Data processing center simulation for the BM@N experiment based on the probabilistic approach

D. PRIAKHINA

V. TROFIMOV, V. MITSYN, I. PELEVANYUK

G. OSOSKOV, K. GERTSENBERGER

Introduction

The important task

Simulation of data storage and processing centers, both as from the BM@N detector, as for simulated particle collision events for comparison with the expected results of real storage and processing processes.

Probabilistic approach to simulate

Using of probability distributions of significant data acquisition processes:

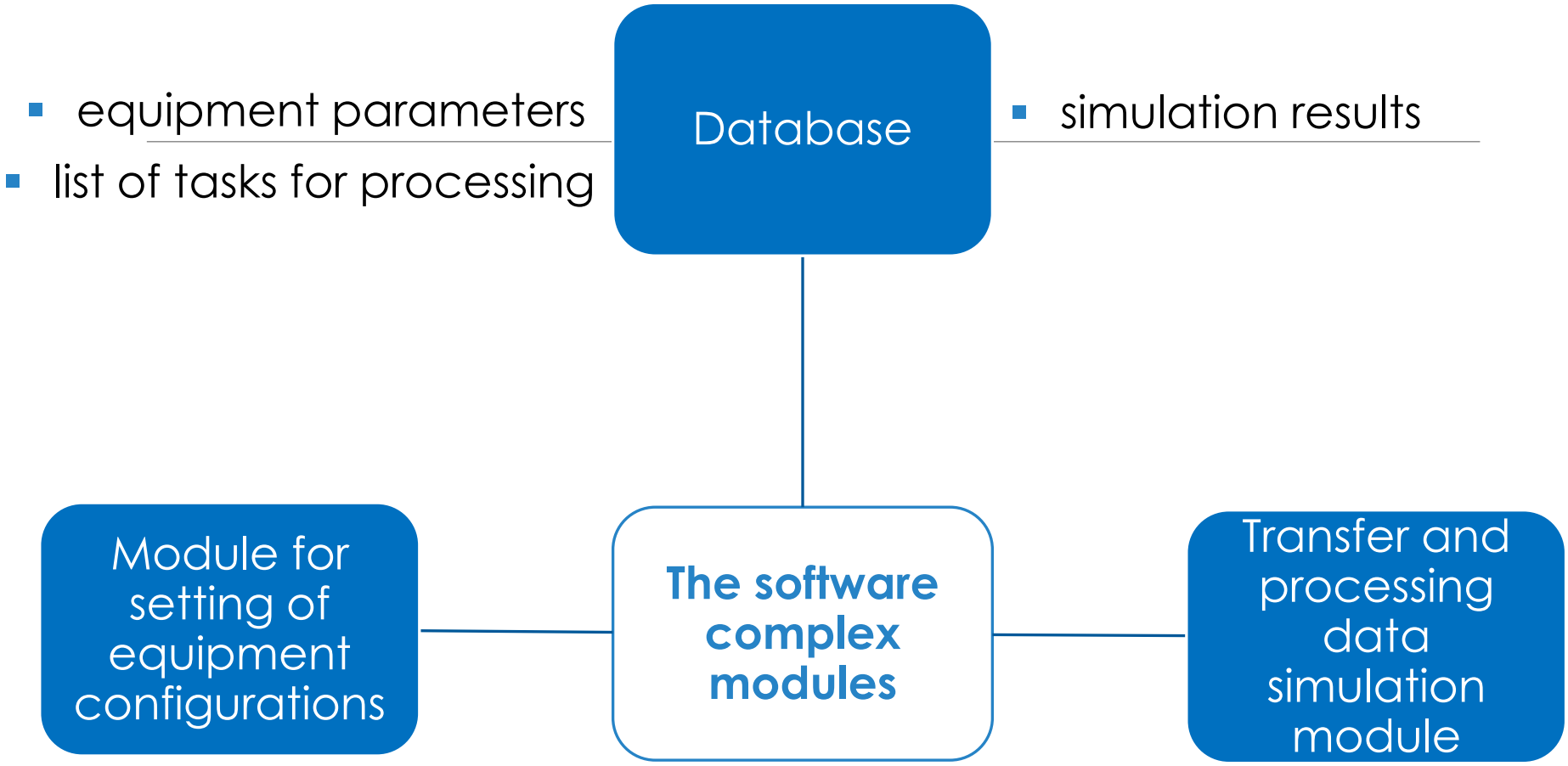
- the probabilities of loss of incoming information for different configurations of the equipment used are defined.

