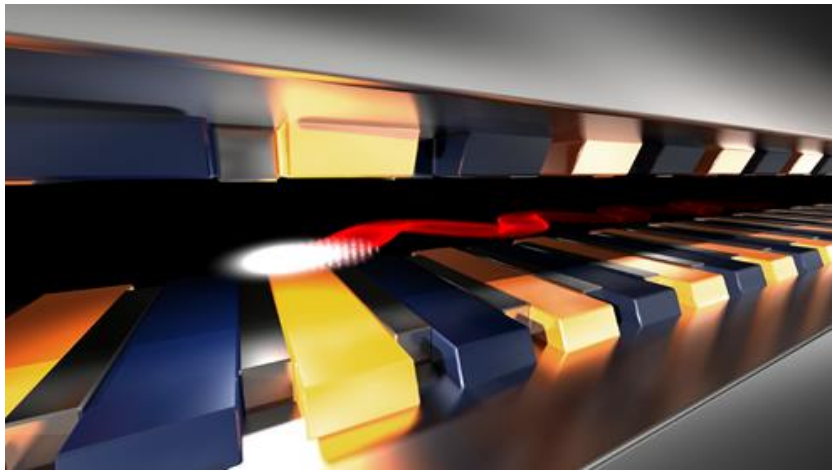
Petabyte scale data capture, control and evaluation at European XFEL

