SPD Computing for analysis

Danila Oleynik, 16.04.2025

JINR computing resources

Suggested resources to perform your analysis jobs in JINR is CICC (also known as 'lxui.jinr.ru') compute cluster managed by SLURM coupled with EOS based data storage system. CVMFS is used for placing and distribution of applied software (SPDRoot containers).

AFS file system is used for ‘home' directories, but take into account that available space in home directory quite low and may disappoint you.

```
fs listquota <path to home directory>
```

Take into account, that CICC is shared among a lot of users, so your jobs may wait in queue sometimes significant time.

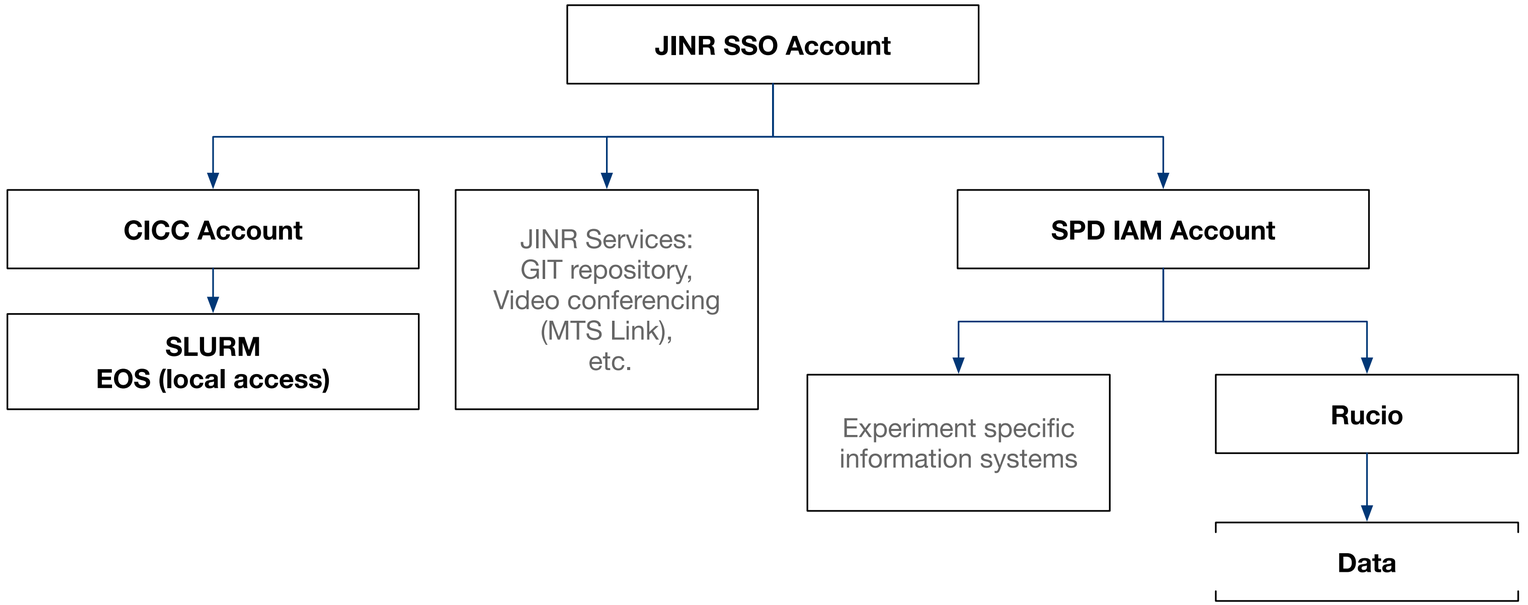
Production data is placed on EOS storage, and have read access for all SPD users. There is no way to navigate in this data through browsing of directories 😜

How to get access

- To get any access to JINR hosted resources and services, you should have JINR SSO account.
- There are different procedures to get for:
    - JINR staff members (https://noc.jinr.ru/ru/registry/sso-reg.php) just come to LIT (blb. 134) room 200 and ask for account
    - Non JINR members: catch your SPDs Team leader and ask to be included in JINR Associated staff. Be patient, paper process will take some time… and finally you will have JINR SSO Account
    - Students should be properly registered in JINR UC
- With JINR SSO account you may easily get CICC account (with same credentials)
- Also, you will be able to apply for access to GIT repository, got access to JINR CA (https://ca.jinr.ru) etc.
- To get access to SPD Specific resources and data you should be also registered in SPD VO through SPD IAM authorization service: http://spd-iam.jinr.ru (*JINR SSO will be required for this too*).

How to get data

- The only way to get production data is using Rucio Client.
- Rucio uses SPD IAM as authorisation backend… so



Data production (Simulation)

Centralized MC data production through SPD Production system in progress

- Current data samples: ~5M 40M events (with 4000 events per file), split by productions. If you need bigger data sample, please contact us.
- Two-step chain:
    - Generation + simulation
    - Reconstruction
- Each step produces dataset of files accordingly.
- If you require producing of particular data, you should agree it with Physics coordinator and contact with Production manager.

Amount of data for analysis

- SPDRoot is stable enough with producing of 4000 events per job
- 2-step production chain of 40M Events produces: ~20000 of files in two datasets (SIM, RECO)
  - Size of dataset may reach a few TB - does not fit well to be analysed on regular workstation
    - will be nice to add derivation step into processing step, to reduce data for final analysis
  - https://docs.google.com/spreadsheets/d/1JWob53dfwMvTlmdsGncwQmPeVbmzFKuD8DHE4rYsJFw/edit?usp=sharing list of Productions
- Good news: data is somehow organised now, and production procedures settled

In plans

- **ProductionDB** - information system which will replace Google sheets
- **Sampo framework** - will replace SPDRoot (please, check our regular reports about progress)
- Production system upgrade expected this year
- Local facilities for data analyses to be discussed

Once upon a time there were mice. Everyone offended them. One day the mice came to the Owl:

– Wise Owl, help! Everyone eats us. Soon we will be gone. What to do?

The owl thought and said:

– Mice! Become hedgehogs! You will be prickly and inaccessible to hunters.

The mice ran joyfully:

– Let's become hedgehogs! Let's become hedgehogs!

Suddenly one stopped:

– ***Does anyone know: how to become hedgehogs?***

Nobody. They ran back to the owl.

– The Owl! And how do we become hedgehogs???

– Mice! Get off! I'm not a tactician, I'm a strategist!