SPD S&C warmup

SPD S&C Subprojects

Software - frameworks for applied software and data processing algorithms (offline and online)

Distributed computing

   Middleware - Production system, PanDA, Rucio, FTS, CRIC etc.

   Infrastructure - computing sites, storages etc.

Online Filter

   Middleware - Visor [Workflow management, Data management, Workload management]

   Infrastructure - dedicated computing facility, right after DAQ

Information system and databases

Software

Two frameworks:

- Sampo - steady progress with development, expecting step by step move to production. Lack of manpower - at least one more core developer is needed, contribution from generators, GEANT4, reconstruction appreciated.
    - will be great to have a kind of delivery plan
    - requirements for online processing should be collected
- SPDRoot - instead of glitches and slow bug fixing cycle, steel main machinery for data production
    - should move to face out

Distributed computing

Distributed data processing step by step goes to operation mode.

Increases requirements for different kinds of monitoring tools

Rucio - the only tool to interact with production data

Production data naming convention was reviewed and changed to have more human oriented format

Current data production terms

Production - amount of work produced by one production chain (set of related tasks).

Task - a part of production, split by type of processing and produced one dataset

Dataset - an amount of data limited by size in terms of number of files

File - a unit of work which process in one job

Datasets naming convention for 2025

Field	Description
MC2025_S1 MC2025_S2	Scope
[desc]	Short name of physics aim, software type (version)
[energy][polarization]	
[ProductionID]	ID of Production (from production system)
[data type]	GEN - Generators SIM - Simulation RECO - Reconstruction AOD - Data for analysis
[Version]	Version is used in case something goes wrong with production, and it fixed without changing of production ID.

GEN datatype does not exist for SPDRoot (should be existed for Sampo)
AOD - results of processing of whole dataset aimed to extract only required information

Full dataset name example:
    MC2025_S1:minbias-spdroot4161.27GeV.PROD01.SIM.1
    MC2025_S1:minbias-sampo0001.27GeV.PROD12.SIM.2

Files naming convention for 2025

Field	Description
Data type	GEN - Generators (EVNT) SIM - Simulation (HITS) RECO - Reconstruction (AOD) DER - Derivation of data for analysis (DAOD)
[software type]	
[TaskID]	PanDA TaskID
_[Job number]	Serial job number in task (not equal to PanDA JobID)
[file format]	
[Version]	Version is used in case something goes wrong with job, and it fixed without changing of job serial number.

File name example:
   MC2025_S1:SIM.spdroot_4161.465.000027.root.1
   MC2025_S1:SIM.sampo_0001.1380.000003.root.1