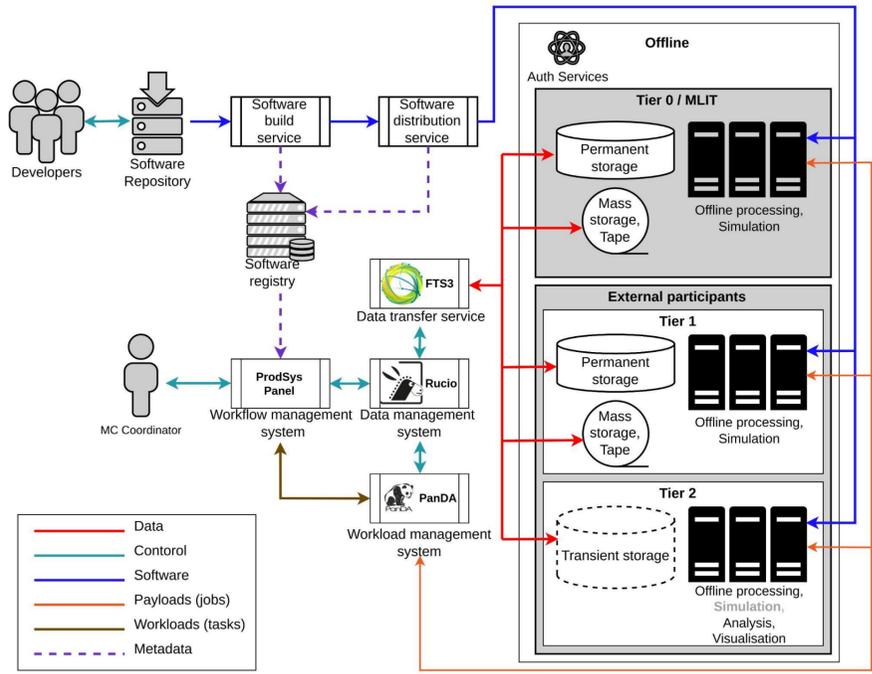
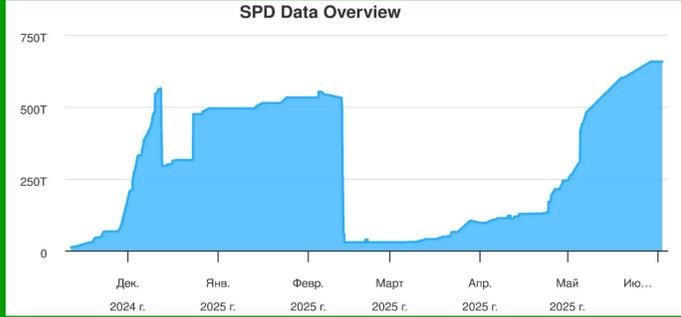


Offline Data Processing System for the SPD Experiment

A. Petrosyan, A. Konak, N. Monakov

SPD (Spin Physics Detector) is a universal detector of the NICA collider (Nuclotron-based Ion Collider fAcility), built at the Joint Institute for Nuclear Research (Dubna), and designed to study the spin structure of the proton, deuteron and other spin phenomena with polarized beams of protons and deuterons. The expected volume of data collected from the SPD experimental facility alone is measured in petabytes per year, to this volume it is also necessary to add secondary data generated during processing and data that will arise during the modeling of the studied physical processes.



Offline data processing system is a combination of systems and services that allows you to create a single computing environment from geographically distributed heterogeneous resources.

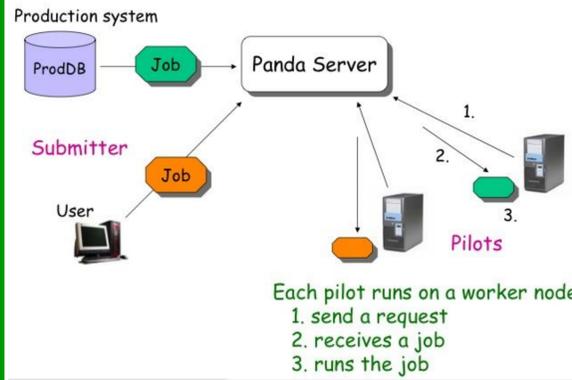
It includes the following components:

- workload management system (WMS) – PanDA,
- workflow management system (WFMS) – ProdSys Panel,
- data management system (DMS) – Rucio,
- data transfer service (DTS) – File Transfer Service 3 (FTS3),
- information system (IS) – Computing Resource Information Catalog (CRIC)

SPD Identity and Access Management (SPD IAM) provides authentication to all services and systems.

Workload Management System

PanDA (Production and Distributed Analysis) is a job management system designed for processing large amounts of data in distributed computing environments such as GRID.



PanDA is used in areas that require large-scale computing:

- Processing of experimental data (event reconstruction, modeling, analysis).
- Optimization of resource usage (grid computing, clouds, supercomputers).
- Automatic distribution of tasks between multiple computing centers.