Simulation goal

Determine the hardware configuration that will ensure the operability of the data storage and processing system:

- takes into account hardware parameters and expected data flows and tasks;
- takes into account parameters of data transfer protocols : the type, the number of parallel threads etc.

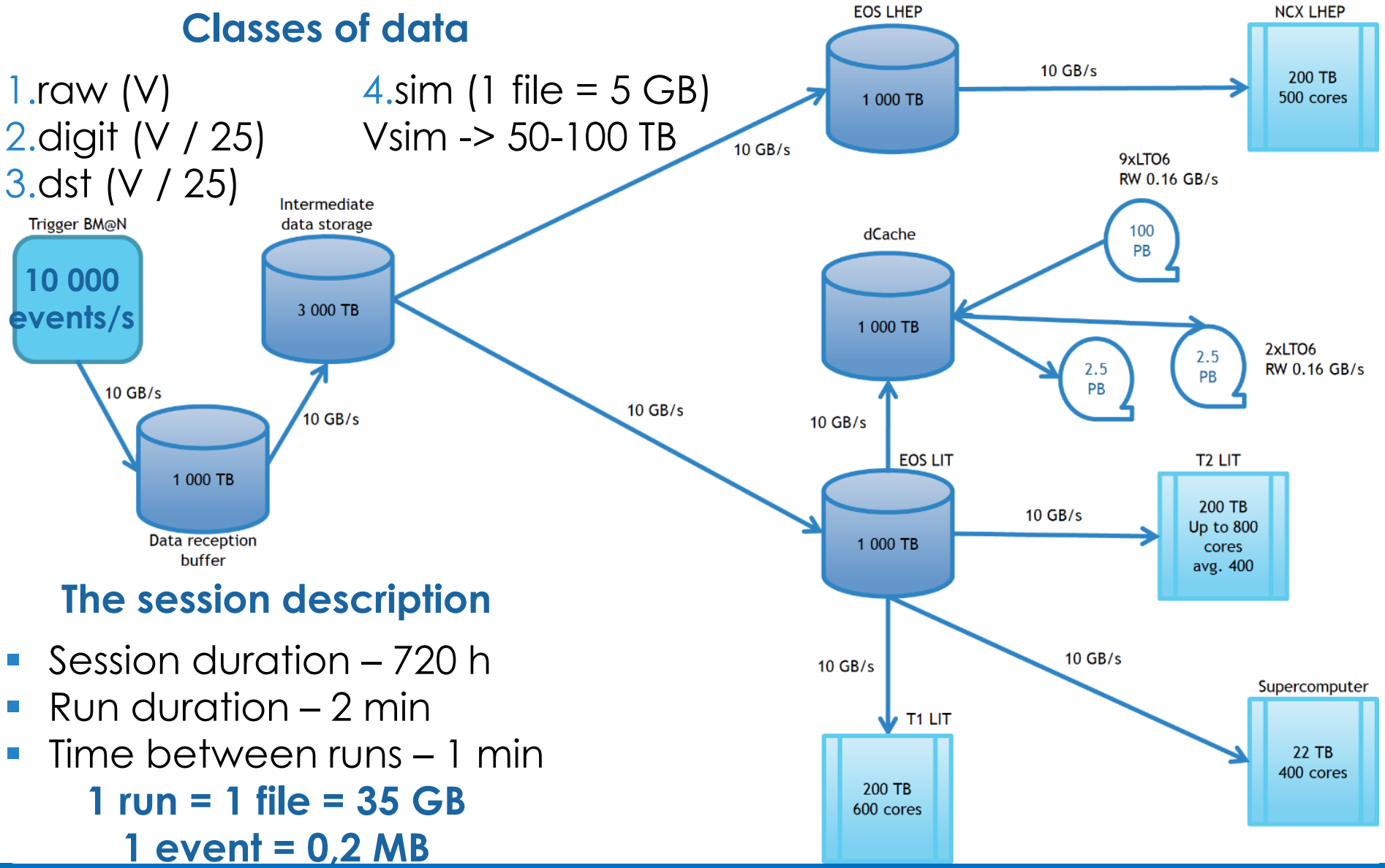
The simulation software complex



The simulated structure

Classes of data

1. raw (V)
2. digit (V / 25)
3. dst (V / 25)
4. sim (1 file = 5 GB)
   Vsim -> 50-100 TB



