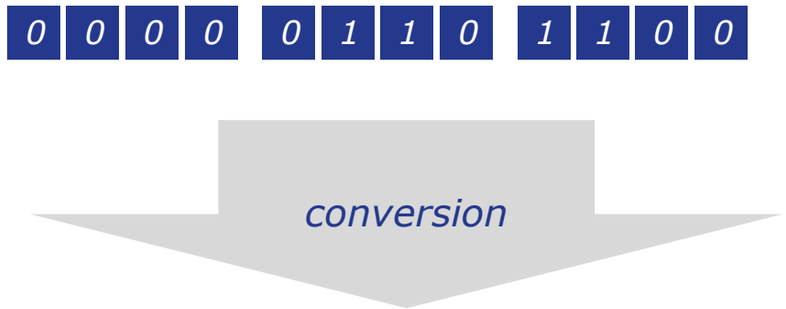
- Lossless data compression algorithm focused on compression & decompression speed.

=> optimized by and for DECTRIS specific data

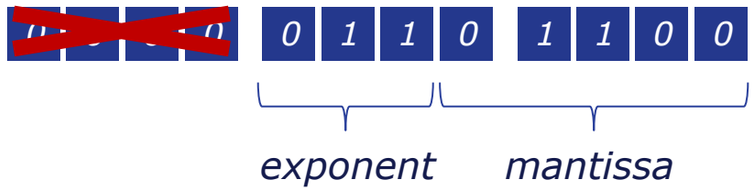


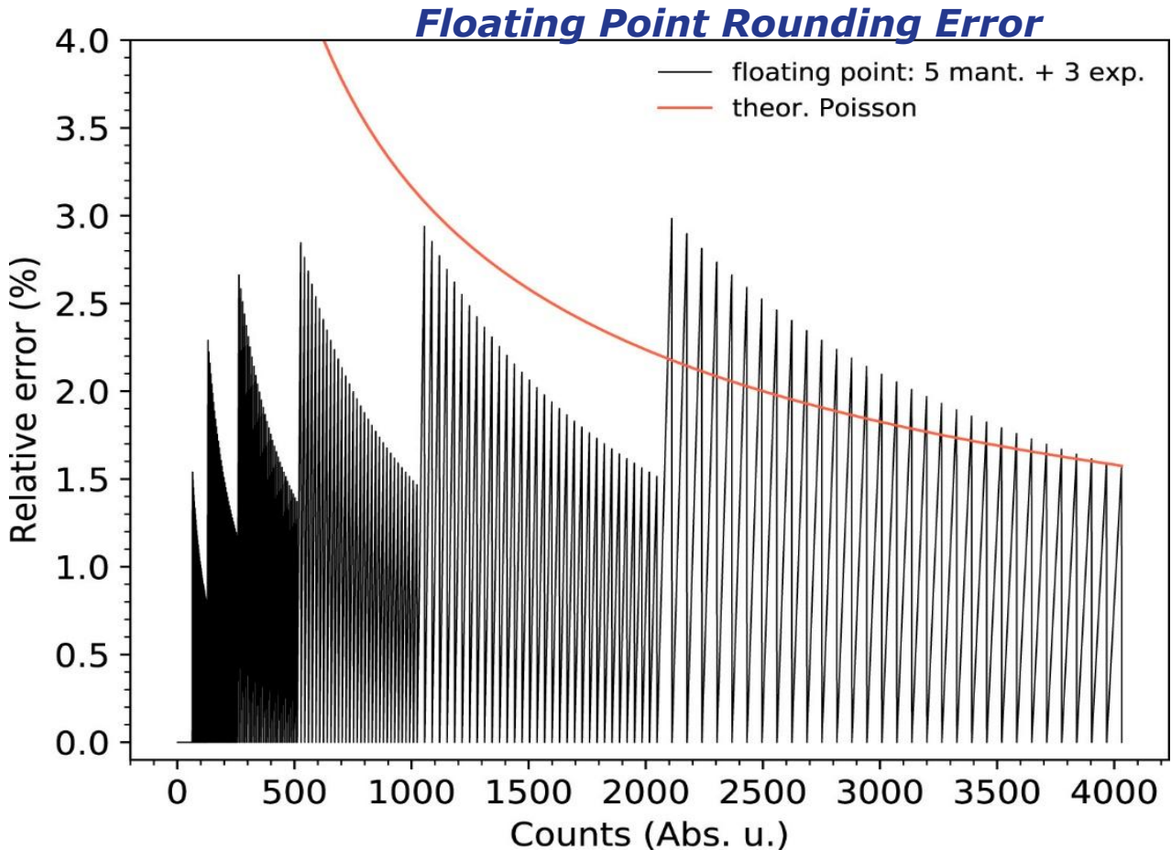
8-bit Floating-Point Compression

Standard 12 Bit UINT counter:



