SPD Production System Status

Artem Petrosyan, MLIT, JINR
VIII SPD Collaboration Meeting
November 7, 2024

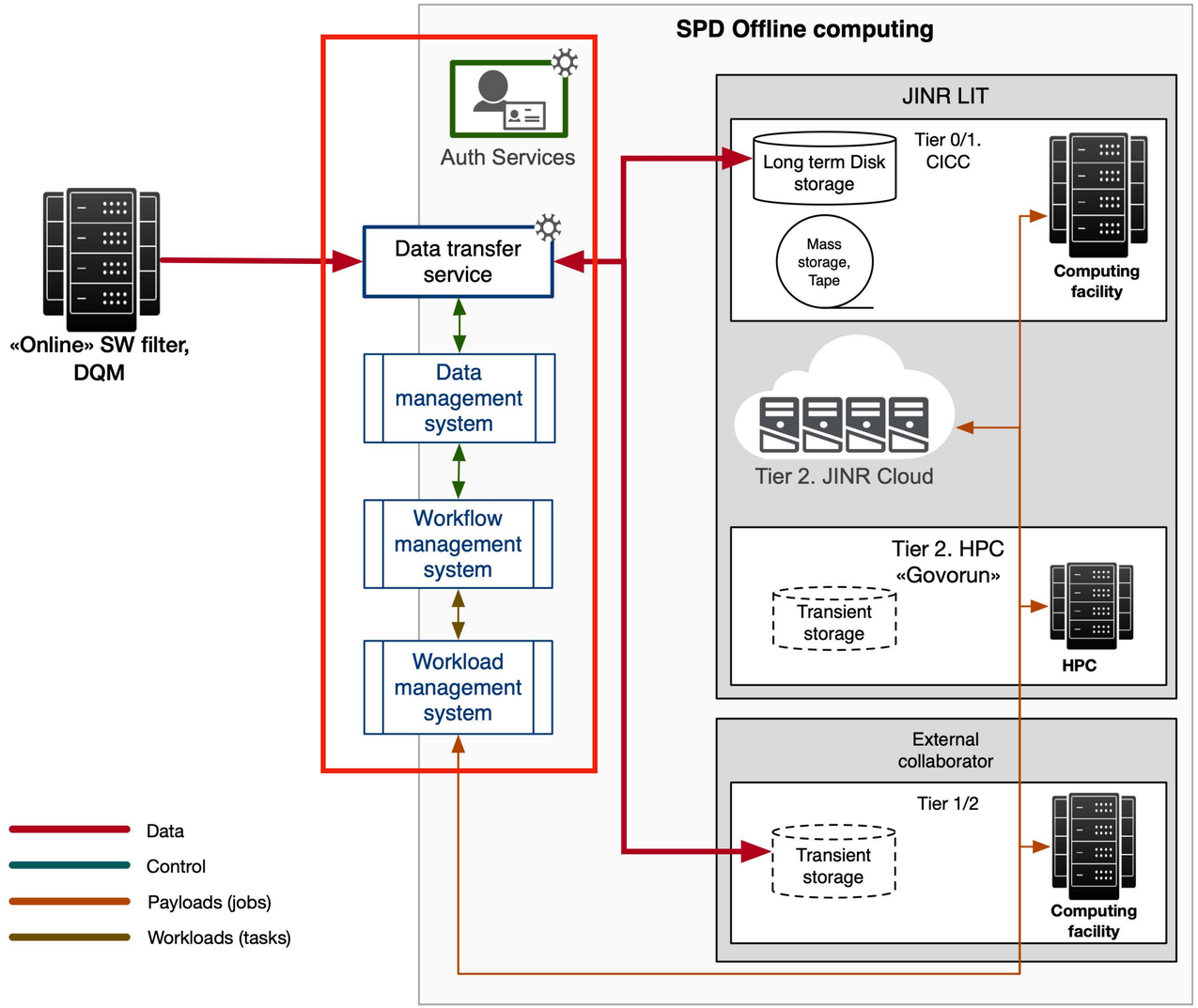
Progress since the previous collaboration meeting