5th October 2022

Steve Aplin

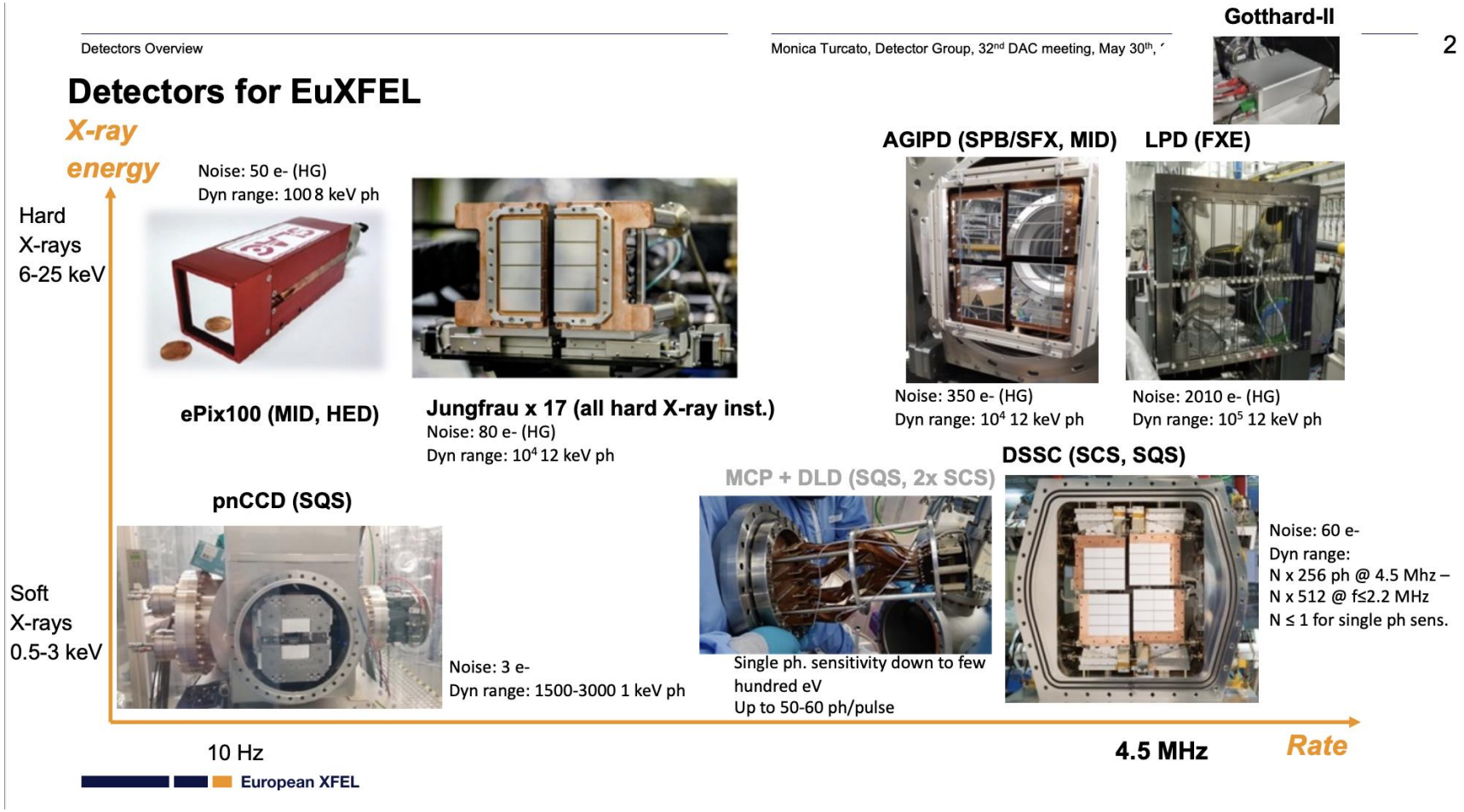
Department Head Data at European XFEL



Big Science Business Forum 2022

Big Science Business Forum 2022

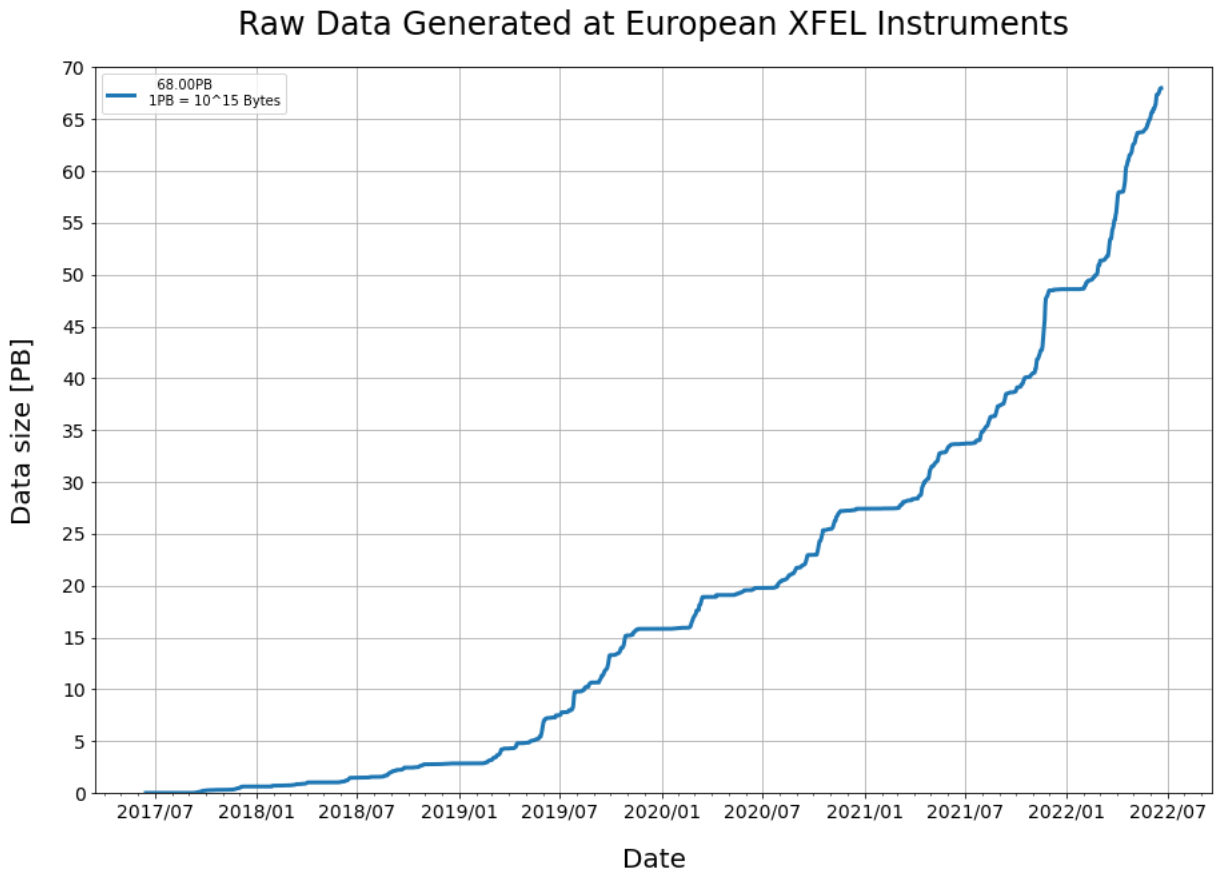
We need more Data ! … be careful what you wish for …



