SAMARA UNIVERSITY

SPD NICA

VI SPD Collaboration Meeting and Workshop on Information Technology in Natural Sciences

Digital Twin of SPD Online filter

DARIA PRIAKHINA

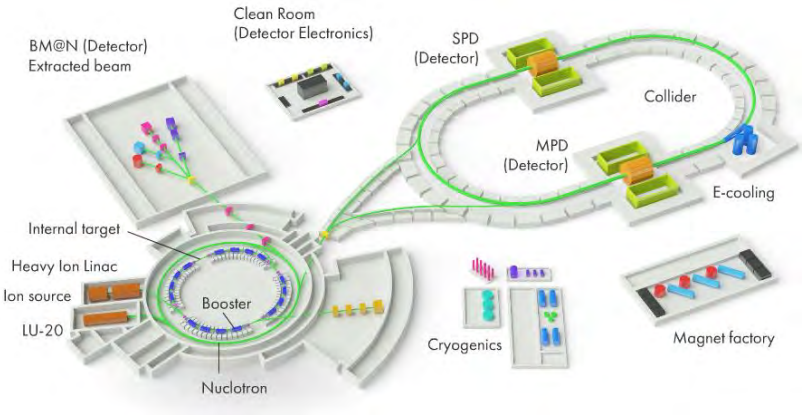
MESHCHERYAKOV LABORATORY OF INFORMATION TECHNOLOGIES

JOINT INSTITUTE FOR NUCLEAR RESEARCH

26.10.2023

Introduction

Nuclotron-based Ion Collider fAcility (NICA, JINR, Dubna, Russia)



- Large amounts of data per second are generated during the experiments.
- Efficient computing systems are required to store and process data in a relatively short time!
- The system should provide enough performance through balanced loading of resources.

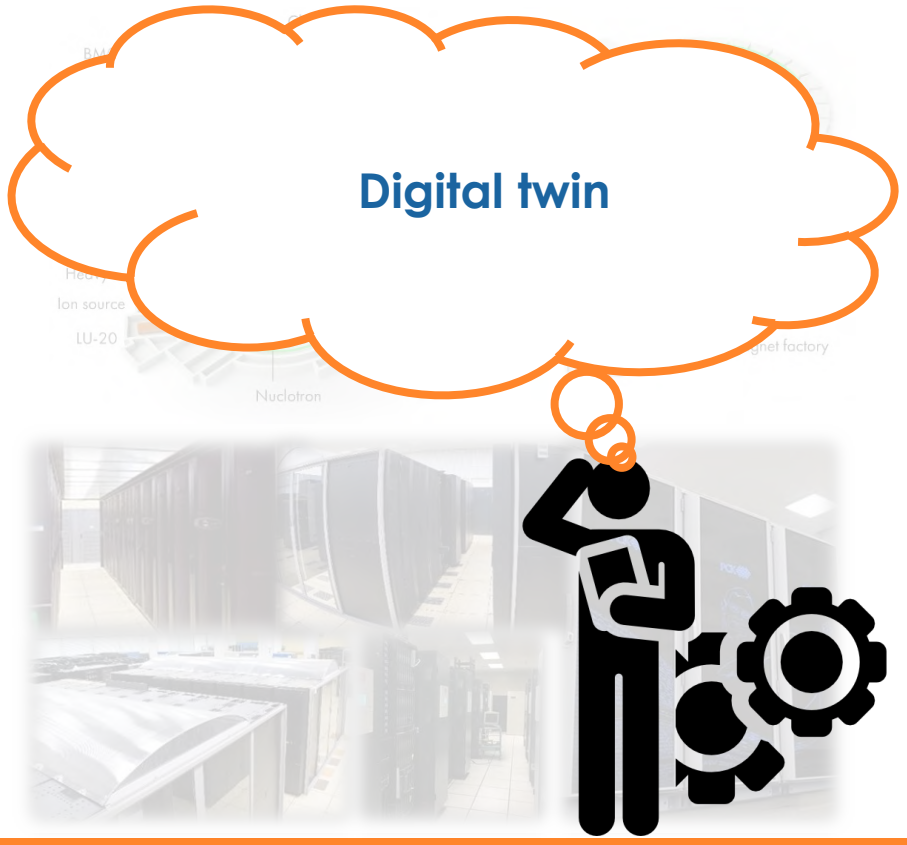
Introduction

Nuclotron-based Ion Collider fAcility (NICA, JINR, Dubna, Russia)

- Large amounts of data per second are generated during the experiments.
- Efficient computing systems are required to store and process data in a relatively short time!
- The system should provide enough performance through balanced loading of resources.