The session description

- Session duration – 720 h
- Run duration – 2 min
- Time between runs – 1 min

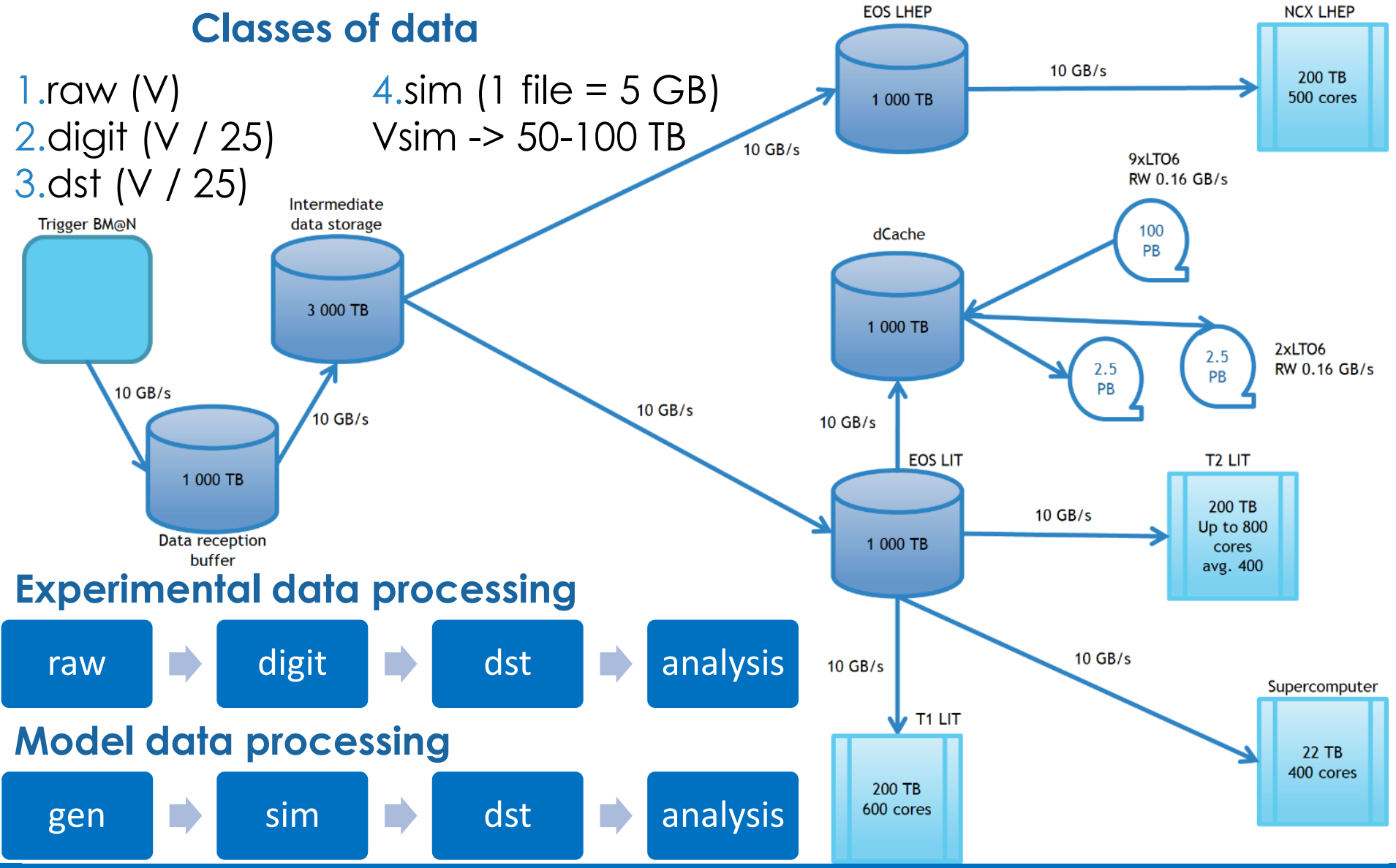
1 run = 1 file = 35 GB

1 event = 0,2 MB

The simulated structure

Classes of data

1. raw (V)
2. digit (V / 25)
3. dst (V / 25)
4. sim (1 file = 5 GB)
   Vsim -> 50-100 TB



