SIMULATION OF DATA PROCESSING FOR THE BM@N EXPERIMENT OF THE NICA COMPLEX

D. Priakhina[1,2,3 a], V. Korenkov[1,2,3], K. Gertsenberger[1], V. Trofimov[1]

[1] *Joint Institute for Nuclear Research, Russia, Moscow region, Dubna, 141980, 6 Joliot-Curie*

[2] *Dubna State University, Russia, Moscow region, Dubna, 141980, 19 Universitetskaya*

[3] *Plekhanov Russian University of Economics, Russia, Moscow, 115093, 36 Stremyanny per.*

E-mail: [a] pryahinad@jinr.ru

The paper considers the application of a software complex for data processing simulation in computing systems. The BM@N experiment data storage and processing system of the NICA complex is used as a simulated infrastructure. Simulation is performed in order to obtain recommendations on the organization of experimental data processing during a session with available allocated resources. The paper presents the results of simulating the processing of data that will be received during the BM@N experiment session, with several scenarios for distributing job flows across existing processing centers. In addition, some recommendations for organizing experimental data processing are proposed. The status of the work and future plans for the development of the software complex are formulated.

Keywords: simulation, data center, BM@N experiment, NICA complex

Daria Priakhina, Vladimir Korenkov, Konstantin Gertsenberger, Vladimir Trofimov

1. Introduction

A software complex for simulating the processing of data coming from the experimental facility of the NICA complex is being developed at the Meshcheryakov Laboratory of Information Technologies of the Joint Institute for Nuclear Research [1]. The software complex allows finding out how the data storage and processing system will work with the available computing power, as well as calculating the load on computing farms and communication links with the specified parameters of data flows and job flows. Unlike the previously developed simulation program [2,3] the software complex consists of a database, a module for setting the simulated structure and equipment configurations, a stable core for the simulation of data transmission and processing, a module for presenting results in the form of graphs. The simulation core is implemented on top of an approach based on the representation of information processes as byte streams [4].

Currently, work on the data processing simulation of the BM@N experiment is in progress [5]. The simulation of the facility session is performed for various hardware configurations, data flow and job flow parameters, scenarios for running data processing jobs [6]. The simulation results presented in this article made it possible to draw conclusions that are the base of some formulated recommendations for organizing the process of experimental data processing during the BM@N experiment session with available allocated resources.

Requirements for the improvement of the software complex are formed in the process of its using. The main aim of the improvement is its further application for data processing simulation in any data storage and processing centers.

2. Simulation input parameters

Before starting to simulate data processing, it is necessary to determine the structure of the data storage and processing center, the configurations of various equipment, as well as the characteristics of data flows and job flows. All these parameters are input data for simulation.

2.1 Distributed computing system for storing and processing data

The simulated distributed data storage and processing system with the equipment parameters is shown in Figure 1.
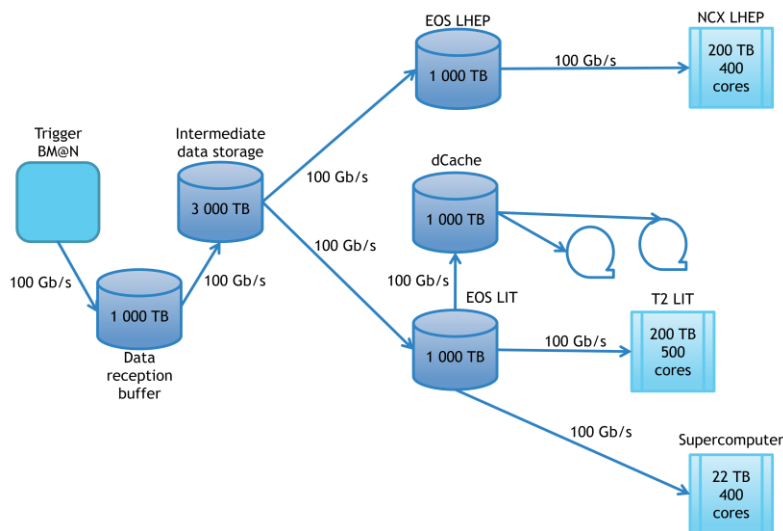

Figure 1. Scheme of the simulated distributed computing system

The system consists of two centers and includes several levels of data storage. The data of particle collision events with a fixed target (raw data), which are selected by the trigger (Trigger BM@N), is first written to the Data reception buffer, then it arrives at the Intermediate data storage level, after which the full volume of raw data is sent to the local buffers (EOS LHEP and EOS LIT) of the respective data processing centers. All storage devices have a limited volume, and data

transmission channels have a limited bandwidth. Data processing jobs are executed on the compute nodes of the above processing centers (NCX LHEP, T2 LIT and Supercomputer). Data processing jobs arrive at the compute nodes and are performed if there are free slots, i.e. processor cores. If all slots are occupied, the job goes to the queue and waits until some slot is free. Each job needs a certain amount of data to execute. Data is transferred from the storage to the compute node before the job starts.