European XFEL

Big Science Business Forum 2022

We need more Data ! … be careful what you wish for …

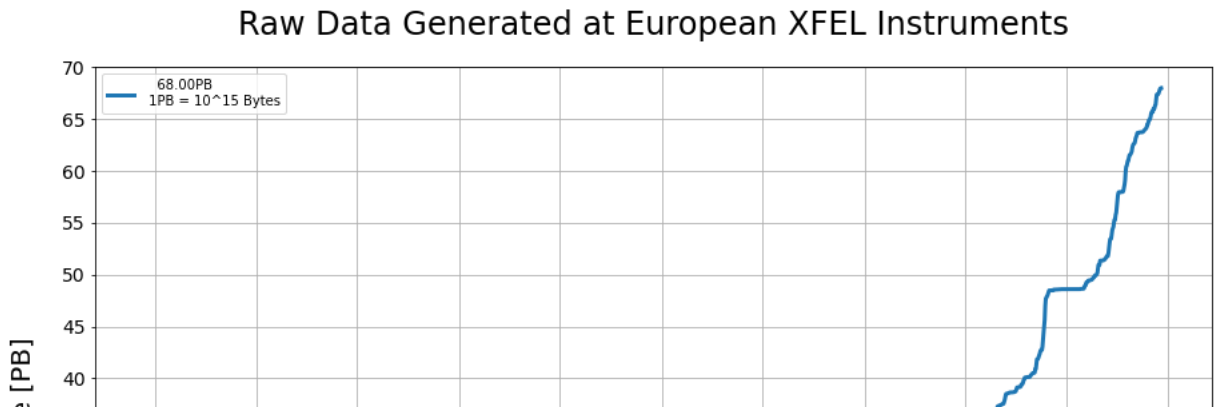


Detector type	Data/sec
AGIPD 1Mpxl	~7 GB/s
AGIPD 1Mpxl Double images	~14 GB/s
AGIPD 4Mpxl	~30 GB/s *
LPD 1Mpxl	~10 GB/s
DSSC 1Mpxl	~16 GB/s

European XFEL

Big Science Business Forum 2022

We need more Data ! … be careful what you wish for …



Detector type	Data/sec
AGIPD 1Mpxl	~7 GB/s
AGIPD 1Mpxl Double images	~14 GB/s
AGIPD 4Mpxl	~30 GB/s *
LPD 1Mpxl	~10 GB/s
DSSC 1Mpxl	~16 GB/s

The data flow from all four experiments at LHC for Run 2 was anticipated to be about *25 GB/s after* data reduction

- **ALICE: 4 GB/s (Pb-Pb running)**
- **ATLAS: 800 MB/s – 1 GB/s**
- **CMS: 600 MB/s**
- **LHCb: 750 MB/s**

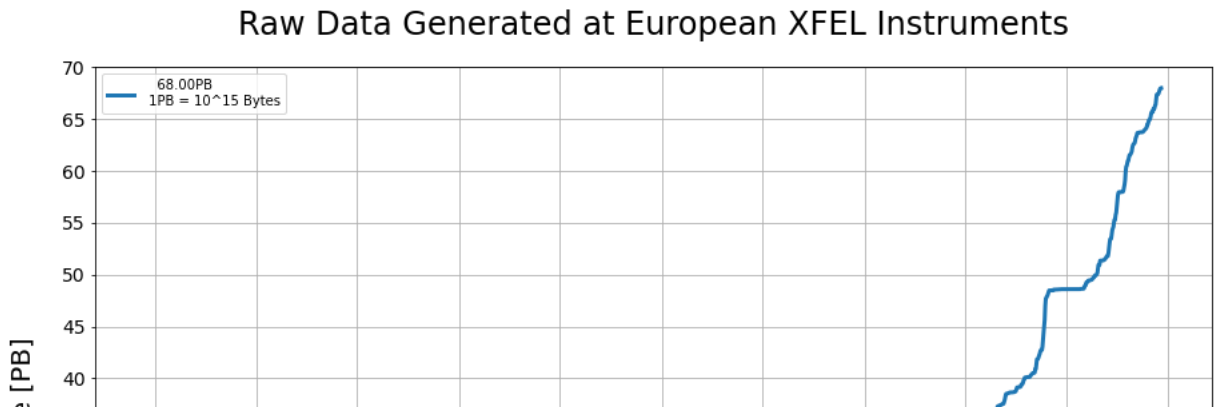
Data reduction in particle physics is built into it's DNA, it is intrinsic to the field's experimental viability.