- IAM identity and access management service now in production
- Several full scale productions performed during last summer and fall
- Rucio data management system in production and integrated with PanDA
- Production system operated in JINR and PNPI
- Control panel of the production manager under development

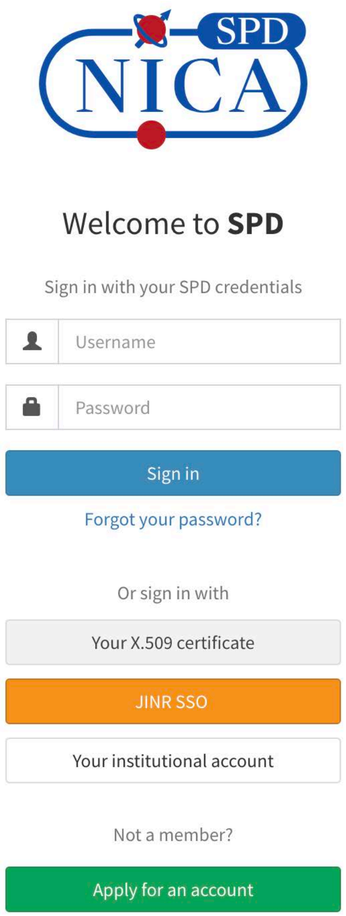
Components of the Production System

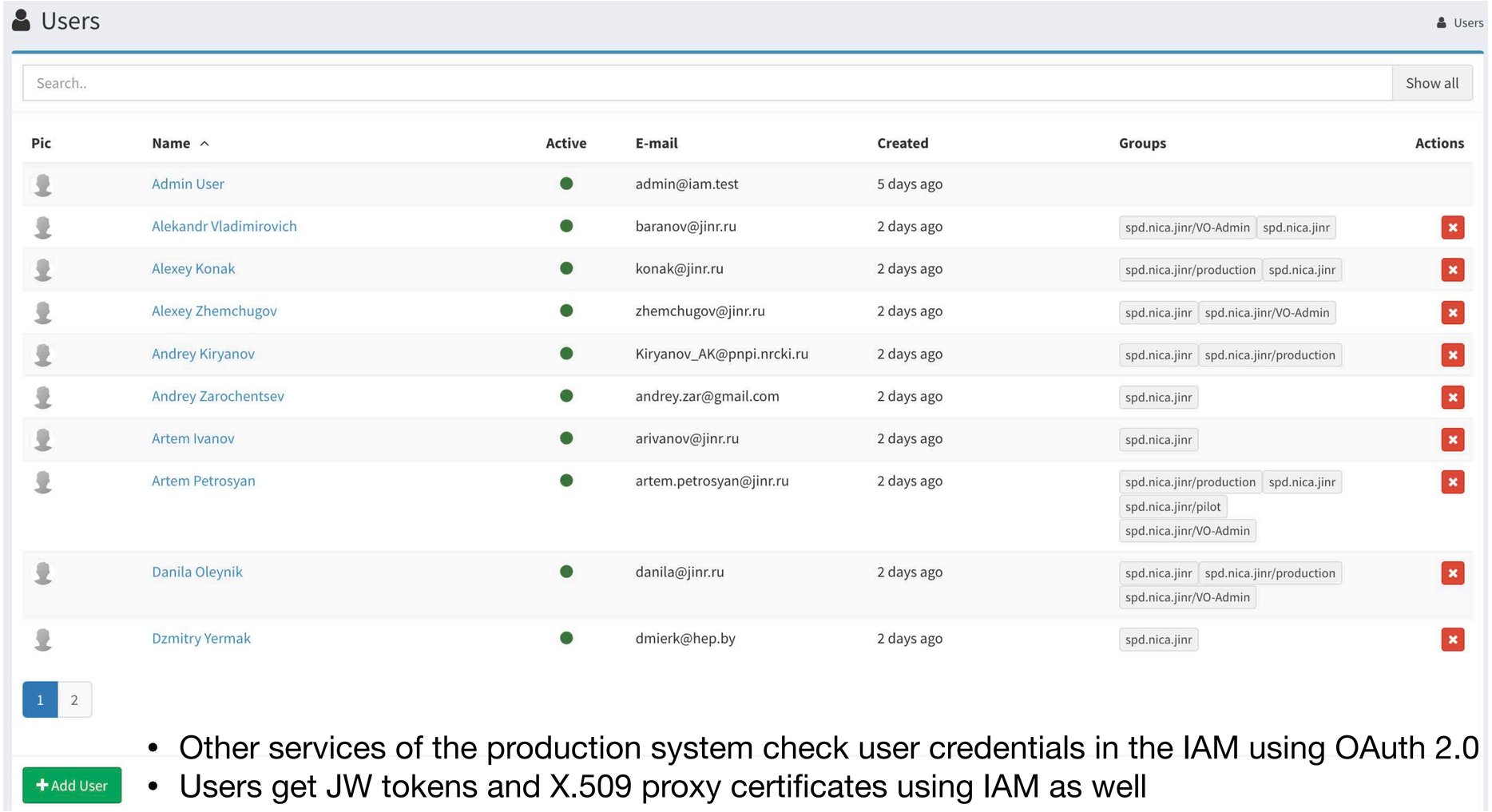
- IAM — an entry point to all members of the computing services of the collaboration: stores user profiles, their roles and rights to perform certain actions
- CRIC information system — the main integration component of the computing system: contains info about all computing and storage resources, access 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