Introduction

Nuclotron-based Ion Collider fAcility
(NICA, JINR, Dubna, Russia)



- Modeling of complex computing systems.
- Describes the system.
- Reflects the processes taking place in the system.
- Testing of the system with different variants of equipment parameters, data flows and jobs.

Priakhina D., Korenkov V. The relevance of creating a digital twin for managing distributed data acquisition, storage and processing centers (accepted in print)

Digital twin (DT)

Real-time operation throughout the entire computing infrastructure life cycle.

COMPUTER MODEL

INPUT DATA

- Architecture and hardware parameters of computing infrastructure.
- Characteristics of data flows and jobs flows.

FUNCTIONAL PURPOSE

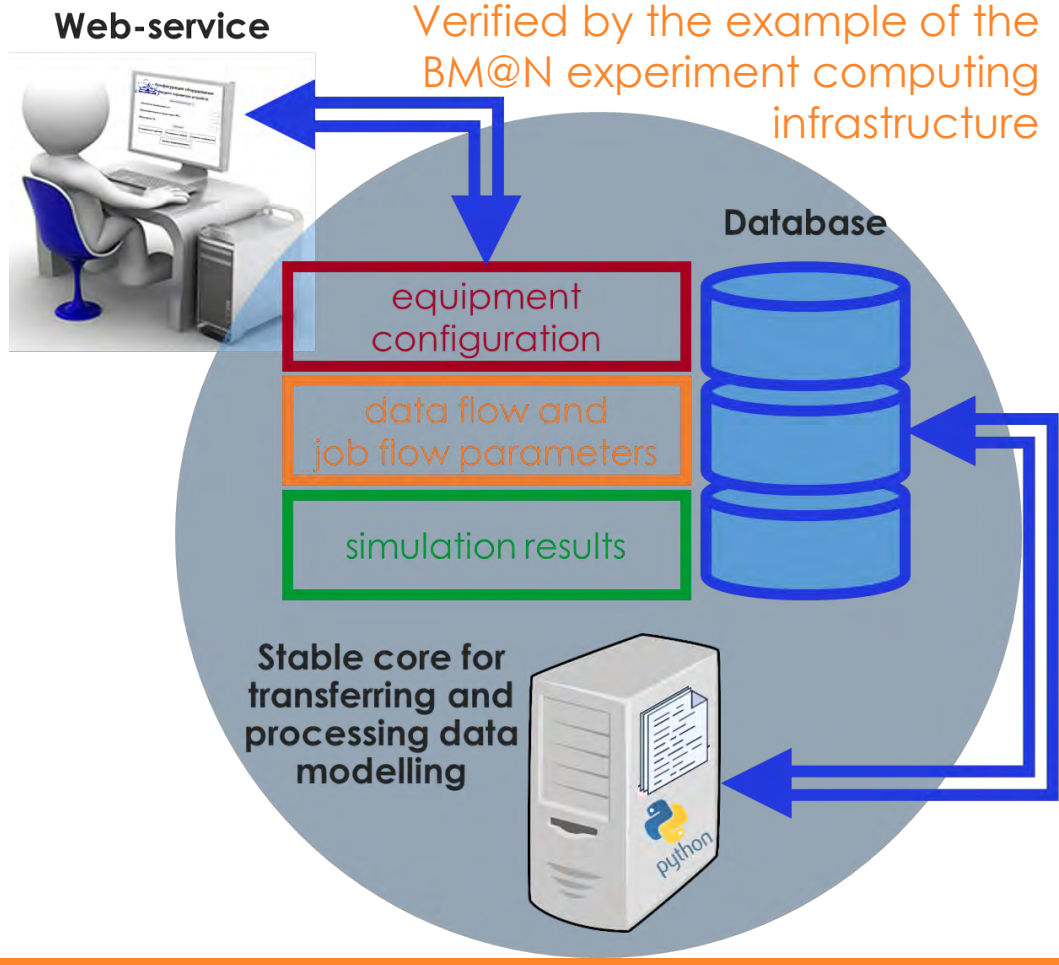
- Designing of computing infrastructure.
- Analysis of the efficiency and reliability of computing infrastructure.
- Testing scaling scenarios based on data flows and jobs flows requirements.
- Assessment of the required amount of resources for specific tasks.
- Checking jobs flows management strategies.

