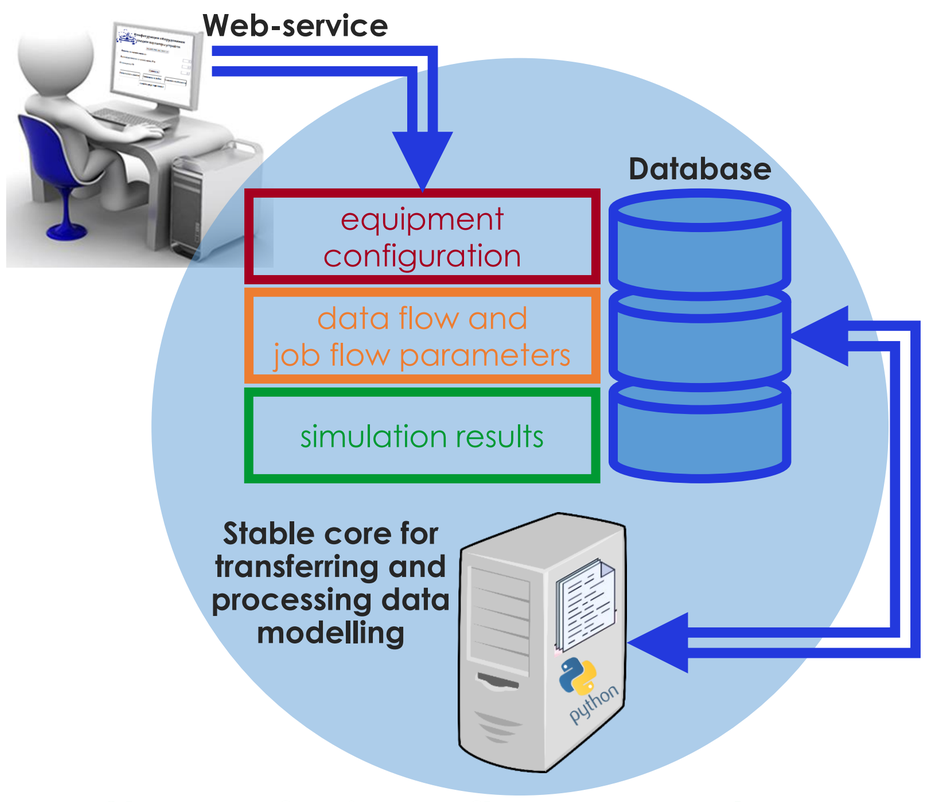
Software complex for creating digital twins of large-scale distributed computer systems for megascience projects

D. Priakhina, V. Korenkov, V. Trofimov
pryahinad@jinr.ru

Digital twin (DT) – a computer model that can be used in real time throughout the entire life cycle of distributed data acquisition, storage and processing center (DDC) [1, 2].

Software complex architecture



Functional purposes of digital twins

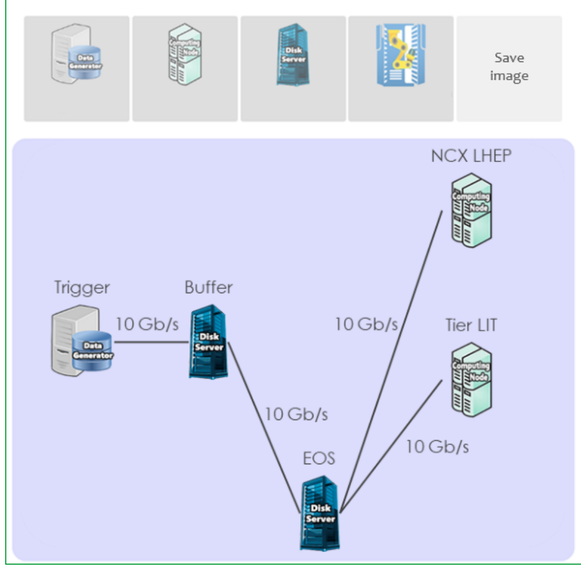
- Designing a DDC.
- Analyzing the efficiency and reliability of DDC functioning.
- Testing scaling scenarios taking into account the requirements for data flows and job flows.
- Assessment of the required amount of resources for specific tasks.
- Checking strategies for managing job flows.