8 Bit Floating-Point counter:



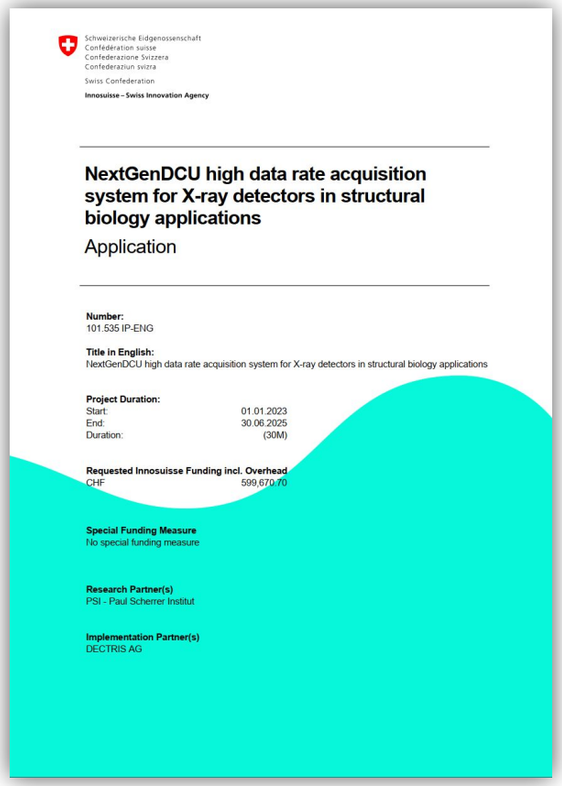


See P.Zambon, et.al. (2023), doi:10.1016/j.nima.2022.167888

“NextGenDCU” Project

Innosuisse

Research & Commercialization Project:



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

NextGenDCU high data rate acquisition system for X-ray detectors in structural biology applications

Application

Number:
101.535 IP-ENG

Title in English:
NextGenDCU high data rate acquisition system for X-ray detectors in structural biology applications

Project Duration:
Start: 01.01.2023
End: 30.06.2025
Duration: (30M)

