DRAM HW requirements: Status update

Documents:

- *Estimation of hardware requirements for data reduction with Mantid at ESS* (Simon)

- *Report on the data volume and rate during a ToF imaging experiment (Manuel)*

- Confluence:
  - UX: Data processing infrastructure (Thomas)
  - CPU requirements (Torben, Wojtek)

Data reduction

- Comprehensive report compiled by Simon

- 3 modes of operation:
  1. Live-reduction (online):
     - Threaded version, can handle most cases on a powerful workstation.
     - MPI version will be needed in order to handle all cases (not implemented yet).
  2. Interactively through GUI using threading
     1. Poor performance
     2. Typical #cores ~10
     3. ~5 simultaneous sessions per instrument
  3. Script based using MPI
     1. Scales well

User requirements

Compromize between good UX and feasibility for offline MPI runs:

$$t_{\text{reduction}} = \max\left(\frac{t_{\text{run}}}{5}, 30 \text{ sec}\right)$$

+ upper limit for very slow experiments.

In practice reduction time is given by:

$$t_{\text{reduction}} = t_0 + \frac{N_{\text{bin}} t_{\text{bin}} + N_{\text{event}} t_{\text{event}}}{N_{\text{core}}} + \frac{N_{\text{event}}}{B_{\text{max}}}$$

And memory by:

$$M_{\text{core}} \equiv \frac{M}{N_{\text{core}}} = 1 \text{ GByte} + N_{\text{spec}} \cdot 256 \text{ Byte} + \frac{N_{\text{workspace}}}{N_{\text{core}}} (24 N_{\text{bin}} + 16 N_{\text{event}}) \text{ Byte}$$

Examples of hardware requirements

ESTIA

Pixels	Mode	p-beam [MW]	p_{use}	f_{event} [s^{-1}]	t_{run} [s]	$t_{reduction}$ [s]	N_{core}	$\langle N_{core} \rangle$	M_{core} [GByte]
250000	reference-high-flux	0.2	0.001	$2 \cdot 10^6$	500.0	93	32	1	3
		0.5	0.001	$5 \cdot 10^6$	200.0	38	256	1	2
		1.0	0.001	$1 \cdot 10^7$	100.0	34	512	1	2
		2.0	0.001	$2 \cdot 10^7$	50.0	34	512	1	2
		5.0	0.001	$5 \cdot 10^7$	20.0	34	512	2	2
	reference-normal	0.2	0.01	$8 \cdot 10^4$	12500.0	367	6	1	10
		0.5	0.01	$2 \cdot 10^5$	5000.0	367	6	1	10
		1.0	0.01	$4 \cdot 10^5$	2500.0	367	6	1	10
		2.0	0.01	$8 \cdot 10^5$	1250.0	214	11	1	6
		5.0	0.01	$2 \cdot 10^6$	500.0	93	32	1	3
	specular-high-flux	0.2	0.049	$4 \cdot 10^4$	250.0	24	3	1	4
		0.5	0.049	$1 \cdot 10^5$	100.0	24	3	1	4
		1.0	0.049	$2 \cdot 10^5$	50.0	24	3	1	4
		2.0	0.049	$4 \cdot 10^5$	25.0	24	3	1	4
		5.0	0.049	$1 \cdot 10^6$	10.0	24	3	1	4
	specular	0.2	0.47	$1.6 \cdot 10^4$	625.0	24	3	1	4
		0.5	0.47	$4 \cdot 10^4$	250.0	24	3	1	4
		1.0	0.47	$8 \cdot 10^4$	125.0	24	3	1	4
		2.0	0.47	$1.6 \cdot 10^5$	62.5	24	3	2	4
		5.0	0.47	$4 \cdot 10^5$	25.0	24	3	3	4
	off-specular	0.2	0.47	$1.6 \cdot 10^4$	62500.0	367	6	1	10
		0.5	0.47	$4 \cdot 10^4$	25000.0	367	6	1	10
		1.0	0.47	$8 \cdot 10^4$	12500.0	367	6	1	10
		2.0	0.47	$1.6 \cdot 10^5$	6250.0	367	6	1	10
		5.0	0.47	$4 \cdot 10^5$	2500.0	367	6	1	10
500000	reference-high-flux	0.2	0.001	$4 \cdot 10^6$	250.0	46	128	1	2
		0.5	0.001	$1 \cdot 10^7$	100.0	34	512	1	2
		1.0	0.001	$2 \cdot 10^7$	50.0	34	512	1	2
		2.0	0.001	$4 \cdot 10^7$	25.0	34	512	2	2
		5.0	0.001	$1 \cdot 10^8$	10.0	34	512	4	2
	reference-normal	0.2	0.01	$1.6 \cdot 10^5$	6250.0	371	6	1	12
		0.5	0.01	$4 \cdot 10^5$	2500.0	371	6	1	12
		1.0	0.01	$8 \cdot 10^5$	1250.0	216	11	1	7
		2.0	0.01	$1.6 \cdot 10^6$	625.0	119	23	1	4
		5.0	0.01	$4 \cdot 10^6$	250.0	46	128	1	2
		0.2	0.049	$8 \cdot 10^4$	125.0	26	4	1	6
		0.5	0.049	$2 \cdot 10^5$	50.0	26	4	1	6