The experiments are designed from the ground up on data reduction.

European XFEL

Big Science Business Forum 2022

We need more Data ! … be careful what you wish for …



Detector type	Data/sec
AGIPD 1Mpxl	~7 GB/s
AGIPD 1Mpxl Double images	~14 GB/s
AGIPD 4Mpxl	~30 GB/s *
LPD 1Mpxl	~10 GB/s
DSSC 1Mpxl	~16 GB/s

The data flow from all four experiments at LHC for Run 2 was anticipated to be about *25 GB/s after* data reduction

- **ALICE: 4 GB/s (Pb-Pb running)**
- **ATLAS: 800 MB/s – 1 GB/s**
- **CMS: 600 MB/s**
- **LHCb: 750 MB/s**

Data reduction in particle physics is built into it's DNA, it is intrinsic to the field's experimental viability.

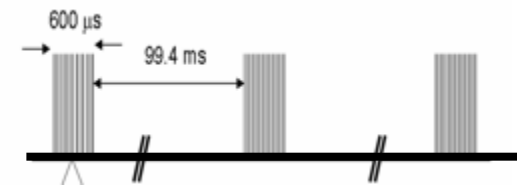
Required Ingest Rate of Data Center
100 GB/s

European XFEL

Maturing Systems 7 PB in 7 days 17th Nov. to 23rd Nov.

Controls – Hardware and Software

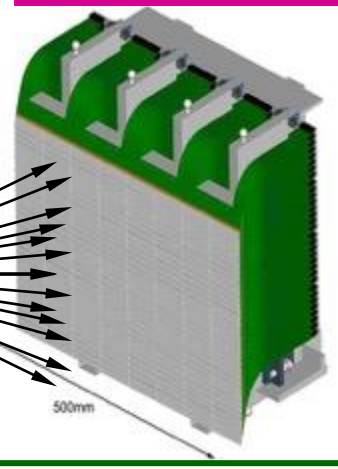
X-ray optics
Shape x-ray pulse
in space & spectrum



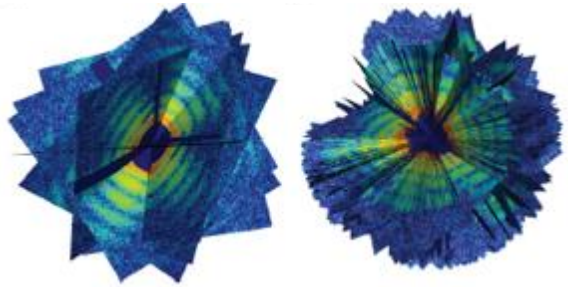
Photon diagnostics
Measure
x-ray pulse properties

Optical laser
Excite sample systems

Sample delivery
Refresh & reposition
samples



Detectors / DAQ
Take and store large number
of large data data sets



Data Analysis
Analyse immense amounts
of data

Large number of moving parts, many of them protein based

IT Infrastructure

Big Science Business Forum 2022

Current Approach = High Tech Brute Force

Schenefeld

Online GPFS

Cache

- Extremly high performance
- Data available immediately
- Optimised for concurrency
- High redundancy
- Dedicated storage for each SASE
- Very high cost per PB
- Capacity for a few days

Offline GPFS

Performance

- High performance
- Large scale data analysis
- High redundancy
- High cost per PB
- Shared within XFEL
- Large capacity

DESY Data Center

dCache

Capacity

- Lower performance
- Lower cost
- Scalability
- Shared within XFEL
- High capacity

Tape Archive

Safety

- Very slow
- Even lower cost
- Very high capacity
- Safety (second copy)
- Shared within DESY campus
- Long term