- Control panel as top level orchestrator and user interface



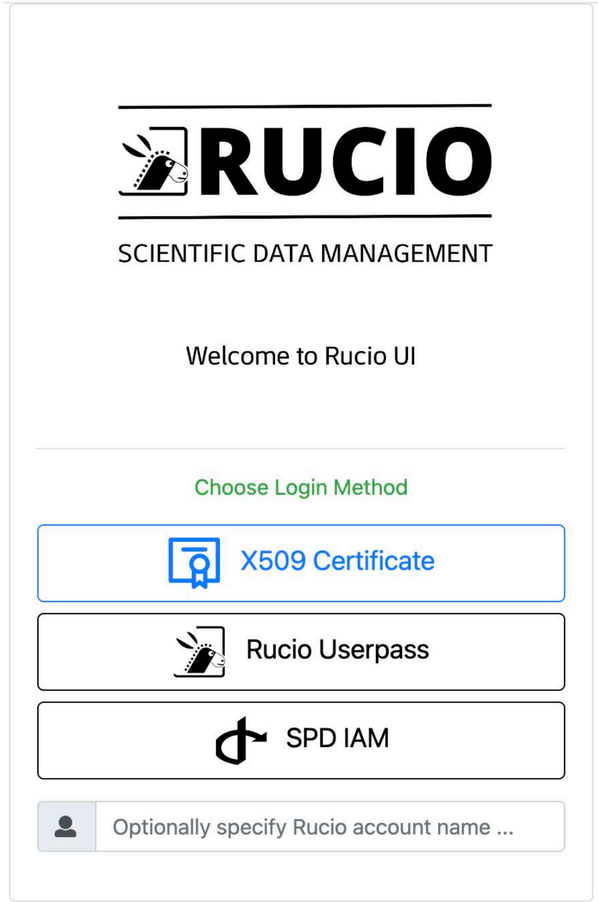
Identity and Access Management Service

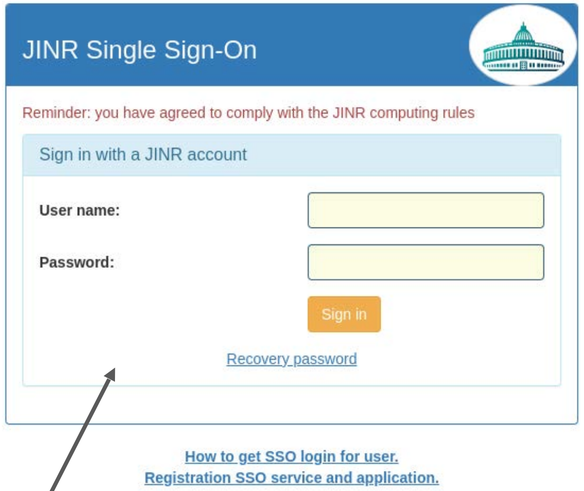


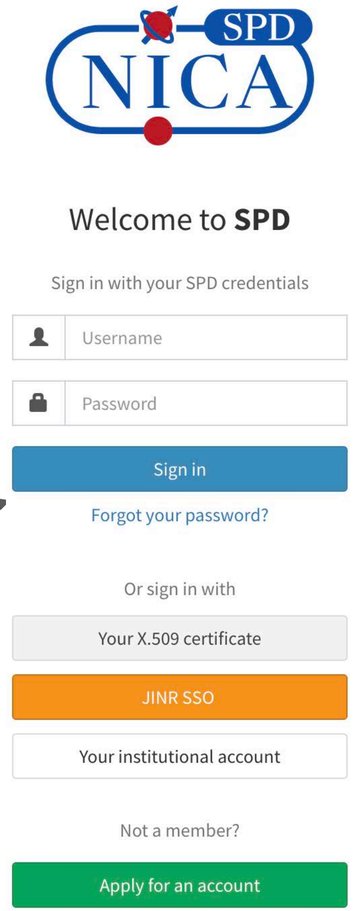


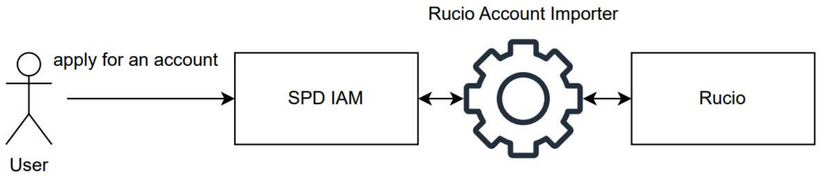
- Other services of the production system check user credentials in the IAM using OAuth 2.0
- Users get JW tokens and X.509 proxy certificates using IAM as well

DDM and IAM Integration