DREAM

Pixels	Mode	p-beam [MW]	p_{use}	f_{event} [s^{-1}]	t_{run} [s]	$t_{reduction}$ [s]	N_{core}	$\langle N_{core} \rangle$	M_{core} [GByte]
4000000	high-resolution	0.2	0.33	$1.7 \cdot 10^4$	28846.2	246	256	2	77
		0.5	0.33	$4.3 \cdot 10^4$	11538.5	246	256	4	77
		1.0	0.33	$8.7 \cdot 10^4$	5769.2	246	256	8	77
		2.0	0.33	$1.7 \cdot 10^5$	2884.6	246	256	15	77
		5.0	0.33	$4.3 \cdot 10^5$	1153.8	180	362	38	55
	medium	0.2	0.33	$1.3 \cdot 10^5$	3750.0	220	45	2	63
		0.5	0.33	$3.3 \cdot 10^5$	1500.0	220	45	5	63
		1.0	0.33	$6.7 \cdot 10^5$	750.0	119	91	10	32
		2.0	0.33	$1.3 \cdot 10^6$	375.0	70	181	23	17
		5.0	0.33	$3.3 \cdot 10^6$	150.0	38	512	85	8
	high-flux	0.2	0.33	$1 \cdot 10^6$	500.0	87	32	4	15
		0.5	0.33	$2.5 \cdot 10^6$	200.0	37	128	16	6
		1.0	0.33	$5 \cdot 10^6$	100.0	28	256	48	4
		2.0	0.33	$1 \cdot 10^7$	50.0	28	256	96	4
		5.0	0.33	$2.5 \cdot 10^7$	20.0	28	256	240	4
12000000	high-resolution	0.2	0.33	$5.2 \cdot 10^4$	9615.4	355	512	13	116
		0.5	0.33	$1.3 \cdot 10^5$	3846.2	355	512	32	116
		1.0	0.33	$2.6 \cdot 10^5$	1923.1	355	512	63	116
		2.0	0.33	$5.2 \cdot 10^5$	961.5	355	512	125	116
		5.0	0.33	$1.3 \cdot 10^6$	384.6	355	512	312	116
	medium	0.2	0.33	$4 \cdot 10^5$	1250.0	215	128	15	67
		0.5	0.33	$1 \cdot 10^6$	500.0	89	362	43	27
		1.0	0.33	$2 \cdot 10^6$	250.0	69	512	94	20
		2.0	0.33	$4 \cdot 10^6$	125.0	69	512	187	20
		5.0	0.33	$1 \cdot 10^7$	50.0	69	512	466	20
	high-flux	0.2	0.33	$3 \cdot 10^6$	166.7	32	362	47	8
		0.5	0.33	$7.5 \cdot 10^6$	66.7	29	512	146	7
		1.0	0.33	$1.5 \cdot 10^7$	33.3	29	512	291	7
		2.0	0.33	$3 \cdot 10^7$	16.7	29	512	581	7
		5.0	0.33	$7.5 \cdot 10^7$	6.7	29	512	1451	7

There are still many uncertainties

- Notably MAGIC and DREAM but also …

- ~ 3 months of effort to get this far

Regardless, we will have large uncertainties until we are in operation.
We thus need a strategy for how to handle this!