European XFEL

Big Science Business Forum 2022

Current Approach = High Tech Brute Force

Schenefeld

Online GPFS

Cache

DESY Data Center

Offline GPFS

Performance

dCache

Capacity

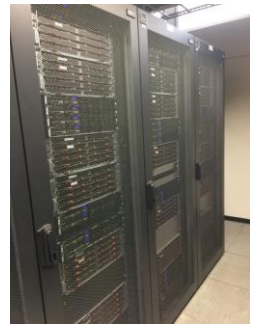
Tape Archive

Safety

European XFEL

Big Science Business Forum 2022

Current Resources and Investment Plan



- IBM GPFS (IBM Spectrum Scale) Online Filesystem 5 PB
- IBM GPFS (IBM Spectrum Scale) Online Filesystem 45 PB

- dCache Offline storage 110 PB

- 200 PB Tape Based Archive (LTO8, LTO9, Jagger)

- Online Compute: 60 nodes, ~ 50/50 split between CPU and GPU
- Offline Compute: 350 nodes, ~15000 cores Intel + AMD, theoretical Rpeak 1Peta Flop, 20 GPU nodes
  - (2022: 100 node extension delayed)

- InfiniBand fabrics long-range backbone HDR (200Gb/s) 1 Tbit/s to connect GPFS clusters between two sites

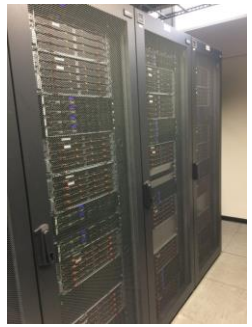
- 5 year system lifetime including support and warranty purchased for both compute and storage systems

Big Science Business Forum 2022

Current Resources and Investment Plan

- IBM GPFS (IBM Spectrum Scale) Online Filesystem 5 PB
- IBM GPFS (IBM Spectrum Scale) Online Filesystem 45 PB

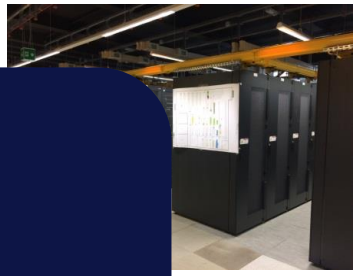
IBM Spectrum Scale



- dCache Offline storage 110 PB

dCache.org

- 200 PB Tape Based Archive (LTO8, LTO9, Jagger)



Annual Budget for Hardware Invest 2.5 Million Euro
Typicall split:

1.5 MEuro for Storage
1.0 MEuro for Compute

3 year framework contracts resulting from tenders managed by DESY (Host CC)