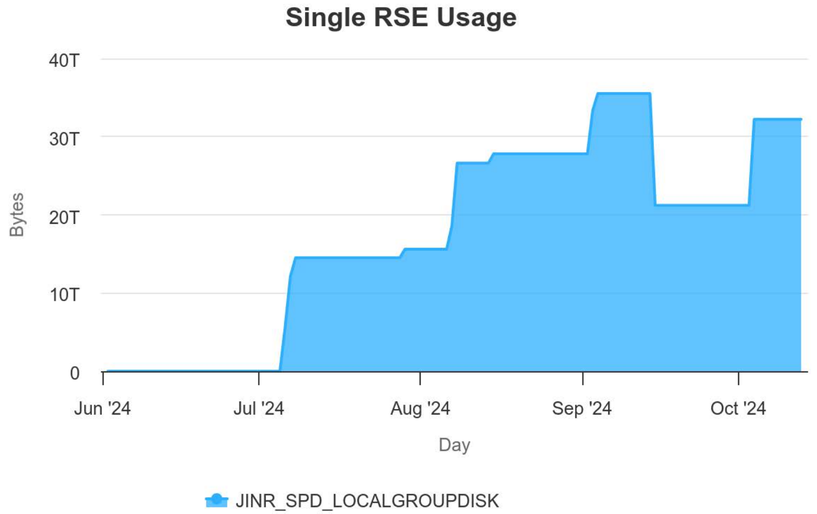
The Rucio Account Importer was implemented to import accounts and their user identification information from SPD IAM to Rucio

Recent developments in PanDA WMS/WFMS

- Brokerage: clouds definition, computing site automatic selection
- Simplification in tasks definition
- Monte-Carlo processing chain tested in several full scale productions during the summer and fall
- Pilot: errors handling, processed events from logs, etc.

