BM@N Run 8 data processing on a distributed infrastructure with DIRAC

K. Gertsenberger[1], I.S. Pelevanyuk[2]
[1] Veksler and Baldin Laboratory of High Energy Physics
[2] Meshcheryakov Laboratory of Information Technologies

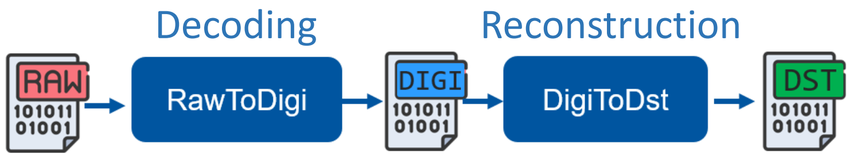
JINR MLIT — LABORATORY OF INFORMATION TECHNOLOGIES — MESHCHERYAKOV

31 k	400 TB	12	363 k	86 years	~3 PB
amount of raw files	BM@N Run8 RAW data size	data processing campaigns	jobs completed	total computation time	data transferred

BM@N Data

The 8th BM@N run with the xenon ion beam was conducted between December 2022 and February 2023. BM@N collected about 500M Xe+CsI interactions at the beam kinetic energy of 3.8 AGeV and 50M interactions at the kinetic energy of 3 AGeV. All these events were recorded in the form of **around 31000 files** with total size of **400 TB** on the EOS storage system of the NICA cluster.

In order to allow physics analysis each file is converted from RAW format to DST (data summary tape) format. According to BM@N data processing model this transition is done in two steps. On the first step raw files converted and decoded to event trees with detector digits in the CERN ROOT format (generally called DIGI format). On the second step, a special macro restores information on the particles registered by the detectors, their tracks and other parameters that are relevant for further physics analysis and saves them in the DST format. The general flow of data transformation is shown in the following schema:



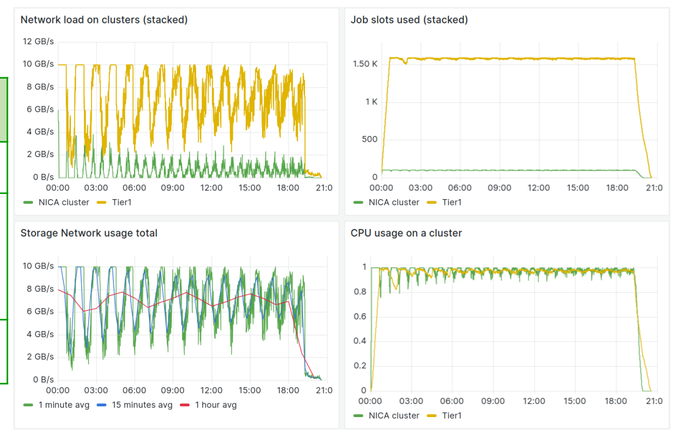
The RAW files contain information on registered signals and metadata from different detectors within the BM@N facility.

Simulation of Jobs Execution

Specially developed simulation system was used. It allows simulating job execution on a distributed infrastructure taking into consideration **network throughput** of different components, and **CPU cores performance** on different clusters.

Simulation parameters

Resources	Cores	CPU speed	Network speed
EOS MLIT	-	-	10 GB/s
Tier1 (slow cores)	1125	13 DB12	10 GB/s
Tier1 (fast cores)	375	24 DB12	
NICA cluster	100	24 DB12	10 GB/s

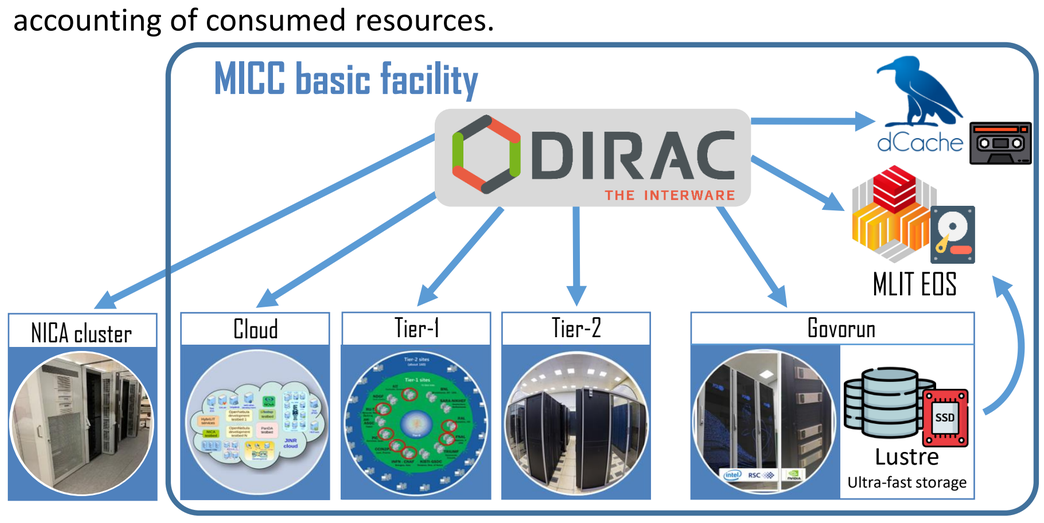


The graph shows occasional full network usage, but CPU usage is still acceptable. No severe "cold start" issues were observed with the default DIRAC pilot submission rate for RawToDigi jobs. Since DigiToDst jobs are much less data intensive, no simulation was required for them.