Priakhina D., Korenkov V., Trofimov V. A method of constructing digital twins for solving problems of effective management and development of distributed data acquisition, storage and processing centers (accepted in print)

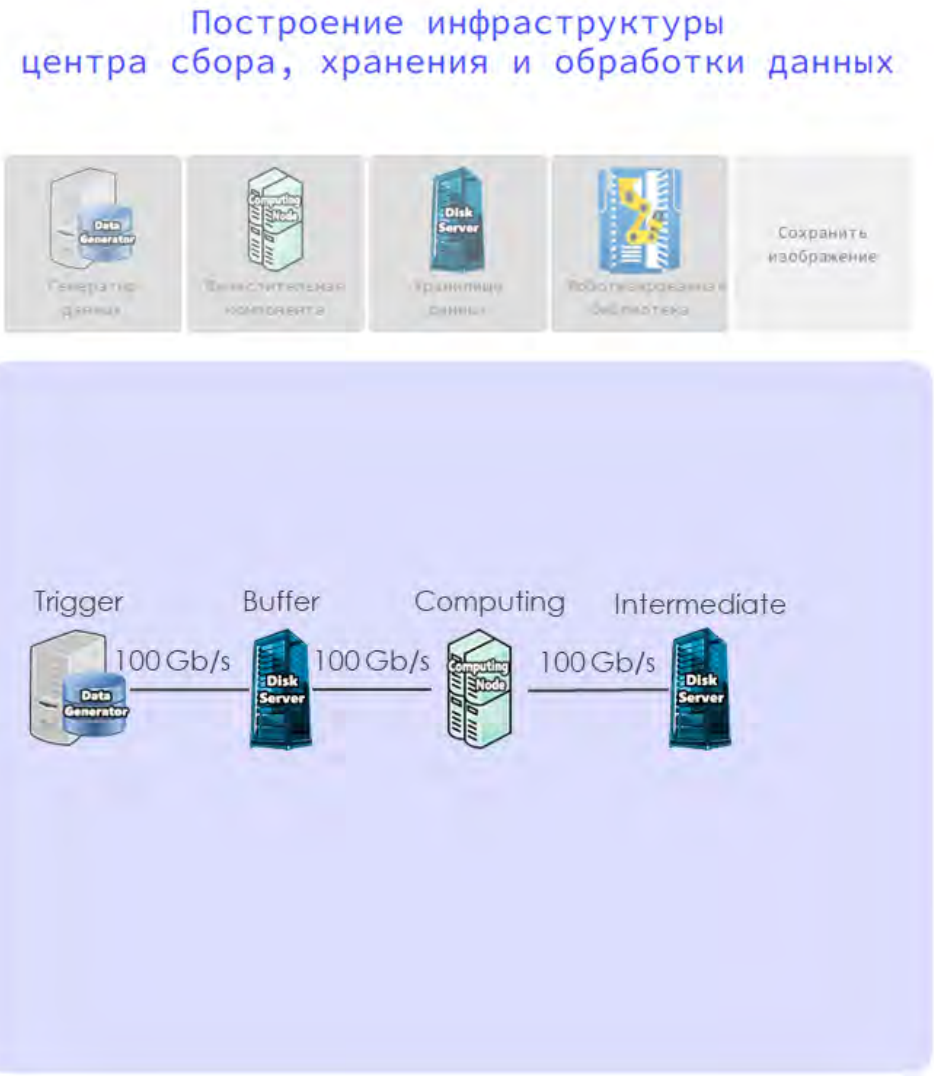
Software complex for creating digital twins

Modeling core

- Universal – applicable for modeling computing infrastructure without changing the program code.
- Probabilistic distributions are taken into account when forming data flows, jobs flows, and criteria for the functioning of equipment.
- Used for design tasks, system scaling during operation, searching for problem areas when data flows and jobs flows change.



Functionality of the web service



Create a digital twin

Building the computing infrastructure

- Setting the equipment parameters.
- Setting characteristics of data flows and jobs flows.