Experimental data processing

raw → digit → dst → analysis

Model data processing

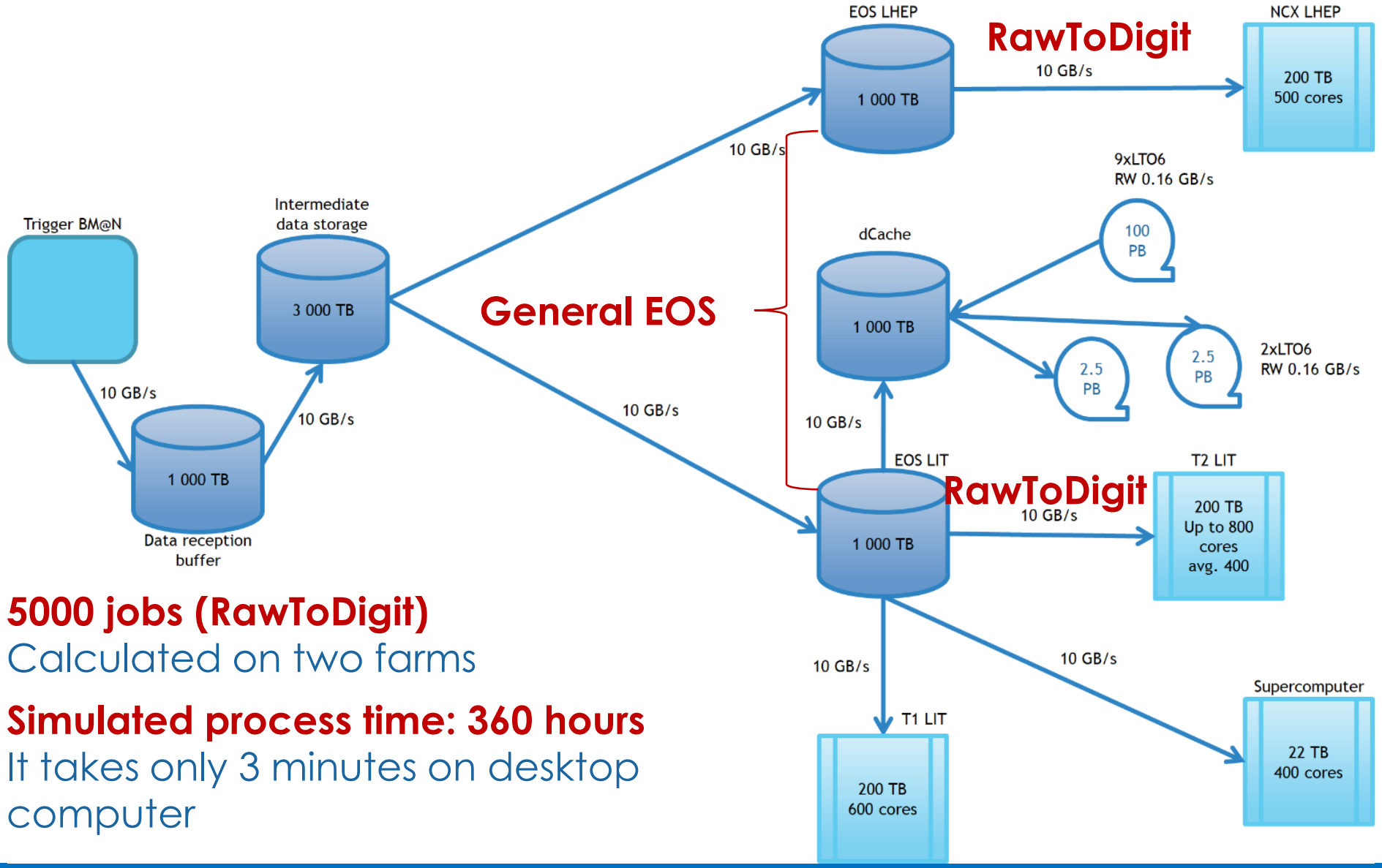
gen → sim → dst → analysis

Classes of jobs

№	Class	Event processing time on one processor (ms)	The average amount of input (TB)	Number of events in the file (1 file = 1 job)	Job execution time (s)	The average amount of output (TB)	Number of jobs
1	RawToDigit	20	0,035	175 000	3 500	0,004375	2 000
2	DigitToDst	10	0,004375	175 000	1 750	0,000875	6 000
3	GenToSim	20	0,035	175 000	3 500	0,005	600
4	SimToDst	10	0,005	175 000	1 750	0,0001	200
5	DstToAnal	10			1 750		

- ✓ Full processing of session results – getting reconstructed data for all the original session data.