Data Production

- We ran several scale productions using resources and storages at JINR and PNPI, generated ~60 millions of events and produced ~30TB of data
- Two patches were applied to SPDRoot to solve issues exposed during mass processing exercises
- We expect to store two copies of the important data – origin at JINR and replicas somewhere else, at the moment at PNPI



Name	Account	RSE Expression	Creation Date	Remaining Lifetime	State	Locks OK	Locks Replicating	Locks Stuck
2024:2024.MC.27GeV.test-minbias.00001.SIMUL.6f25043e-689f-40f7-951f-ba6e0c9f4d14.0.log	panda	JINR_SPD_LOCALGROUPDISK	2024-08-07T07:55:11.000Z	-	OK	5006	0	0
2024:2024.MC.27GeV.test-minbias.00001.SIMUL.6f25043e-689f-40f7-951f-ba6e0c9f4d14.0.P	panda	JINR_SPD_LOCALGROUPDISK	2024-08-07T07:55:12.000Z	-	OK	5000	0	0
2024:2024.MC.27GeV.test-minbias.00001.SIMUL.6f25043e-689f-40f7-951f-ba6e0c9f4d14.0.S	panda	JINR_SPD_LOCALGROUPDISK	2024-08-07T07:55:12.000Z	-	OK	5000	0	0
2024:2024.MC.27GeV.test-minbias.00001.RECO.6f25043e-689f-40f7-951f-ba6e0c9f4d14.1.R	panda	JINR_SPD_LOCALGROUPDISK	2024-08-15T09:07:38.000Z	-	OK	1426	0	0
2024:2024.MC.27GeV.test-minbias.00001.RECO.6f25043e-689f-40f7-951f-ba6e0c9f4d14.0	panda	JINR_SPD_LOCALGROUPDISK	2024-09-26T11:43:57.000Z	-	OK	23	0	0
2024:2024.MC.27GeV.test-minbias.00001.RECO.6f25043e-689f-40f7-951f-ba6e0c9f4d14.1.log	panda	JINR_SPD_LOCALGROUPDISK	2024-09-26T11:44:52.000Z	-	OK	1451	0	0
2024:2024.MC.27GeV.test-minbias.00001.RECO.2.log	panda	JINR_SPD_LOCALGROUPDISK	2024-10-02T12:36:46.000Z	-	OK	4785	0	0
2024:2024.MC.27GeV.test-minbias.00001.RECO.2.R	panda	JINR_SPD_LOCALGROUPDISK	2024-10-02T12:36:46.000Z	-	OK	4753	0	0

More details about DDM service in the talk of Alexey Konak later today

Example of the MC task definition 1/2

- Step 1: Simulation
- User defines an output dataset name
- Desired total number of events and events per job
- The system divides the total number of events by the number events per job and generates the required number of jobs
- User can specify either a specific computing queue or a cloud; in the second case, the jobs will be distributed among the queues of the specified cloud
- Jobs execution is performed in the container