The prototype of the web service has not yet been localized.

Functionality of the web service

Configuration of computing infrastructure scaling scenarios

Parameters for modeling:

- experiment name;
- description;
- duration of work;
- speed up of modelling;
- objects and events for logging.

Functionality of the web service

Информация об эксперименте

Дата создания: 7 февраля 2023 г. 10:36

Название эксперимента

Test 1

Описание эксперимента

Поиск оптимального количества ресурсов для хранения данных

Параметры моделирования

- *Продолжительность работы моделируемой инфраструктуры - 800 ч.*
- *Ускорение процесса моделирования в 1000 раз.*

Параметры логирования

- *Объекты моделируемой инфраструктуры*
  - ☑ *Хранилища данных*
  - ☐ *Вычислительные компоненты*
  - ☐ *Каналы связи*
- *События*
  - ☑ *Генерация данных*
  - ☑ *Потери данных*
  - ☑ *Работа с файлами*
  - ☐ *Генерация, запуск, выполнение задач*

Посмотреть результаты

Выбрать другой эксперимент

Базовая конфигурация

Хранилища данных

Название	Описание	Объем (ТБ)
trigger	Trigger BM@N	10000,0
buffer	Data receptipon buffer	5400,0
eoslhep	Main storage LHEP	1000,0
eoslit	Main storage LIT	1000,0
dcach	pp	1000,0

Вычислительные компоненты

Название	Описание	Количество ядер
t2lit	LIT T2 farm	500
ncxlhep	LHEP main farm	1200
super	Govorun	190

Каналы связи

Название	Описание	Пропускная способность (Гб/с)
raw0	trigger - buffer	100,0
raw1	buffer - lhep	10,0
raw2	buffer - lit	10,0
compute0	lhep - farm lhep	10,0
compute1	lit - Govorun	10,0
compute2	lit - farm lit	10,0
dataeosLhepLit	eoslhep - eoslit	10,0
dataeosLitLhep	eoslit - eoslhep	10,0

Добавить модификацию

Starting the digital twin

№	Статус	Дата обновления			
16	NEW	9 марта 2023 г. 14:52	Просмотр	Запуск	Результаты
15	DONE	10 марта 2023 г. 10:18	Просмотр	Запуск	Результаты

Simultaneous run of all modifications is possible

Functionality of the web service

The digital twin results

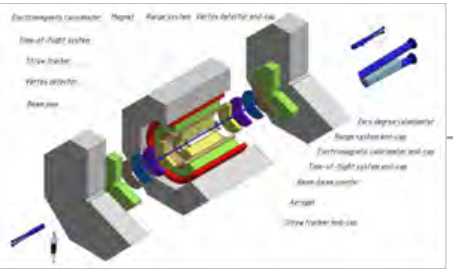
Available for viewing:

- data storage load volume;
- using cores on computing components;
- load on communication links;
- job queues, the number of completed jobs;
- distribution of files in storages.

Digital Twin of SPD Online filter

First experience

Raw data rate
20 GB/s



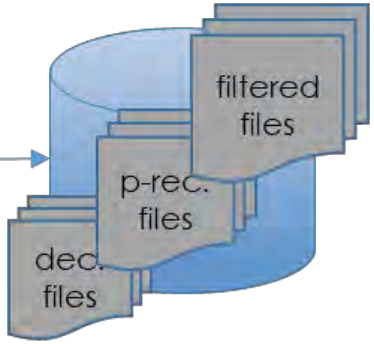
1 raw event = 7 KB

Buffer for
data received
from the detector

1 raw file = 4 GB

Computing
resources

Storage for
intermediate
data



Processing raw file
to filt. file: **10 min**

1) Decryption:
raw file → dec. file — 1 dec. file = 4 GB

2) Partial reconstruction:
dec. file → p-rec. file — 1 p-rec. file = 8 GB

3) Filtering:
p-rec → filt. file — 1 filtered file = 450 MB

Experiment duration: 24 hours

To calculate:

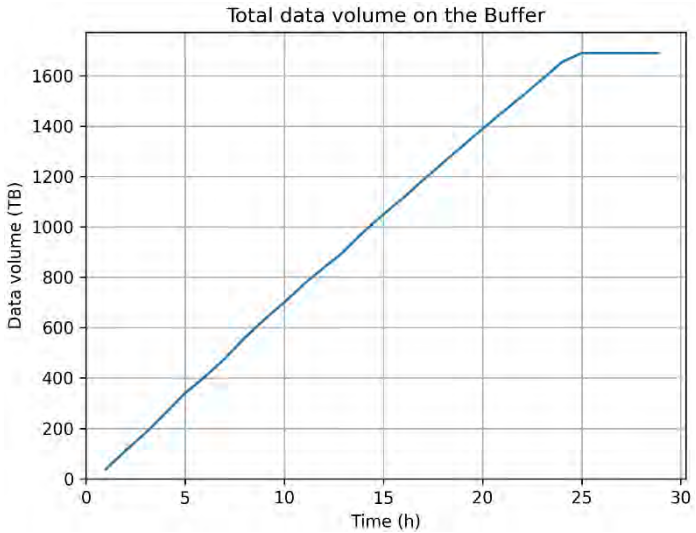
- ? data storage volumes;
- ? network load;
- ? load of computing resources etc.

Digital Twin of SPD Online filter

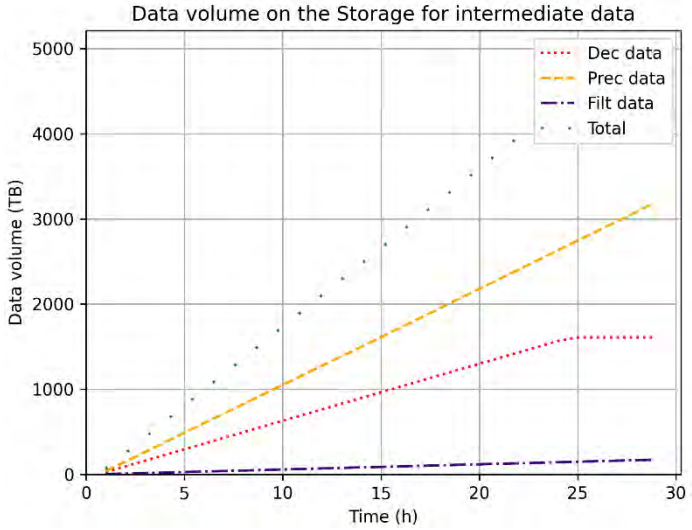
First experience

Results 1. Constant data generation

Data storages



~ 1 700 TB on the Buffer



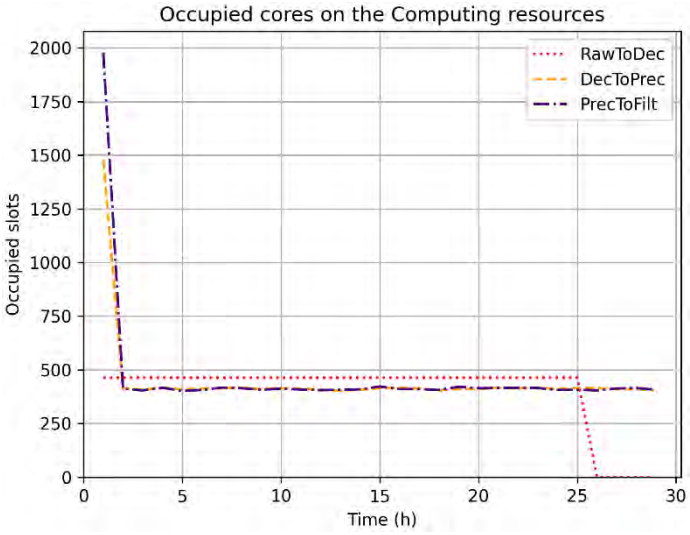
~ 5 000 TB on the Storage for intermediate data