Requested Innosuisse Funding incl. Overhead
CHF 599,670.70

Special Funding Measure
No special funding measure

Research Partner(s)
PSI - Paul Scherrer Institut

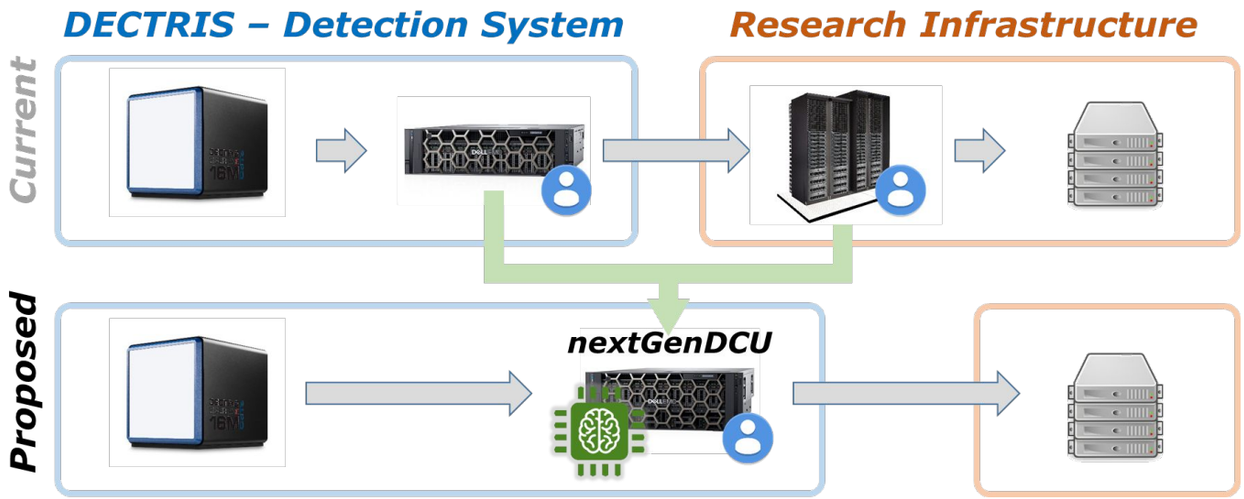
Implementation Partner(s)
DECTRIS AG

Filip Leonarski

Meitian Wang

PAUL SCHERRER INSTITUT

PSI



Community involvement

- DAPHNE4NFDI

- LEAPS-INNOV – work package 7

- Individual exchanges in the community
  - **HDRMX** - High Data-Rate Macromolecular Crystallography
  - **HDF** User Group Meetings
  - **NeXus/NXmx** discussions and implementation

Summary

- **Scientists always push the limit**
  Data bandwidth will always be used
  Data reduction upstream will lead to faster frame rates

- **Compression requires de-compression**
  FAIR principles are key
  de-compression overhead and integration are crucial for usability

- **Edge computing will become more important**
  Data should be transformed into actionable insight as soon as possible
  «Process data where it is generated» □ enable data filtration

- **It's an «infrastructure» problem**
  Detectors can «easily» do more
  But storage, archiving, transfer, sharing, etc.. become paralyzed

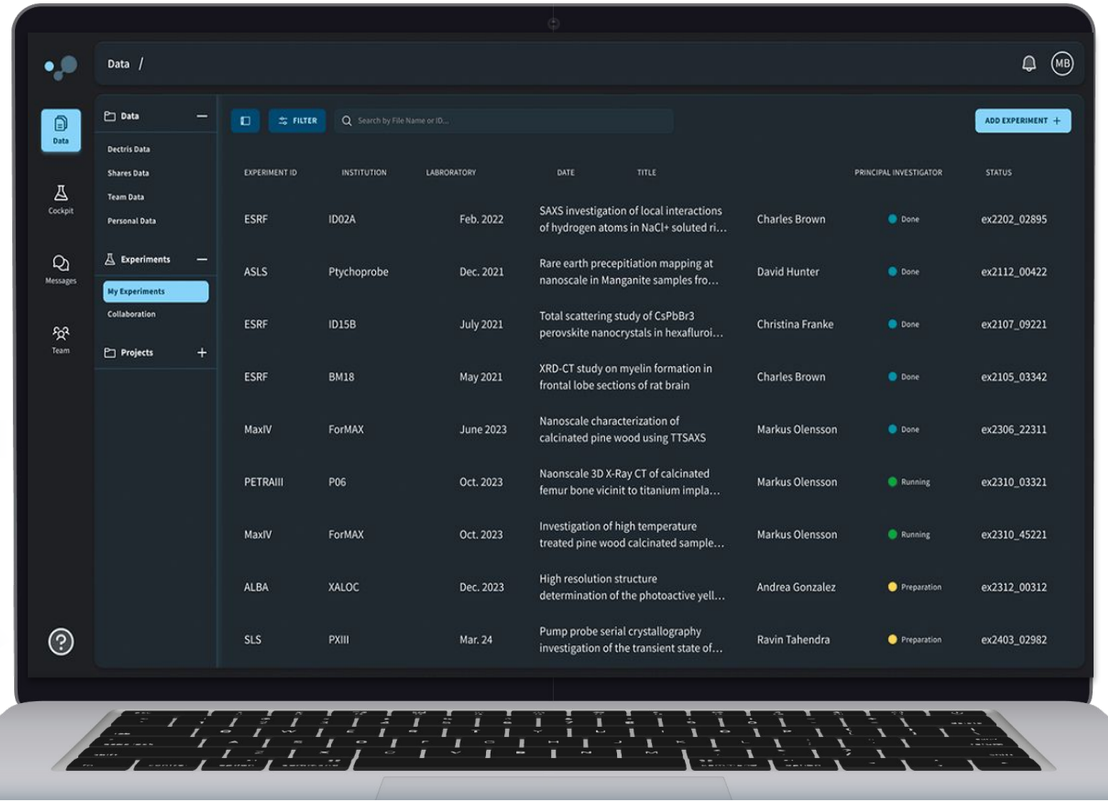
- **Bslz4 is great, but we need more**
  Will the community embrace lossy compression algorithms?