Data from the simulation of job execution let us safely initiate the full raw data processing for BM@N Run 8.

Distributed Infrastructure

Distributed heterogeneous computing infrastructure based on the DIRAC platform, has been being developed at JINR since 2016. It is used to process data from JINR experiments. For BM@N collaboration it **combines most of major resources** that are available: MICC resources, the NICA cluster, the clouds of the JINR Member States. The usage of the DIRAC platform makes it possible to get unified access to all integrated computing and storage resources, to carry out performance analysis and accounting of consumed resources.

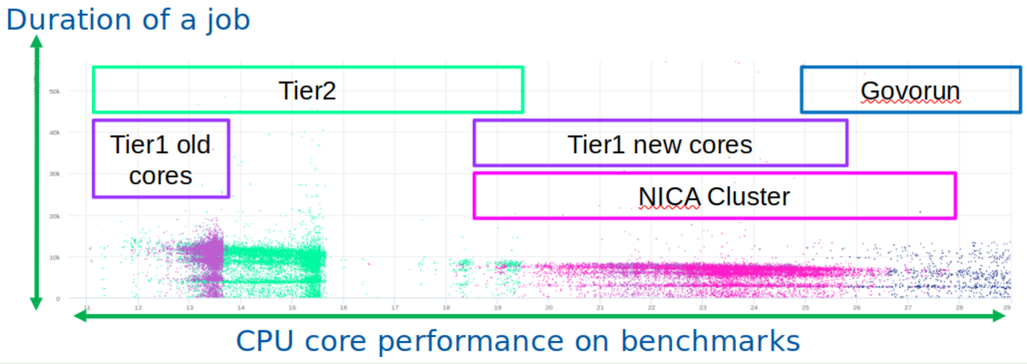


Run 8 Data Processing

First **RawToDigi campaign** took 35 hours. At the beginning, Tier1 was occupied by another workloads, but at the peak there were 1500 jobs on Tier1 and 100 jobs on the NICA cluster. The maximum transfer speed between worknodes and MLIT EOS was **7.5 GB/s**. Average job execution time was around **90 minutes**. The size of the produced DIGI files is **23 TB**.

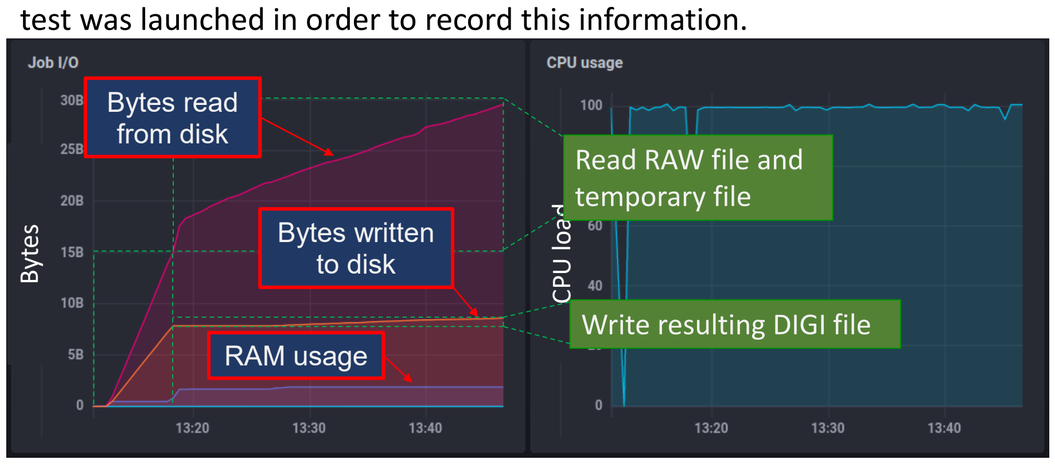
DigiToDst campaign is much less data intensive, disk requirements are around 4 GB, and average job duration was around 3 hours. **All major computing resources were used**: Tier1, Tier2, the NICA cluster and Govorun supercomputer. The peak **amount of simultaneously executed jobs** was around 3100. The full reconstruction campaign took around 30 hours. The total size of the DST files is **53 TB**.

Duration of a job



Raw to Digi Complexity

Most of the RAW files have size around **15 GB**. It is important to know how many computing resources each RawToDigi job consumes. Special test was launched in order to record this information.



The most limiting requirement for this process is the disk size. Not all computing resources can provide **25 GB** of free disk space for a job and therefore allocate the amount of the jobs corresponding to the amount of available CPU cores. With these requirements two computing resources were chosen for this campaign: **Tier1** centre and the **NICA cluster**.

Since RawToDigi transformation is highly **network** and **disk** I/O intensive, it was necessary to perform simulation of job execution before submitting any jobs.

Conclusion

- The task of the mass data processing for BM@N Run 8 was **successfully completed**.
- It was the first usage of the **whole DIRAC infrastructure** in JINR for processing experimental data.
- The process of BM@N data processing was **repeated several times** after changes in reconstruction algorithms: **5 major** and **7 minor** processing campaigns were completed.
- Around **3 PB of data transfers** related to Run8 were performed so far.
- **Requirements** to CPU **grows!** Especially for DigiToDst jobs. Current processing takes at least 300 hours to complete on the whole infrastructure.

	Tier1	Tier2	NICA	Govorun	Total
Executed jobs (k)	194	133	21	15	363
Normalized CPU time (kHS06 days)	362	165	37	29	593
Walltime (years)	54	24	4	4	86