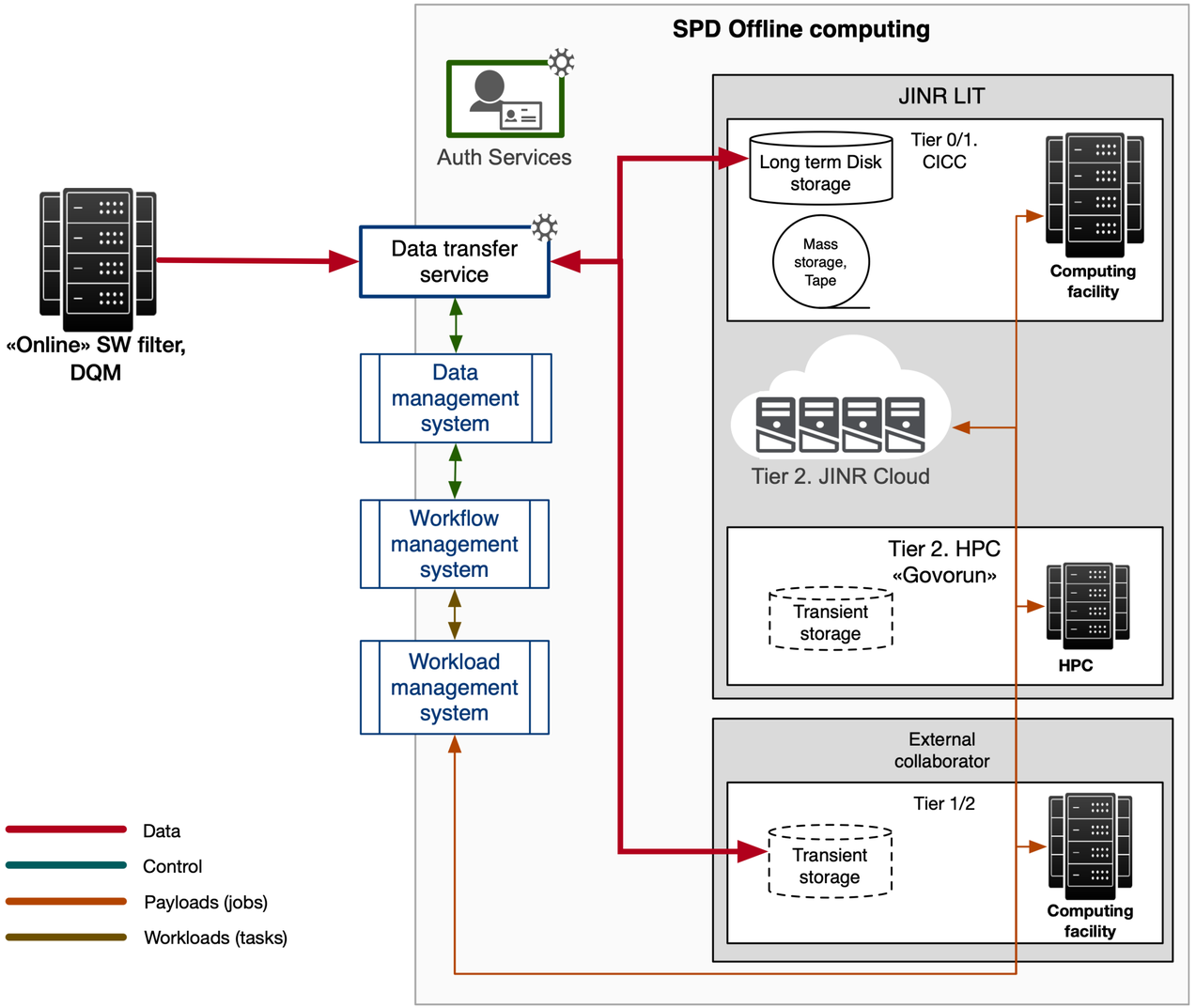
SPD offline computing: status and plans

Artem Petrosyan (MLIT JINR)
October 3, 2023

Computing system components

- CRIC information system — the main integration component of the system: gathers info about all computing and storage resources, access protocols, entry points, and many other things in one place and distributes this info via API to all other components mentioned below
- PanDA WFMS/WMS — manages data processing at the highest level of chains of tasks and datasets or periods and campaigns, finds the best computing resource for task to be executed on, manages individual jobs (usually 1 job means 1 input file) processing
- Rucio DMS — responsible for data management, including data catalog, data integrity and data lifetime management strategies
- FTS DTS — enables massive data transfers