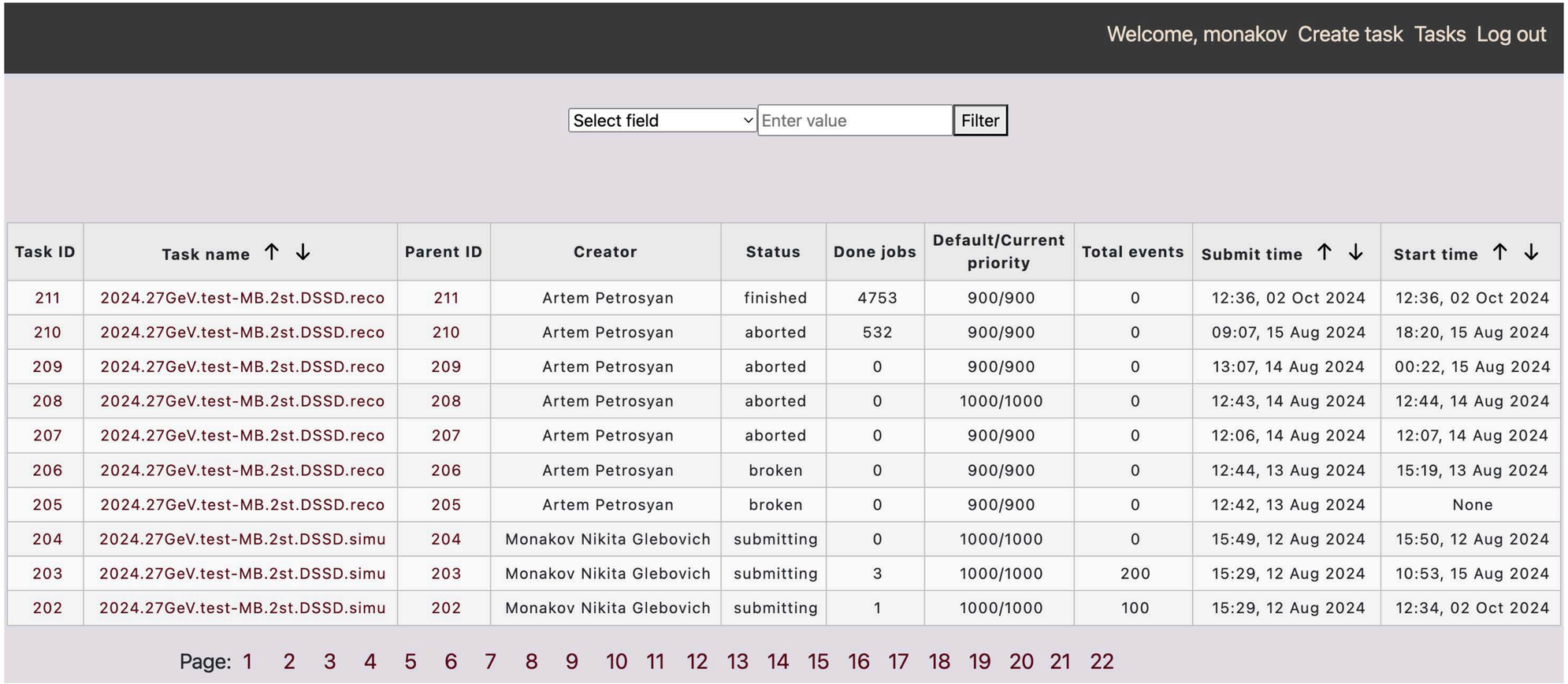
```
TaskName = '2024.27GeV.test-MB.2st.DSSD.simu'
DatasetName = '2024.MC.27GeV.test-minbias.00001.SIMUL.0'
#DatasetName = 'jeditest.000023.simu'

taskParamMap = {}

taskParamMap['nEventsPerJob'] = 4000
taskParamMap['nEvents'] = 20000000
taskParamMap['noInput'] = True
taskParamMap['skipScout'] = True
taskParamMap['taskName'] = TaskName
taskParamMap['userName'] = 'Artem Petrosyan'
taskParamMap['vo'] = 'spd.nica.jinr'
taskParamMap['taskPriority'] = 900
taskParamMap['architecture'] = 'x86_64'
taskParamMap['transUses'] = 'A'
taskParamMap['transHome'] = None
taskParamMap['transPath'] = 'https://159.93.221.125:8080/spd_simu_VA_transform.sh'
taskParamMap['processingType'] = 'step1'
taskParamMap['prodSourceLabel'] = 'managed'
taskParamMap['taskType'] = 'test'
taskParamMap['workingGroup'] = 'spd.nica.jinr'
taskParamMap['cloud'] = 'JINR'
taskParamMap['ramCount'] = 1900

outDatasetNameLog = '{0}.log'.format(DatasetName)
outDatasetNameS = '{0}.S'.format(DatasetName)
outDatasetNameP = '{0}.P'.format(DatasetName)

taskParamMap['log'] = {'dataset': outDatasetNameLog,
                       'type':'template',
                       'param_type':'log',
                       'token':'DATADISK',
                       'value':'{0}.${{SN}}.log.tgz'.format(DatasetName)}

taskParamMap['jobParameters'] = [
    {'type': 'constant',
     'value': '''singularity run --bind /cvmfs/spd.jinr.ru/production/MC/2024.27GeV.test-MB.2st.DSSD:/prod -H
     ./:/WORKDIR
    /cvmfs/spd.jinr.ru/images/spdroot-4.1.6.sif spdroot.py -b -q \'/prod/simu.C({0}, '''.format(taskParamMap['nEventsPerJob']),
     },
```