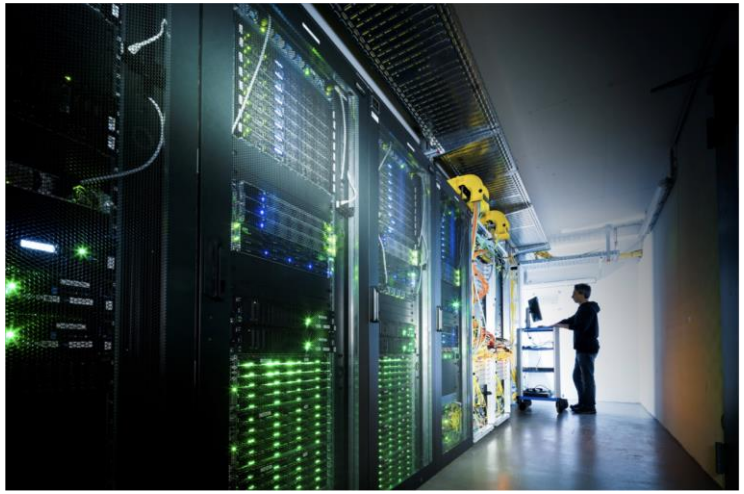
MEGWARE

European XFEL

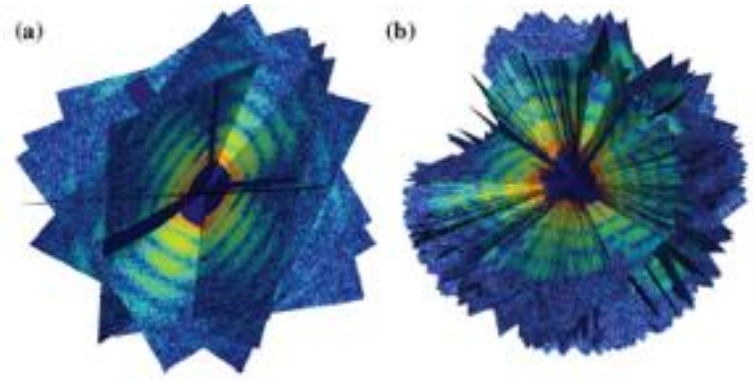
Big Science Business Forum 2022

Data Reduction and Compression is the only hope …

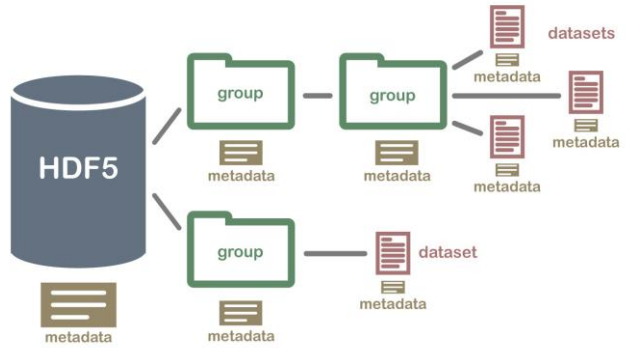
Reduce to Store



Reduce to Process



Reduce to Transport

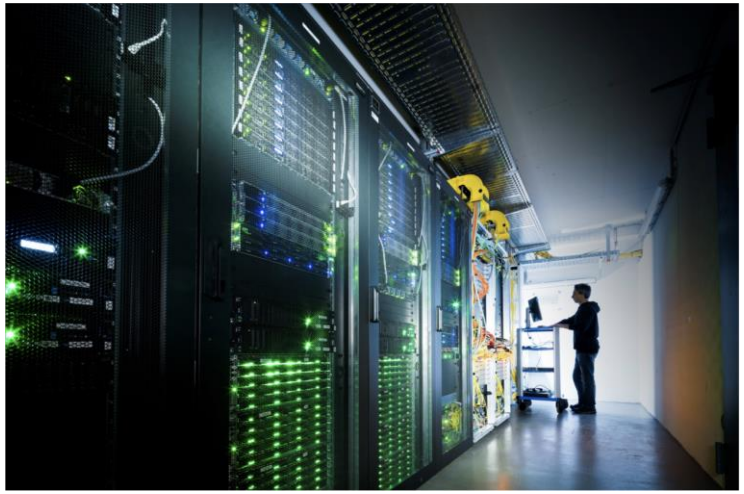


European XFEL

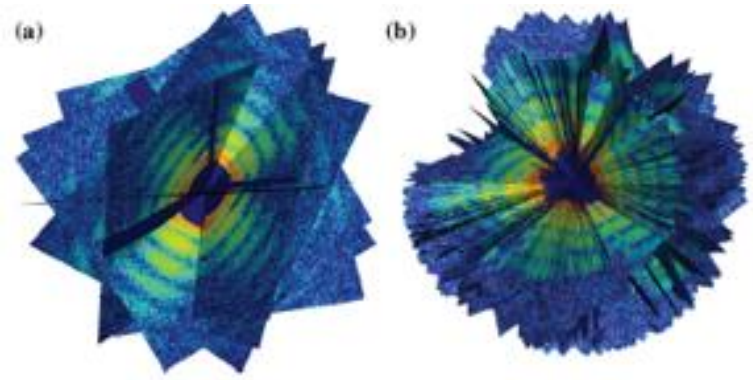
Big Science Business Forum 2022

Data Reduction and Compression is the only hope …

Reduce to Store

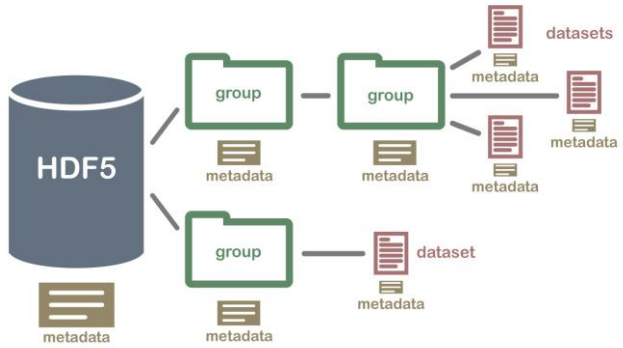


Reduce to Process



Reduce to Transport

Energy price increase means the Storage : I/O : CPU balance will need to be rethough



European XFEL