Digital Twin of SPD Online filter

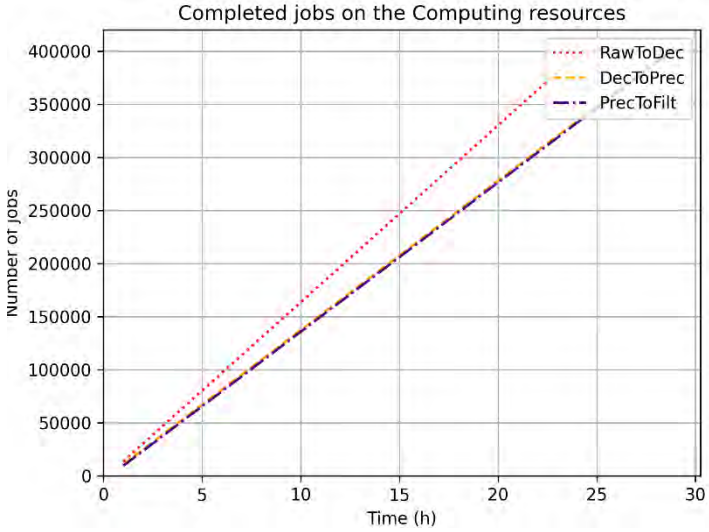
First experience

Results 1. Constant data generation

Computing resources



~ 1 500 cores are occupied by jobs



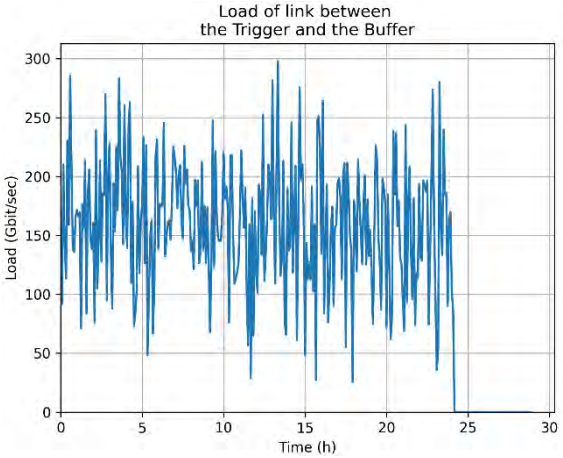
~ 400 000 jobs will be done for 30 hours