Example of the MC task definition 2/2

- Step 2: Reconstruction
- User defines a name of the input dataset, in this example there are two input datasets of the same size (have the same number of files)
- Sets a name of the output dataset
- Set how many jobs needs to be created per each file in the dataset
- At the job generation stage, the workload management system communicates with the data management service, reads the size (number of files) of the dataset and generates the appropriate number of jobs
- The input files will be staged-in from the storage closest to the computing node

```
scope = '2024'
inDatasetName = '2024.MC.27GeV.test-minbias.00001.SIMUL.0'
outDatasetName = '2024.MC.27GeV.test-minbias.00001.RECO.2'

inDatasetNameS = '{0}.S'.format(inDatasetName)
inDatasetNameP = '{0}.P'.format(inDatasetName)
outDatasetNameR = '{0}.R'.format(outDatasetName)
outDatasetNameLog = '{0}.log'.format(outDatasetName)

taskParamMap = {}

taskParamMap['nFilesPerJob'] = 1
taskParamMap['nEventsPerJob'] = 4000
taskParamMap['noInput'] = False
taskParamMap['taskName'] = TaskName
taskParamMap['userName'] = 'Artem Petrosyan'
taskParamMap['vo'] = 'spd.nica.jinr'
taskParamMap['taskPriority'] = 900
taskParamMap['architecture'] = 'x86_64'
taskParamMap['transUses'] = 'A'
taskParamMap['transHome'] = None
taskParamMap['transPath'] = 'https://159.93.221.125:8080/spd_simu_VA_transform.sh'
taskParamMap['processingType'] = 'step2'
taskParamMap['prodSourceLabel'] = 'managed'
taskParamMap['taskType'] = 'test'
taskParamMap['workingGroup'] = 'spd.nica.jinr'
taskParamMap['cloud'] = 'JINR'
taskParamMap['ramCount'] = 1900

taskParamMap['log'] = {'dataset': outDatasetNameLog,
                       'type': 'template',
                       'param_type': 'log',
                       'token': 'DATADISK',
                       'value': '{0}.${{SN}}.log.tgz'.format(outDatasetName)}

taskParamMap['jobParameters'] = [
    {'type': 'constant',
     'value': '''singularity run --bind /cvmfs/spd.jinr.ru/production/MC/2024.27GeV.test-MB.2st.DSSD:/prod -H
     ./:/WORKDIR /cvmfs/spd.jinr.ru/images/spdroot-4.1.6.1.sif spdroot.py -b -q \'/prod/reco.C({0}, '''
     .format(taskParamMap['nEventsPerJob'])
     },
```