Input to required hardware for data analysis

- Feedback from Instrument scientist:
  - LoKI
  - ESTIA
  - BiFrost
  - ODIN

- Instrument scientists were asked to estimate the required compute capacity needed for their beamline

- See Confluence page for details:

Pages / Data Reduction, Analysis and Modelling Home / Instrument Techniques & Software

CPU requirements

LoKI

General req.	Post analysis	Workflow
0.5 users /day	10 cluster users / day	Live data - Yes
SasView	10 simultaneously cluster users / day	Iterative analysis
1-16 CPU	16-32 CPU (total)	
Data file: small		

Typical running device: **laptop/workstation**

ESTIA-Specular

General req.	Post analysis	Workflow
100 experiments / day *	5-10 cluster users / day	Live data - Yes
GenX BornAgain	1-5 simultaneously cluster users / day	Iterative analysis
8-32 CPU	Max 98 CPU (total) Average 4 CPU	
Data file: 0.1 MB-1 MB		

Typical running device: **laptop/workstation**

* Distribution: per day: 10 (25%) / 30 (15%)/ 100(50%) / 1000 (10%)

Bifrost

General req.	Post analysis	Workflow
1-2 experiments / week	2 cluster users / day	Live data – Yes Imperative
NO main analysis software for INS	2 simultaneously cluster users / day	Iterative analysis
	Max ?? CPU (total) Average ??? CPU	
Data file: 20 MB – 5 GB		

Typical running device: **laptop - workstation - cluster**

ESTIA-Others

General req.	Post analysis	Workflow
5 experiments / day	1 cluster users / day	Live data - Maybe
BornAgain	1 simultaneously cluster users / day	Iterative analysis
16 - max CPU Average 32 CPU	Full node Average 8 CPU	
Data file: 0.1 GB – 1 GB		

Typical running device: **cluster**

ODIN

Use case - ToF tomography

DRAFT Sigfried sits with his laptop in the airport connected to the internet. Sigfried *somehow* performs the following tasks on the ESS infrastructure:

- accesses X GB of raw ToF data stored at ESS from a past experiment performed by Hildur
- corrects and normalizes the raw data using Mantid in 1 min. Corrected and normalized data can be retrieved later by people with the right authorization through SciCat.
- reconstruct the data. The resulting energy resolved 3D volume is returned in 20 mins and can be retrieved later by people with the right authorization through SciCat.
- Sigfried installs TensorFlow for later usage.

Sigfried gets on his flight. Back at his home institution Sigfried accesses the ESS infrastructure again and has readily access to all his derived data from his previous session and TensorFlow. Sigfried performs the following task:

- Search for patterns in the energy resolved volumetric data using TensorFlow.

This is a slow operation, so Sigfried lets it run over night, returns next day and inspects the data. He makes the results available in SciCat.

CPU requirements - ToF tomography

DRAFT For return of result in ~20 min, 1 core is required per energy bin. Thus **3000 cores** for 3000 energy bins. Some algorithms treats bin independently from each other, others use result from one bin as a priori to other bins.

For worst case 3000 energy bins (for perfect parallelization on four cores per energy bin):

#cores	Wall time for 3000 energy bins
375 - 749	~160 min
750 - 1499	~80 min
1500 - 2999	~40 min
3000 - 5999	~20 min
6000 - 8999	~10 min
9000 - 11999	~7 min
12000 - 14999	~5 min

We should based on numbers from SNS expect 2000 energy bins to be more normal.

Data requirements - ToF tomography

DRAFT Size of data sets to be processed: #projections × #(energy bins) × #(size per projection). Number of energy bins is 3000 in worst case scenario and size per projection is 0.5 MB for 500 x 500 ToF detector and 2 MB for a 1000 x 1000 ToF detector for *raw* data. The images used for reconstruction will twice those sizes. Thus, 3 GB, 12 GB per projection, respectively. It takes XX mins to acquire the data. #projections = 1.5 × (width of sample in pixels). #projections ~ 300-400 for 500 x 500.