DECTRIS CLOUD is …

Vendor Agnostic

Analysis Toolbox

Built for Scientists



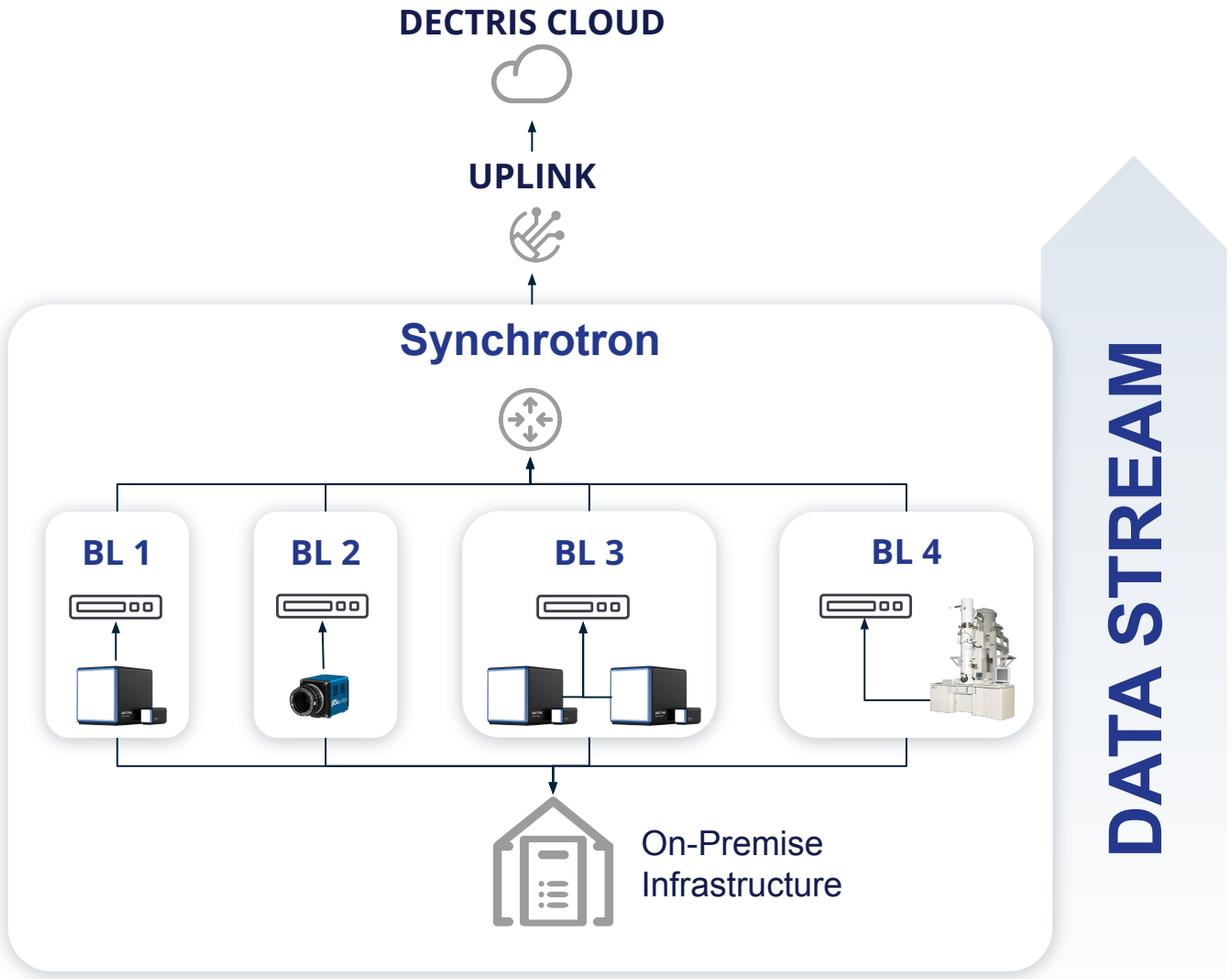
Community Driven

OPEN & FAIR

Globally Accessible

An open platform empowering collaborative discovery

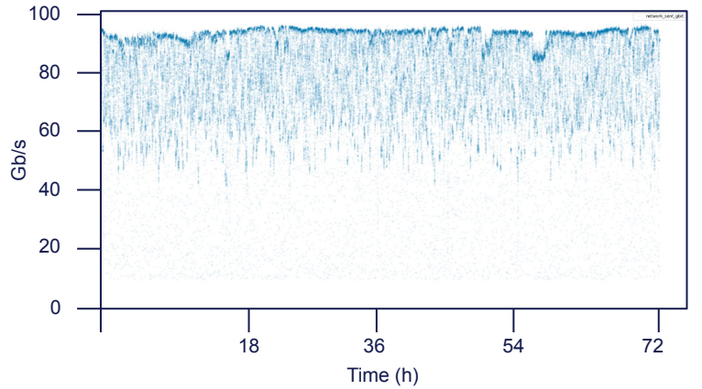
DECTRIS CLOUD Instant Data Availability