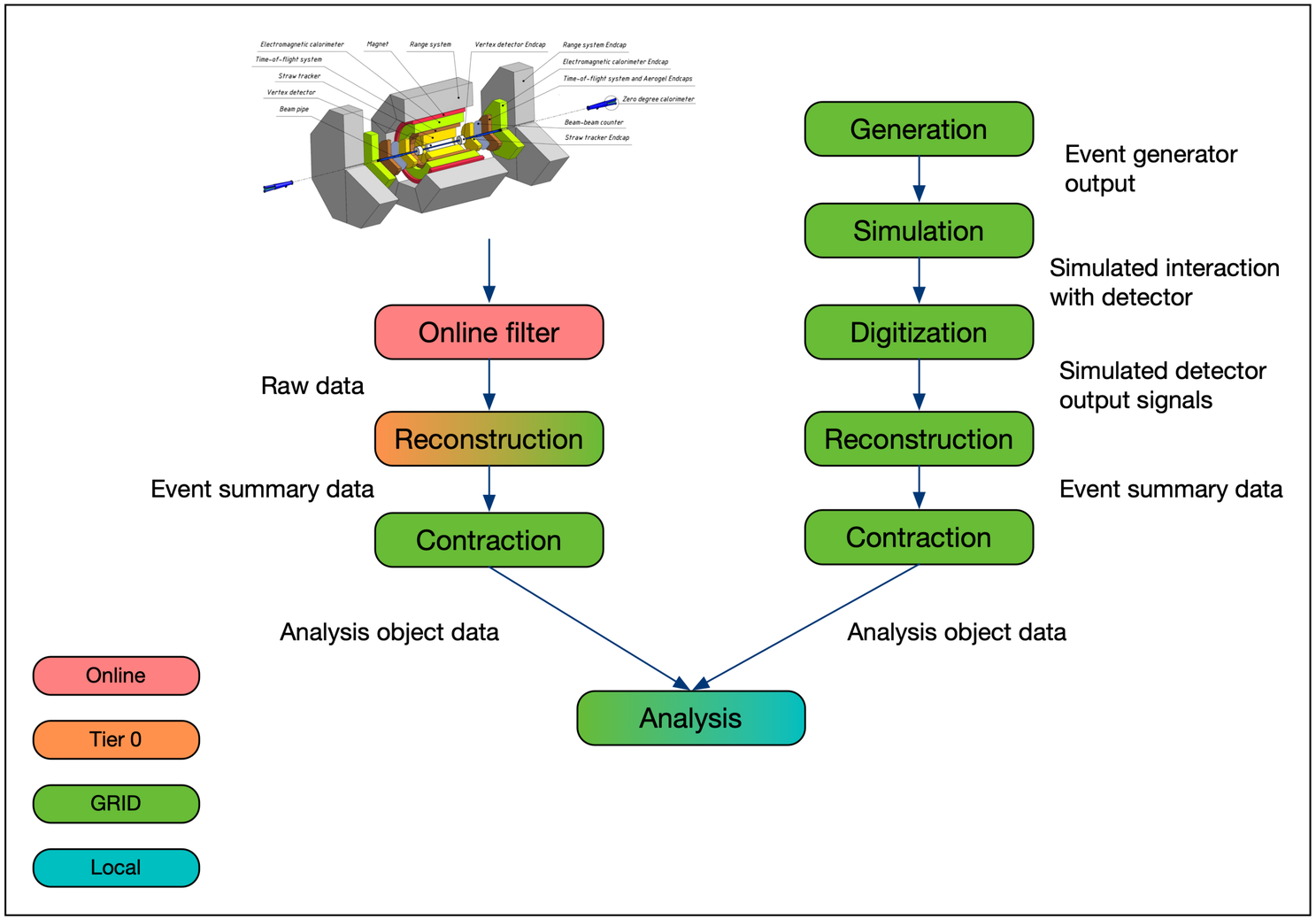


Computing system services status

- CRIC
  - CRIC was deployed at LIT in 2020, tested with BM@N and now is the information system for the SPD experiment
  - Our installation is not the same as ATLAS has, we support our own branch on the top of the CRIC core
- PanDA
  - PanDA was deployed at LIT in 2015 in order to manage COMPASS data processing, another instance of PanDA was deployed in 2020, tested with BM@N jobs and now is the workflow/workload management system for the SPD experiment
  - We are running the latest version of PanDA with SPD-driven extensions
- Rucio
  - Rucio was deployed at LIT in 2022, at the moment we're learning how to work with this system, but it is already integrated with PanDA and is being used as data catalog for jobs and tasks data
- FTS
  - Not yet installed, we are going to install and support FTS as a central service for many our experiments, not only for the SPD