Memory requirements

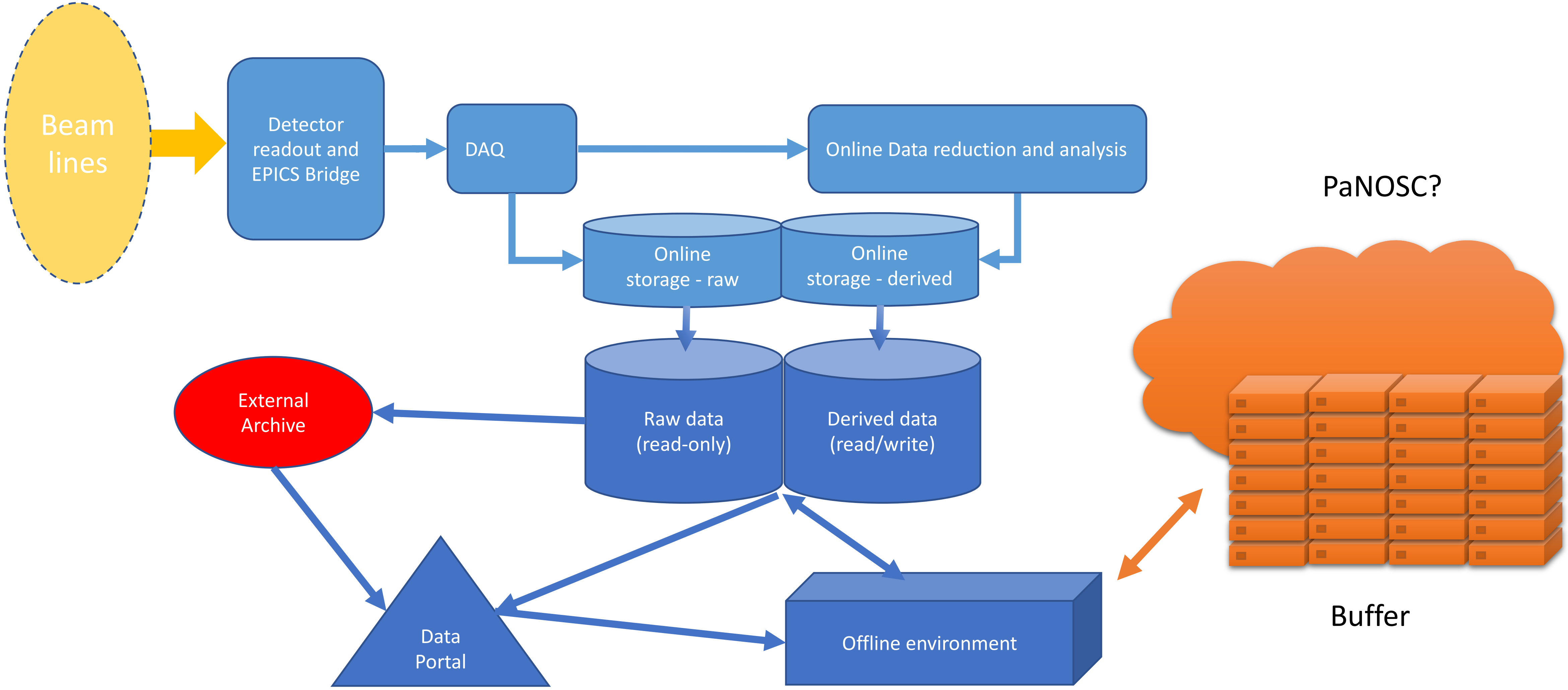
Minimum 128 GB

What to do for analysis?

- Define one or a few standard analyses for each instruments,
- impose the same requirements as for reduction and estimate #cores and #RAM

However, large uncertainties will remain. A strategy for how to handle this will be required.

Everything is work in progress, e.g. still need estimations for data storage


Beam lines
Detector readout and EPICS Bridge
DAQ
Online Data reduction and analysis
PaNOSC?
Online storage - raw
Online storage - derived
External Archive
Raw data (read-only)
Derived data (read/write)
Data Portal
Offline environment
Buffer