Digital Twin of SPD Online filter

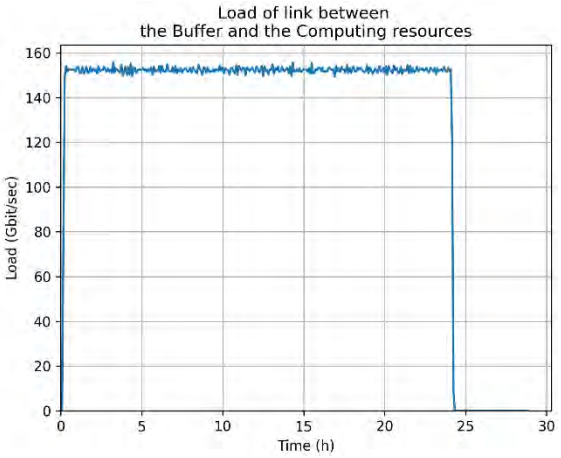
First experience

Results 1. Constant data generation

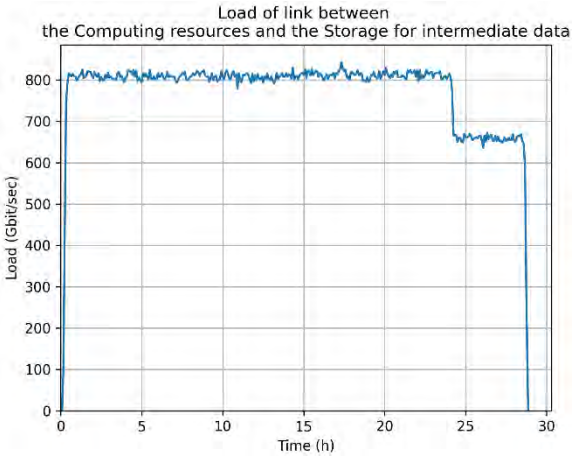
Network



~ 150 Gbit/sec between the Trigger and the Buffer



~ 150 Gbit/sec between the Buffer and the Computing resources



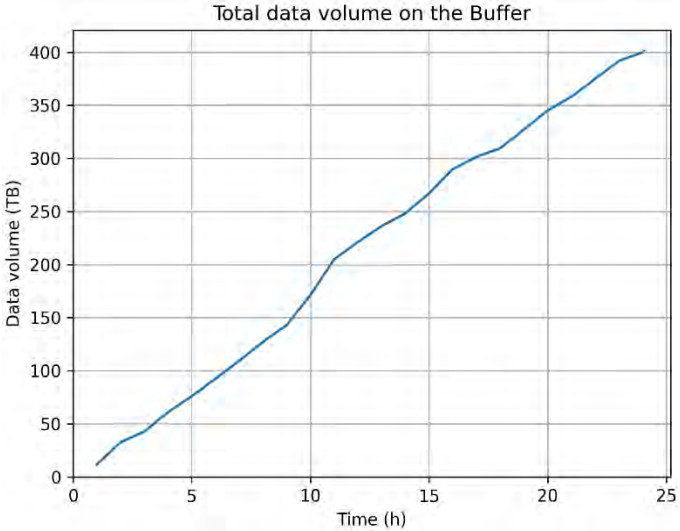
~ 800 Gbit/sec between the Computing resources and the Storage for intermediate data

Digital Twin of SPD Online filter

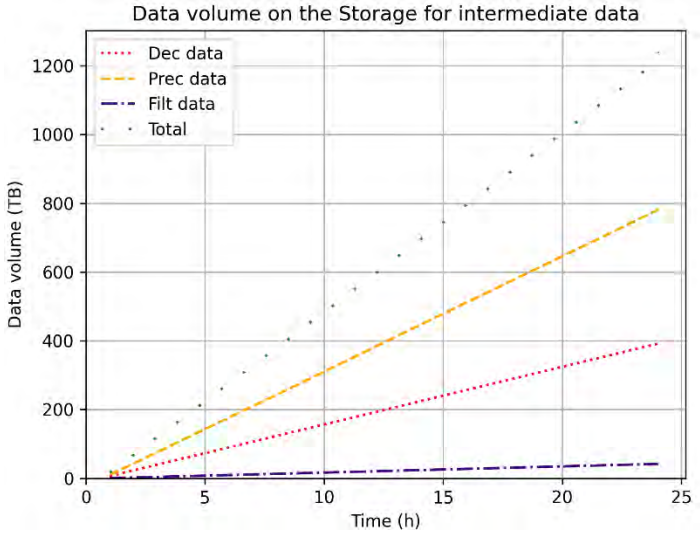
First experience

Results 2. Data generation efficiency – 20%

Data storages



~ 400 TB on the Buffer



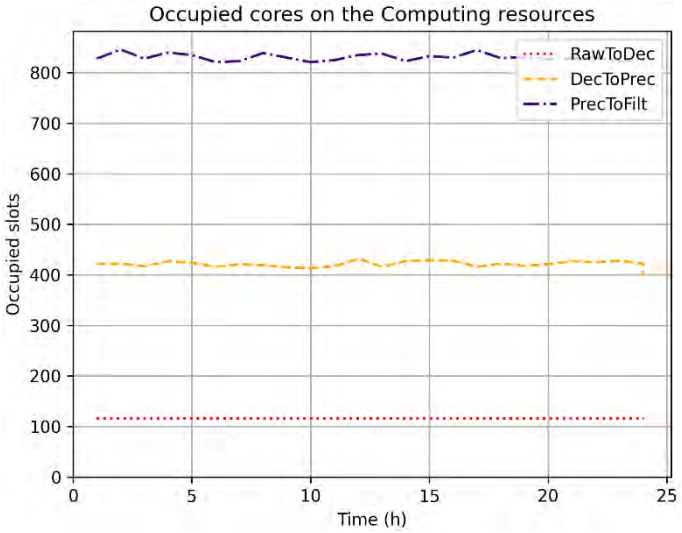
~ 1 200 TB on the Storage for intermediate data