Architecture, Network and Application regularly penetration tested by 3rd party security experts

Cobalt

Data Ingest per Beamline



Uninterrupted upload of 1.2 PetaBytes in 72h

Average bandwidth of 92 GBits

2023, ALBA (ESP), XAIRA Beamline

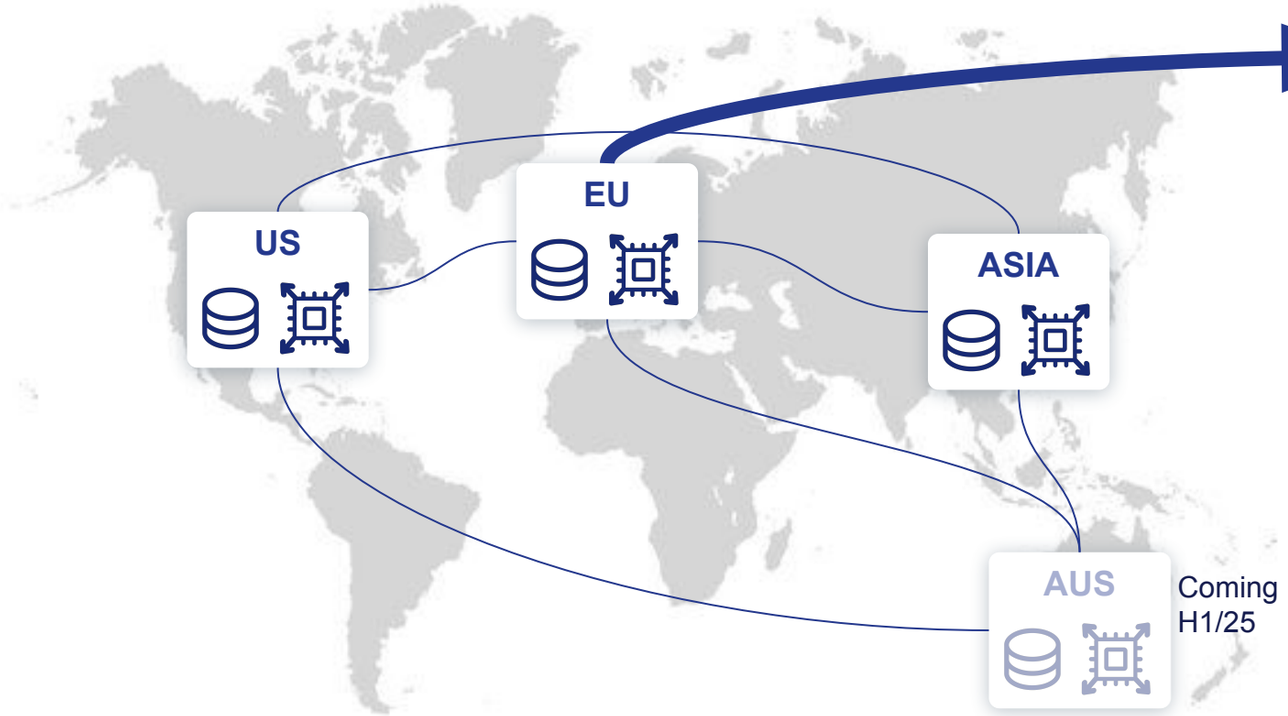
For beamlines & users:

+ Instant remote access to data

+ Real-Time sharing of data

DECTRIS CLOUD Global Data Bandwidth Capacity

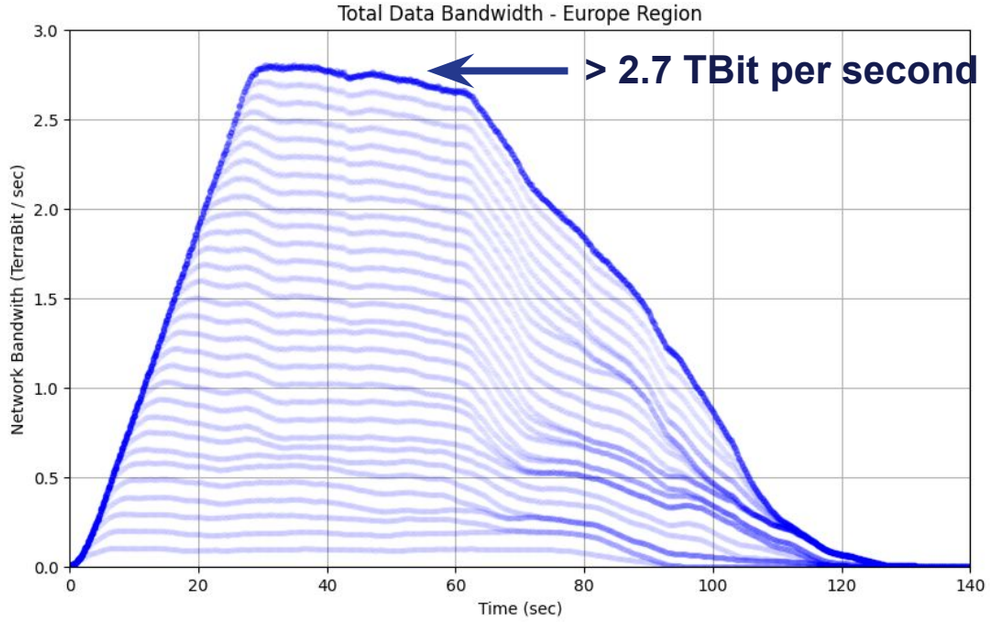
Cloud Architecture Overview



- DECTRIS CLOUD operates in 4 regions of the world.
- Tenant and cluster isolation to ensure data security
- Infrastructure, availability, and SLA provided by hyperscalers

aws Google Cloud

Data Access Bandwidth - EU Region

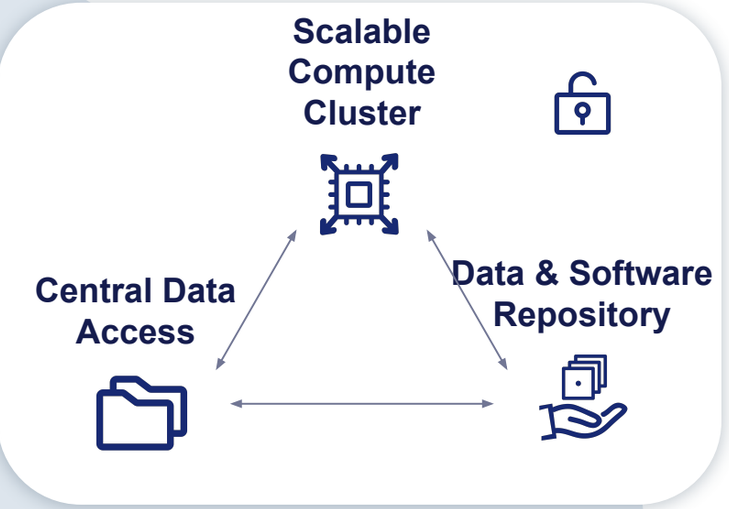


- Each region provides triple geo-redundancy for data storage
- Load balancing per availability zone to ensure data throughput
- Each AZ provides >2.7 TBit of data access BW

Total Global Bandwidth:

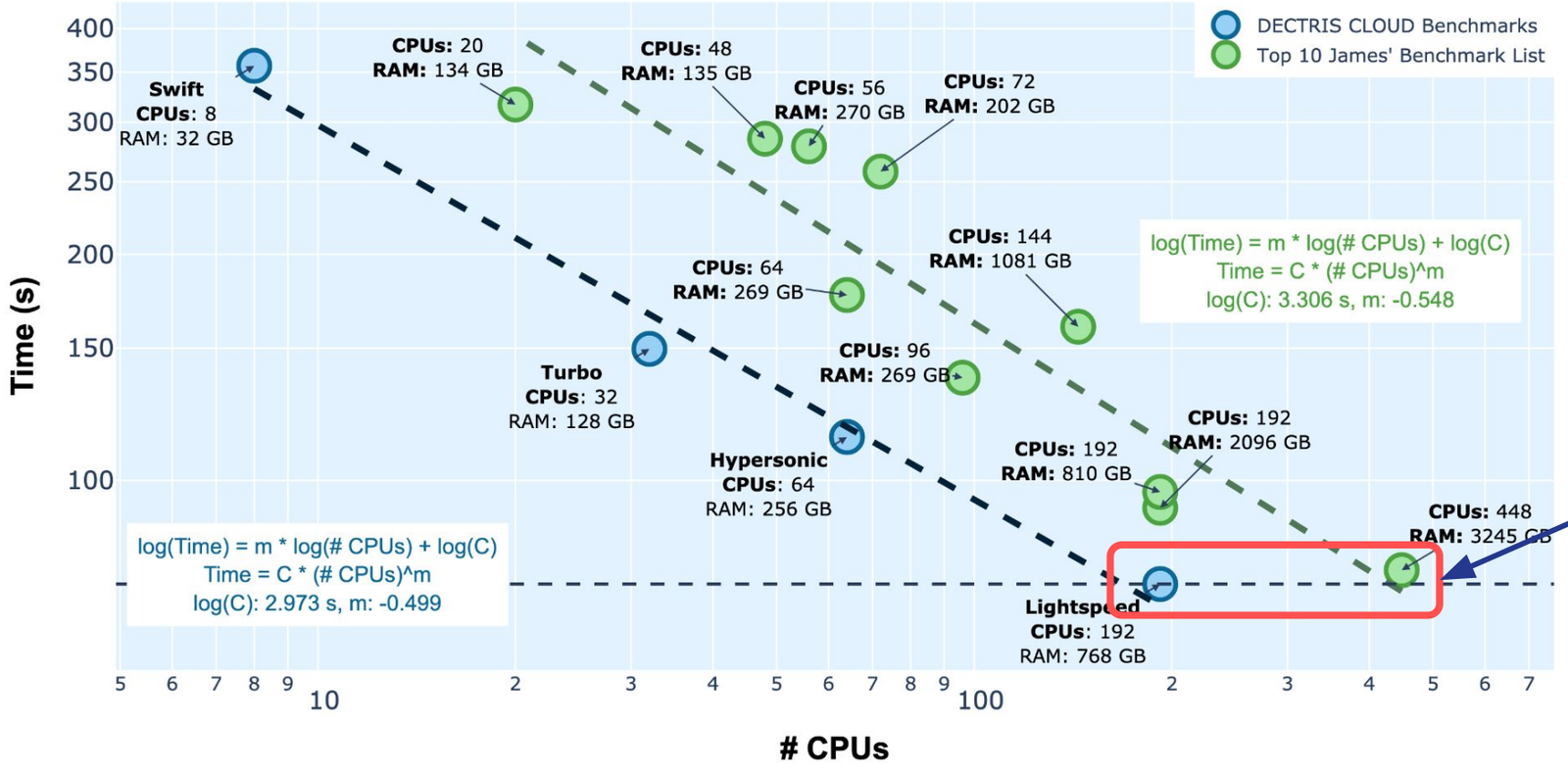
3 Regions **x** 3 Availability Zones **x** 2.7 TBit **=** >20 TBits

DECTRIS CLOUD Data processing



./xds_bench.com

Artificial MX Dataset with controlled Signal, Noise, Resolution, etc..



72.7 sec with ½ the CPUs

on-par with leading infrastructure but 2x more efficient

Courtesy to James Holton for feedback and support

Thank you for your attention!

DECTRIS Ltd.
5405 Baden-Dättwil
Switzerland
www.dectris.com