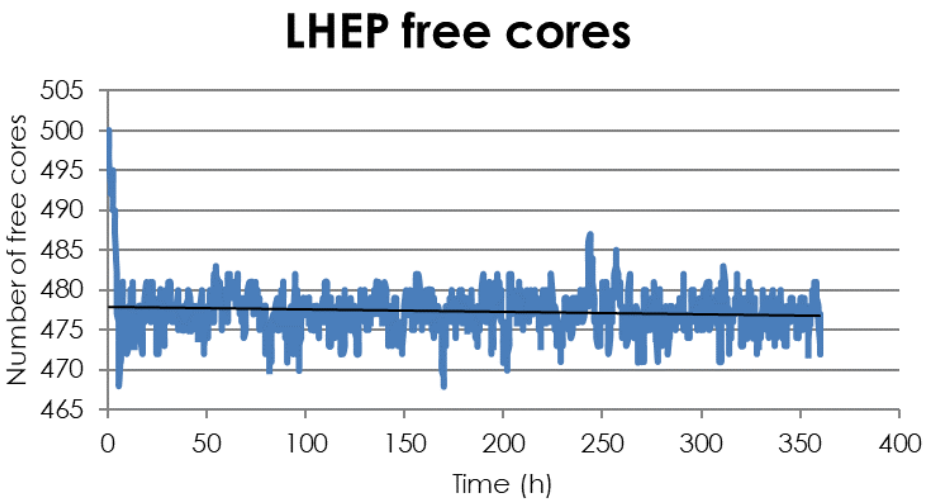
The process of primary data processing



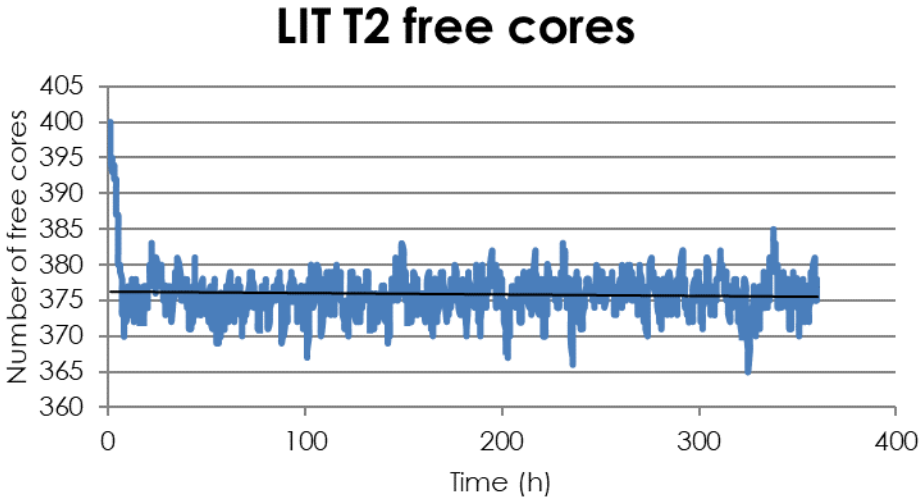
5000 jobs (RawToDigit)
Calculated on two farms

Simulated process time: 360 hours
It takes only 3 minutes on desktop computer

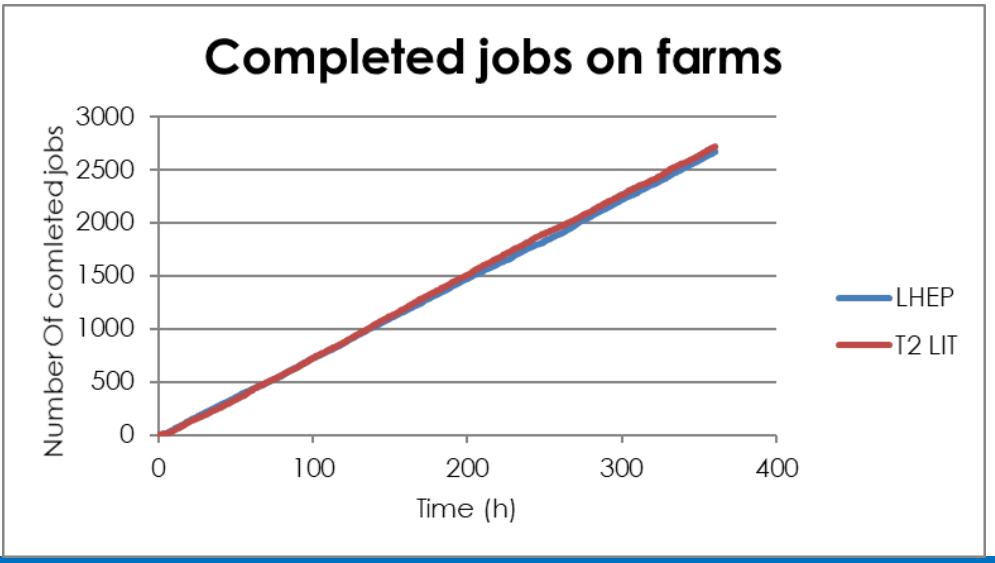
Simulation results



Avg.: 478 free slots



Avg.: 376 free slots



We can convert raw-data to digit, taking up no more than 25 slots on each farm.

Simulated process time: 360 h

5000 jobs (RawToDigit) calculated on two farms

Conclusions and Outlook

- Successfully simulated primary data handling process.
- Converting raw-data to digit-data would require no more than 25 cores per farm (for the hardware and data flow configurations under consideration).
- If the collaboration approves our activity, a full-scale BM@N data center simulation with various hardware and software configurations will be conducted and presented.

Thank you for the attention!

D. PRIAKHINA

V. TROFIMOV, V. MITSYN, I. PELEVANYUK

G. OSOSKOV, K. GERTSENBERGER