2.2 BM@N Run workload and job stream

A 30-day session of the BM@N experiment is considered in the article. It is assumed that the session will last continuously, except for minute breaks between runs. Each run lasts 2 minutes. Experimental data is written to files during each run. The recording is carried out continuously up to the maximum size of each file (35 GB). The speed of data selection by the BM@N trigger is 10 000 events per second. The average size of a single event is about 0.2 MB. The maximum planned volume of experimental data per session, taking into account the described characteristics, will be approximately 350 TB.

Raw binary data, which is received from the installation, is subject to processing, namely, digitization (the final format is digit) and conversion to reconstruction data (DST format), after which physical analysis of experimental data is performed. In addition to processing experimental data, it is planned to process simulation data (sim), which will also be converted into reconstruction data (DST) for subsequent physical analysis.

There are several classes of jobs in accordance with the described process of data processing. The characteristics of classes are described in Table 1. Job flows for the simulation of data processing were formed on the basis of the data from the table.

Table 1. Characteristics of job classes

№	Class	Avg. event processing time on one processor (ms)	Average amount of input (GB)	Number of events in the file (1 file = 1 job)	Avg. job execution time (s)	Average amount of output (GB)	Number of jobs
1	*RawToDigit*	675	35	175 000	118 125	1	15 552
2	*DigitToDst*	290	1	175 000	50 750	1	15 552
3	*GenToSim*	60	0.6	175 000	10 500	8	300
4	*SimToDst*	30	8	175 000	5 250	1	300
5	*DstToAna*	10	1	175 000	1 750	0.1	1 000

3. Simulation

In the example considered below, simulation is performed in order to provide recommendations on the organization of experimental data processing during the session with available allocated resources [fig. 1]. First of all, there were simulated three scenarios, where the data processing jobs described in Table 1 are distributed in different ways across the compute nodes [tab.2].

Table 2. Distribution of data processing jobs (in %) by compute nodes

№	Class	Scenario 1			Scenario 2			Scenario 3		
		NCX LHEP	T2 LIT	Super-computer	NCX LHEP	T2 LIT	Super-computer	NCX LHEP	T2 LIT	Super-computer
1	*RawToDigit*	50	15	35	80	20	-	-	10	90
2	*DigitToDst*	50	15	35	80	20	-	-	10	90
3	*GenToSim*	-	20	80	-	-	100	100	-	-
4	*SimToDst*	-	20	80	-	-	100	100	-	-
5	*DstToAna*	-	20	80	70	10	20	80	20	-

The results of simulating three scenarios showed that one of the computing resources could be fully allocated for processing simulated data up to reconstruction data (GenToSim and SimToDst jobs). As for the processing of experimental data, at best only 20% of raw data can be processed to digit data, and less than 10% of jobs on processing data to reconstruction data will have time to complete before the end of the session, which lasts 720 hours. Figure 2 shows graphs reflecting the number of completed experimental data processing jobs (RawToDigit and DigitToDst jobs) on computing resources in accordance with Scenario 2. Figure 2 does not show graphs of completing DstToAna jobs since the number of completed jobs is too small and is approximately 1%. This is due to the full load of all computing resources.
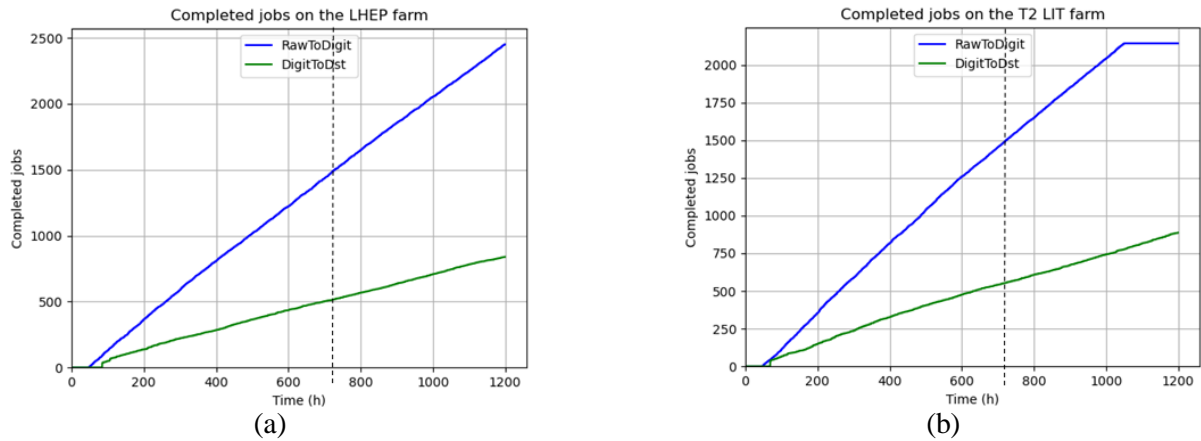

Figure 2. Completed RawToDigit and DigitToDst jobs on the LHEP farm (a) and T2 LIT farm (b)

Surely, the result is unsatisfactory. Such a small amount of experimental data can be processed by the end of the session. After the end of the session, one will have to wait a few more months before the end of processing all the data acquired from the experiment. It may be noticed that a large amount of time is required to fully process a single data file for RawToDigit and DigitToDst jobs [tab. 1]. In this regard, it was decided to simulate a scenario similar to Scenario 2, in which express file processing will be added. Express processing will be run simultaneously with full processing. Express processing consists in processing 1% of the file in RawToDigit jobs to obtain preliminary results.