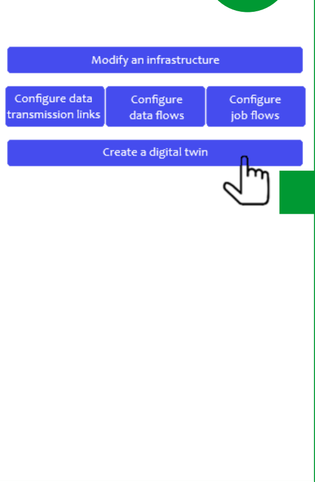
Distinctive features of modelling core

- **Universal** – applicable for modelling any data center without changing the program code.
- **The initial data**: the architecture and hardware parameters of the DDC equipment; the characteristics of data flows and job flows.
- **Probabilistic distributions** are taken into account when forming data flows, job flows, and criteria for the functioning of equipment.
- **Used for** design tasks, data center scaling during operation, searching for problem areas when data flows and jobs flows change.

Stages of using software complex

1 Building an infrastructure for data acquisition, storage and processing centers





2 Configuring the simulation parameters:

- setting the duration of the DDC operation;
- adding probabilistic events that may occur in the system (equipment failure, changes in the amount of computing resources, etc.)
- adding specify objects and events for logging;
- Setting additional possible modifications besides the basic hardware configuration that was set when creating the DDC infrastructure.

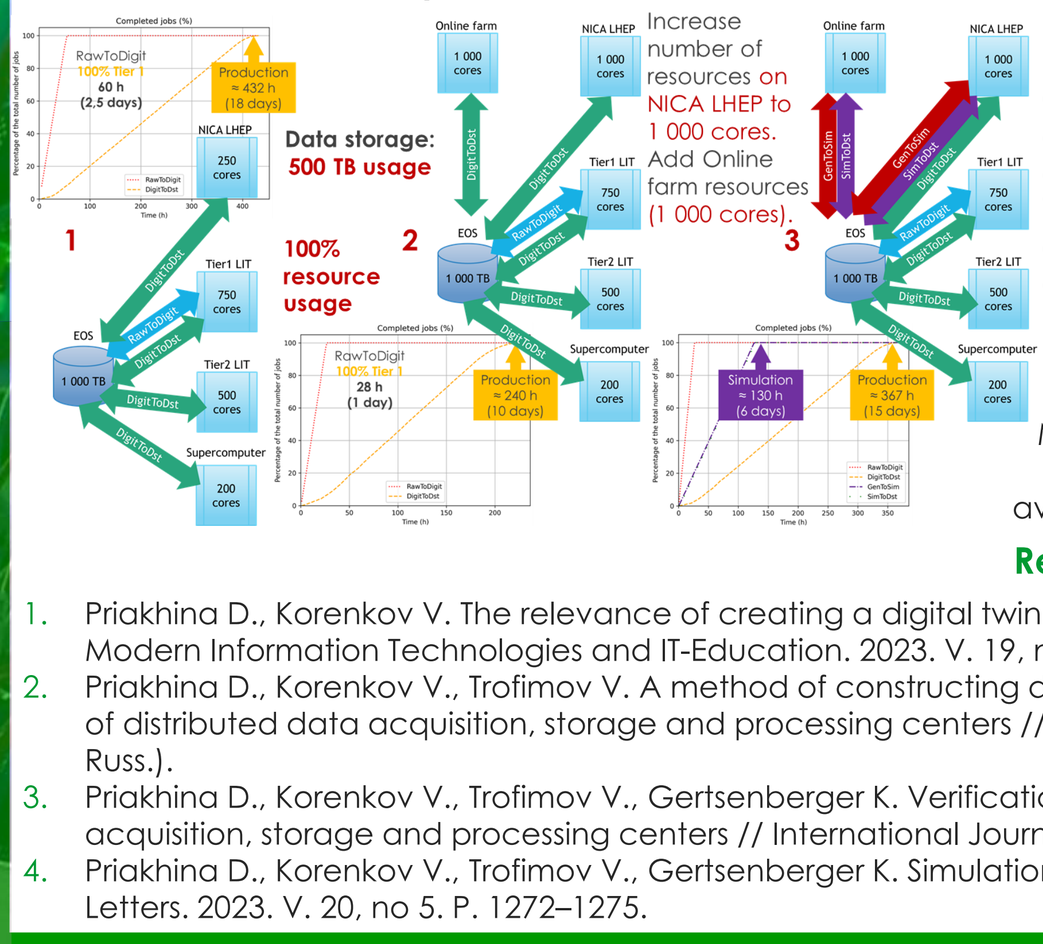
3 DT launch.

4 Analysis of the DT functioning results:

- data storages load;
- using computing resources;
- baud rate;
- job queues, completed jobs;
- distribution of files in storages.