Auth System



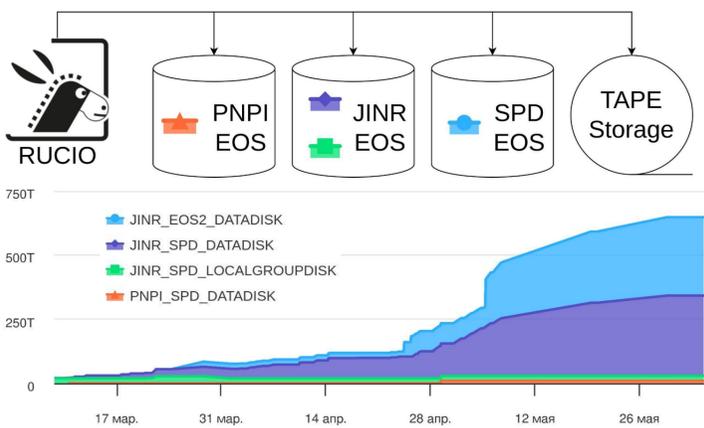
SPD IAM allows you to log in to a third-party application or system using an access token and an ID token obtained during authorization in IAM or JINR SSO.

In the foreseeable future, each member of the collaboration will be registered there and will use this system to log in to other systems and applications. This allows for a single entry point and makes it easier to account for user activity, as well as access control.

INDIGO IAM for SPD

Data Management System

Rucio is an open source software package that provides scalable functionality for organizing, managing, and accessing data. Rucio organizes a single namespace, realizing the possibility of data management, interacting with the entire storage space as one.



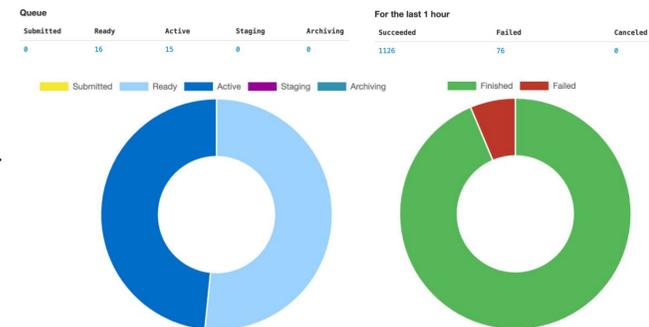
The Computing Resource Information Catalog (CRIC) is a system designed to manage and provide information about computing resources used in distributed computing infrastructures.

VO	NICA Site	State	Tier	Site	Storage Units
spd	JINR-SPD	ACTIVE	T0	JINR	SPD-JINR-DATA
spd	PNPI-SPD	ACTIVE	T1	PNPI	SPD-PNPI-DATA
spd	SSAU-SPD	ACTIVE	T2	SSAU	

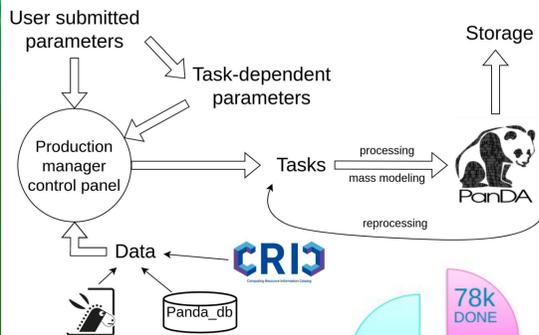
Data Transfer Service

FTS3 (File Transfer Service 3) is a service for reliable and managed copying of large amounts of data between distributed storages. FTS3 is used:

- for mass data transfers between grid sites;
- automatic channel loading planning and optimization;
- monitoring and logging of all transmission operations;
- support for various protocols (xrootd, https, etc.)

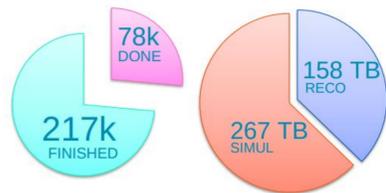


WorkFlow Management System



This is a system responsible for managing data processing processes, managing and interacting with other implemented services, such as Rucio, PanDA, and CRIC.

- Successfully processed about 300k jobs
- Total output datasets volume – more than 425 TB



Task ID	Task name	Status	Start date	End date	Walltime	Total events	Events per job	Total Jobs	Out DS size, GB	Out Log size, GB
371	PROD2025-017.RECO	finished	22 May 2025	23 May 2025	11	None	None	4996	12840.12	2.60
370	PROD2025-017.SIM	done	20 May 2025	21 May 2025	37392	20000000	4000	5000	13114.05	1.12
369	PROD2025-016.RECO	finished	10 May 2025	13 May 2025	30	None	None	9999	18557.43	5.07
368	PROD2025-016.SIM	done	08 May 2025	09 May 2025	22931	40000000	4000	10000	18365.39	1.92
367	PROD2025-015.RECO	finished	07 May 2025	09 May 2025	23	None	None	9992	18543.30	5.05
366	PROD2025-015.SIM	done	06 May 2025	07 May 2025	24486	40000000	4000	10000	18362.72	1.97
365	PROD2025-014.RECO	finished	04 May 2025	06 May 2025	24	None	None	9991	18540.20	5.05
364	PROD2025-014.SIM	done	03 May 2025	04 May 2025	24612	40000000	4000	10000	18358.86	1.96
363	PROD2025-013.RECO	finished	02 May 2025	04 May 2025	20	None	None	9992	18536.20	5.06
362	PROD2025-013.SIM.2	done	30 Apr 2025	02 May 2025	24899	40000000	4000	10000	18357.95	1.93

MESHCHERYAKOV LABORATORY OF INFORMATION TECHNOLOGIES