Digital Twin of SPD Online filter

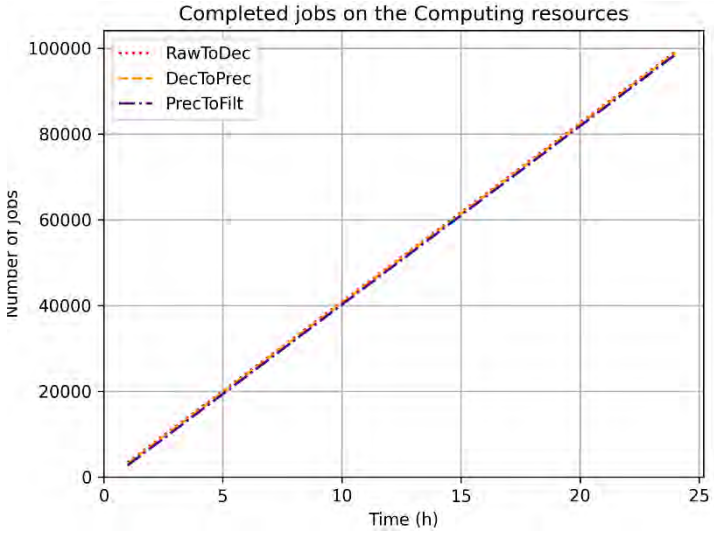
First experience

Results 2. Data generation efficiency – 20%

Computing resources



~ 1 500 cores are occupied by jobs



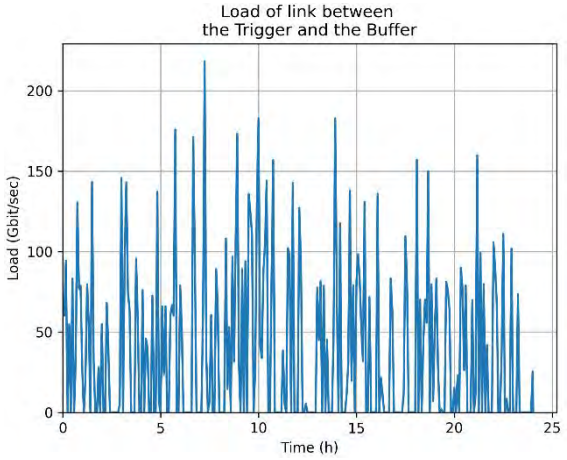
~ 100 000 jobs will be done "on the fly"

Digital Twin of SPD Online filter

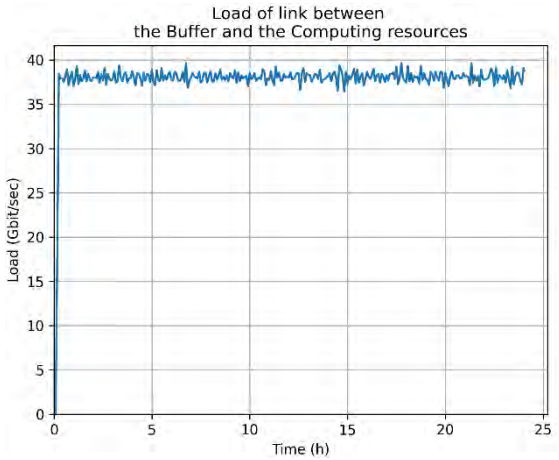
First experience

Results 2. Data generation efficiency – 20%

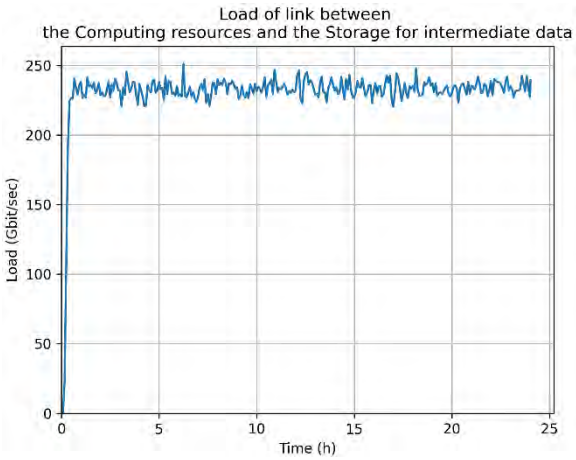
Network



~ 50 Gbit/sec between the Trigger and the Buffer



~ 40 Gbit/sec between the Buffer and the Computing resources



~ 250 Gbit/sec between the Computing resources and the Storage for intermediate data

Conclusions

- Software complex can be used for create digital twins of SPD computing infrastructure.
- Digital twin of SPD online filter can help calculate:
  - the required storages volumes;
  - time for data processing;
  - use of computing resources;
  - loading of communication links, etc.
- Future plans:
  - adding the probabilities of various events in the system;
  - running digital twins for different scenarios of experiment.

SAMARA UNIVERSITY

SPD NICA

VI SPD Collaboration Meeting and Workshop on Information Technology in Natural Sciences

Thank you for the attention!

D. PRIAKHINA

MESHCHERYAKOV LABORATORY OF INFORMATION TECHNOLOGIES

JOINT INSTITUTE FOR NUCLEAR RESEARCH