Implementation tools

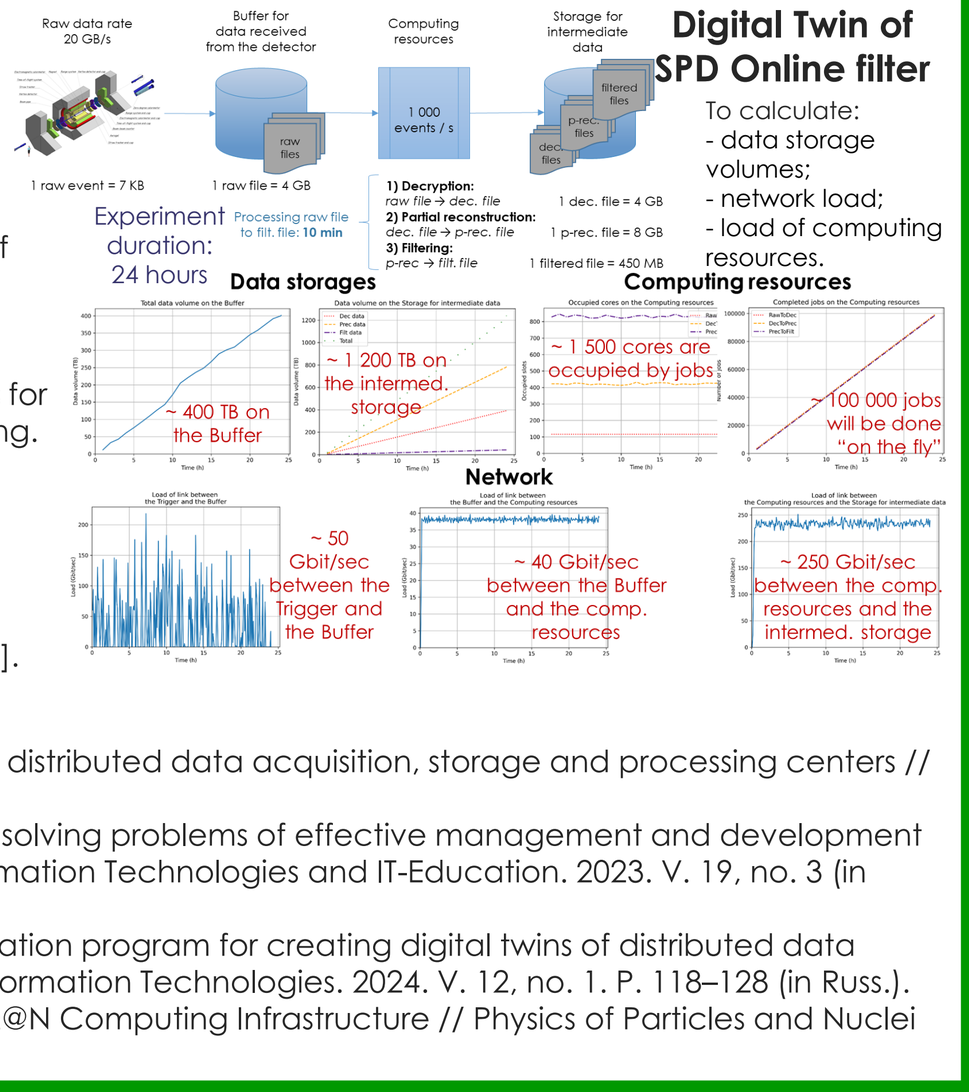
Digital Twin of BM@N computing infrastructure for data production



Comparison of different computing infrastructure configurations for data processing.

More detailed information is available in [3, 4].

Software complex application



Digital Twin of SPD Online filter

To calculate:
- data storage volumes;
- network load;
- load of computing resources.

References

1. Priakhina D., Korenkov V. The relevance of creating a digital twin for managing distributed data acquisition, storage and processing centers // Modern Information Technologies and IT-Education. 2023. V. 19, no. 3 (in Russ.).
2. 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 // Modern Information Technologies and IT-Education. 2023. V. 19, no. 3 (in Russ.).
3. Priakhina D., Korenkov V., Trofimov V., Gertsenberger K. Verification of the simulation program for creating digital twins of distributed data acquisition, storage and processing centers // International Journal of Open Information Technologies. 2024. V. 12, no. 1. P. 118–128 (in Russ.).
4. Priakhina D., Korenkov V., Trofimov V., Gertsenberger K. Simulation Results of BM@N Computing Infrastructure // Physics of Particles and Nuclei Letters. 2023. V. 20, no 5. P. 1272–1275.

MESHCHERYAKOV LABORATORY OF INFORMATION TECHNOLOGIES