Processing steps distribution over computing resource types

- Execution of events reconstruction and reprocessing jobs is accompanied by intensive I/O operations and will be done mostly on the dedicated farms on JINR site as Tier 0 component of the distributed computing system
- The use of Tier 0 is dictated by huge amount of initial data, gathered by the physics facility — data must be reduced as much as possible in order to be ready for distribution
- Less I/O intensive steps, especially Monte-Carlo production, can be performed on the remote computing centres
- User analysis can be run on every close to user resource



How we recommend to use our computing resources

- Personal computers and laptops: development, initial testing
- JINR cloud service: development, testing in the environment, profiling, testing with containers
- Batch service: personal analysis, small tasks
- Production system over the distributed computing resources: large (thousands jobs) tasks over large datasets, chain of tasks and workflows, mass productions in the interest of the collaboration

Production system

- In two words production system is a database with the web UI which is used to manage SPD data processing and store the history:
- What must be kept for each task?
  - Location of the applied software (SpdRoot) at CVMFS
  - Location of input parameters, like input files, datasets or parameters (number of events, flags and so on) to be passed to SpdRoot or other applied software
  - Location and metadata of input and output data
- Once being prepared in the ProdSys, the task being sent to the WFMS and, as soon as it’s done, the task closes and can be used as an input data by for later processing
- The ProdSys database in combination with frozen sandboxes at CVMFS allow to reproduce any processing at any time

Production setup

- CVMFS as an entry point to the “official” versions of SpdRoot:
  - /cvmfs/spd.jinr.ru/images/spdroot-4.1.5.sif
- Production setups on the CVMFS in form of frozen sandboxes
  - /cvmfs/spd.jinr.ru/production/testOpenCharm/simu_VA_OC.C
- Each new production means new directory with all dependencies on CVMFS
- Each production on CVMFS corresponds to path with the same name on EOS
  - /eos/nica/spd/production/testOpenCharm
- Directory to store results of the production on EOS with strict access rights in order not to be deleted accidentally

Distributed infrastructure test suite

- As the distributed infrastructure grows, the testing system becomes increasingly important
- What do we expect this suite to do?
  - Get a list of CE endpoints from some external source (CRIC or other)
  - Get a payload to be executed on them
  - Run this payload with a given frequency over the CE endpoints and store results in some round robin storage in some common format (JSON)
- There is an intention to make a standalone experiment agnostic project designed which then can be distributed as rpm
- There is also a request for such system from Juno experiment, we think that this activity may be organised as a common effort in the interest of several experiments using distributed computing infrastructure to process and store their data

Manpower

- CRIC 0/0.25 FTE
- WMS/WFMS/Harvester/Pilot 1/2 FTE, Artem Petrosyan
- Rucio 1/1 FTE till September 2024, Alexey Konak
- FTS 0/0.25 FTE
- Production system 0/1 FTE
- Production manager 0/0.5 FTE
- Test suite 0/0.5 FTE is being discussed with Andrey Kiryanov (PNPI) and Andrey Zarochentsev (SPbSU)

Summary

- Almost all components of the production system, which will be responsible for running massive calculations in the distributed environment have already deployed and configured
    - First test production (samples of D-meson decays and minimum bias) is now ongoing
    - Last week I finally managed to run a test task of containers with SpdRoot through the system
- We have several external participants willing to participate in the software development and data processing: PNPI, SPbSU and INP BSU, their CE are already connected to the distributed computing infrastructure of the experiment
- Our next steps will lay in the field of data and workflow management: continued Rucio integration, FTS deployment, Production system development
- Soon integration work of JINR CTA will be started, at the end of the story our most sensitive data will be stored on tapes. Once it’s done our data processing scheme will be changed: in order to be processed files will be first copied from CTA to EOS and after processing results will be migrated on tapes. EOS will be used mostly as a disc pool for data during processing. After the work is done all temporary data on EOS will be kept for a declared period of time and then removed.

Thank you!