Production and analysis with SpdRoot in PANDA (physicist's view)

Igor Denisenko

iden@jinr.ru

SPD Software & Computing Weekly Meeting

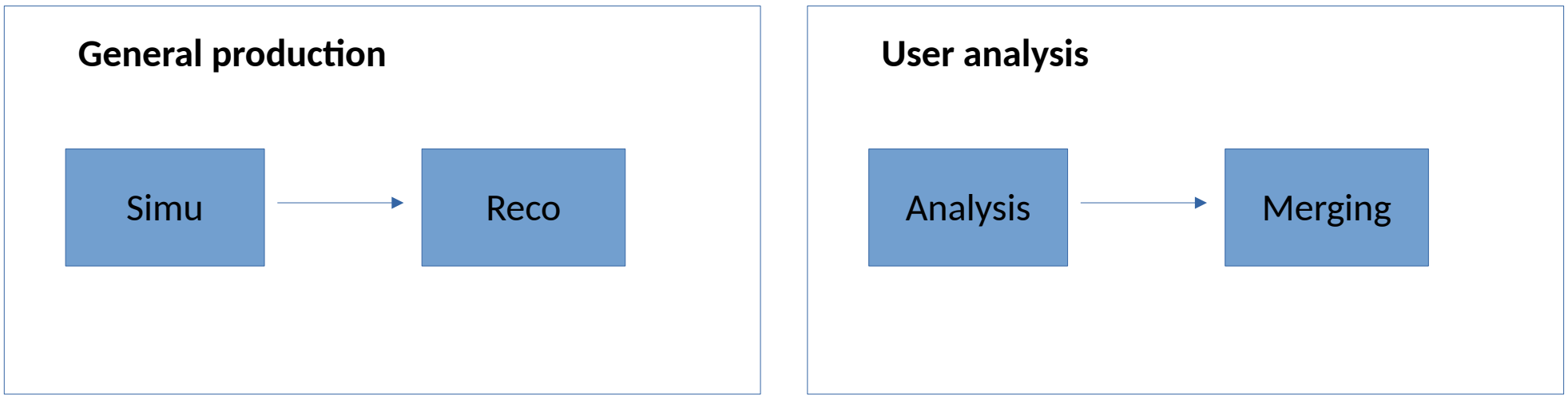
15.09.2026

Motivation of this talk

- Currently, we need to process large amounts of data with SpdRoot (currently 200M MB events ore more).
- The available Slurm queue (400 jobs at best) is not sufficient to process such amounts of data without creating “jams” in the Slurm queue.
- To achieve robust physics results for Phase II we need to proceed to statistics of 1 billion events and more!
- SpdRoot code is not optimal, but Sampo development is a priority.
- The SPD computing resources can be used as
  - slurm-like batch system,
  - PANDA job submission (this talk).
- Here a simplistic user analysis model is proposed to complement the general data production in PANDA.

Production stages (SpdRoot)

Given the software migration plans, SpdRoot can use a simplistic (compared to Sampo) production and physics analysis model.



General data production

The general data production is initiated by Phys. Coord. or his deputy and consists of two successive steps: simulation and reconstruction.

Simulation

Input:
- **software release**
- **simulation script**
- **number of events**
- random seed

Output:
- **"simu"** and **"params"** datasets

→

Reconstruction

Input:
- **simu.root** and **param.root** files

Output
- **"reco"** dataset

What can be improved:
- simulation and reconstruction tasks are now submitted manually (not in chain)
- the jobs are manually split to chunks (acceptable for 200 million, bad for 1 billion), chunks details are only relevant to PM, not for any other user
- production definition can be made minimalistic and more convenient

User analysis (proposal)

A request for user analysis can be initiated by any registered SPD member.

User analysis

Input:

- **"reco"** and "**params**" datasets
- **user script** (the same format as for the simulation and reconstruction jobs: *ana.C(nevents, "reco.root", "params.root"[, seed])*

Output:

- **"analysis"** dataset, consisting from user ROOT files with histograms and trees **only**

→

Merging (with ROOT **hadd**)

Input:

- **"analysis"** dataset
- merging parameter (e.g. 1000 → 1)

Output

- **"final" datatset,** which user can download on his working PC

Issues:

- Procedure of uploading the user script, verifying its safety and safety of merging procedure are not defined.
- No interface to specify the user analysis job.
- Before job submission to PANDA, the script should be tested on the meaningfully large dataset at lxui (test queue is critical).
- Who is responsible for managing user analyses? PM?