Figure 3 shows the results of simulating the scenario with express processing. The graphs show the number of completed jobs of the full (RawToDigit jobs) and express (RawToDigit Express jobs) processing of experimental data. The graphs illustrate that express data processing is completed much faster. It will allow one to obtain preliminary results every hour throughout the experiment, and to conclude that the measurements obtained are correct.
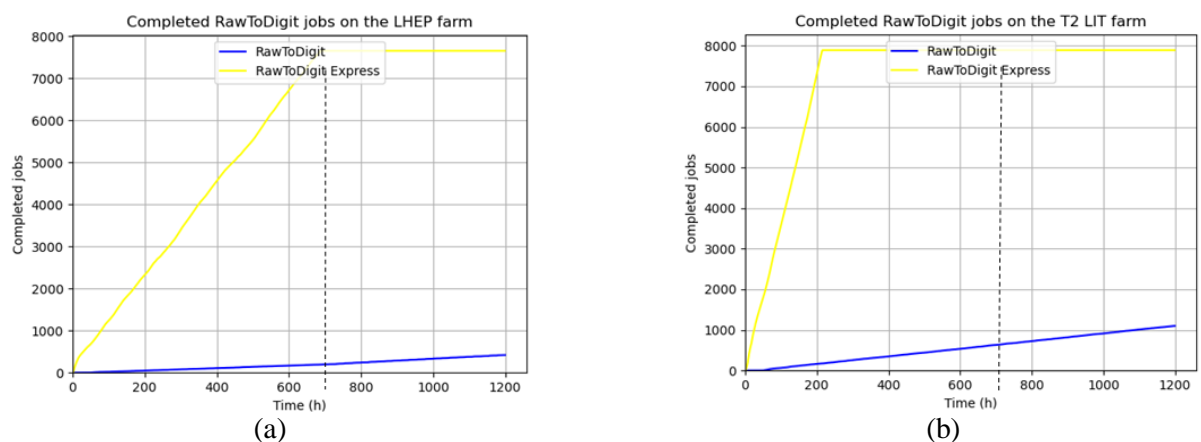

Figure 3. Free slots on the LHEP farm (a) and T2 LIT farm (b)

4. Conclusion

A software complex to simulate the processing of data, which comes from the experimental facility of the NICA complex, is being developed at MLIT JINR. The presented simulation results of the data processing of the BM@N experiment showed that with the described parameters of equipment, data flows and jobs flows, no more than 20% of raw data could be processed by the end of the session, lasting 30 days. It is also proposed to perform express data processing simultaneously with full processing, which will allow obtaining preliminary results every hour throughout the experiment. It is possible to conclude that the measurements obtained are correct.

Requirements for the improvement of the software complex are formed in the process of its using. The main aim of the improvement is its further application for data processing simulation in any data storage and processing centers. Thus, at the next stage, it is planned to develop a module for launching jobs similar to the pilot, as well as to conduct computational experiments taking into account the fact that the equipment does not have absolute reliability, i.e. to calculate the probability of equipment failure and recovery time.

5. Acknowledgement

References

[1] Kekelidze V., Kovalenko A., Lednicky R., Matveev V., Meshkov I., Sorin A. and Trubnikov G. Status of the NICA project at JINR // EPJ Web Conf. 2017. V. 138. P. 01027

[2] Korenkov V.V., Nechaevskiy A.V., Ososkov G.A., Pryahina D.I., Trofomov V.V., Uzhinskiy A.V. Simulation concept of NICA-MPD-SPD Tier0-Tier1 computing facilities // Physics of Particles and Nuclei Letters, 2016, 13(5), pp. 693–699

[3] Nechaevskiy A., Ososkov G., Pryahina D., Trofimov V., Li WD. Simulation approach for improving the computing network topology and performance of the China IHEP Data Center // 23rd international conference on computing in high energy and nuclear physics (CHEP 2018), EPJ Web of Conferences, 2019, 214(08018), DOI: 10.1051/epjconf/201921408018

[4] Korenkov V., Nechaevskiy A., Ososkov G., Priakhina D. and Trofimov V. A Probabilistic Approach to the Simulation of Data Processing Centers // EPJ Web Conf. 2020. V. 226. P. 03012

[5] Kapishin M. Studies of baryonic matter at the BM@N experiment (JINR) // Nuclear Physics A. 2019. V. 982. P. 967

[6] Priakhina D.I., Trofimov V.V., Ososkov G.A., Gertsenberger K.V. Data center simulation for the BM@N experiment of the NICA project // AIP Conference Proceedings 2377, 040007 (2021), https://doi.org/10.1063/5.0063338