Production Manager Control Panel 1/3



Welcome, monakov Create task Tasks Log out

Select field | Enter value | Filter

Task ID	Task name ↑ ↓	Parent ID	Creator	Status	Done jobs	Default/Current priority	Total events	Submit time ↑ ↓	Start time ↑ ↓
211	2024.27GeV.test-MB.2st.DSSD.reco	211	Artem Petrosyan	finished	4753	900/900	0	12:36, 02 Oct 2024	12:36, 02 Oct 2024
210	2024.27GeV.test-MB.2st.DSSD.reco	210	Artem Petrosyan	aborted	532	900/900	0	09:07, 15 Aug 2024	18:20, 15 Aug 2024
209	2024.27GeV.test-MB.2st.DSSD.reco	209	Artem Petrosyan	aborted	0	900/900	0	13:07, 14 Aug 2024	00:22, 15 Aug 2024
208	2024.27GeV.test-MB.2st.DSSD.reco	208	Artem Petrosyan	aborted	0	1000/1000	0	12:43, 14 Aug 2024	12:44, 14 Aug 2024
207	2024.27GeV.test-MB.2st.DSSD.reco	207	Artem Petrosyan	aborted	0	900/900	0	12:06, 14 Aug 2024	12:07, 14 Aug 2024
206	2024.27GeV.test-MB.2st.DSSD.reco	206	Artem Petrosyan	broken	0	900/900	0	12:44, 13 Aug 2024	15:19, 13 Aug 2024
205	2024.27GeV.test-MB.2st.DSSD.reco	205	Artem Petrosyan	broken	0	900/900	0	12:42, 13 Aug 2024	None
204	2024.27GeV.test-MB.2st.DSSD.simu	204	Monakov Nikita Glebovich	submitting	0	1000/1000	0	15:49, 12 Aug 2024	15:50, 12 Aug 2024
203	2024.27GeV.test-MB.2st.DSSD.simu	203	Monakov Nikita Glebovich	submitting	3	1000/1000	200	15:29, 12 Aug 2024	10:53, 15 Aug 2024
202	2024.27GeV.test-MB.2st.DSSD.simu	202	Monakov Nikita Glebovich	submitting	1	1000/1000	100	15:29, 12 Aug 2024	12:34, 02 Oct 2024

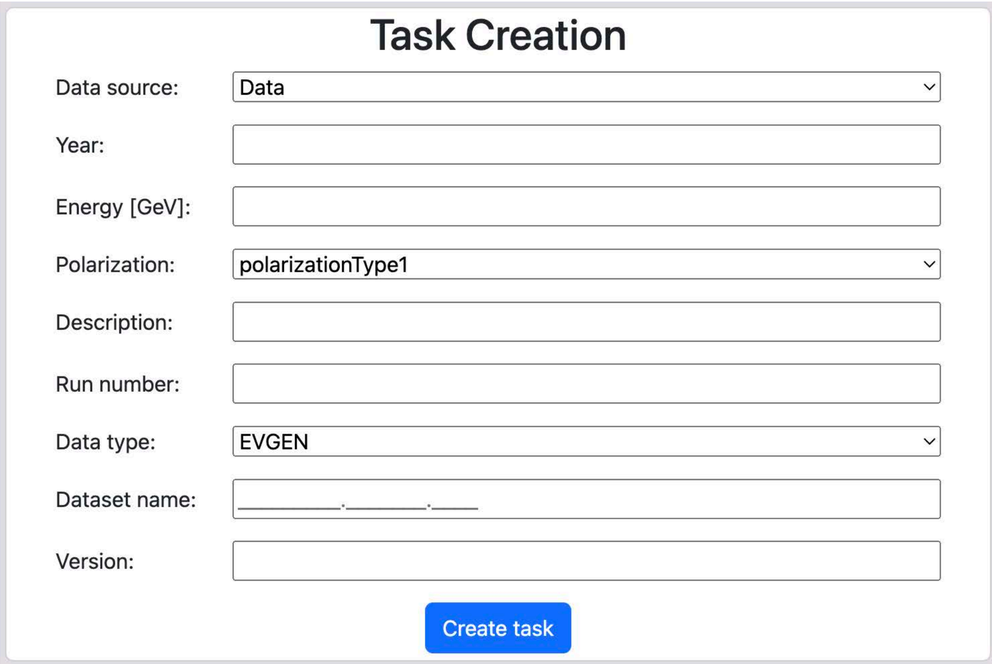
Page: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

- Registered as a client in the IAM, can be accessed only by users who has Production role in the IAM
- A control panel for pipelines submissions under development

Production Manager Control Panel 2/3

Task ID	211
Parent task ID	211
Chain task ID	None
Name	2024.27GeV.test-MB.2st.DSSD.reco
Type	managed
VO	spd.nica.jinr
Creator	Artem Petrosyan
State status	finished
Total events	0
Total jobs done	4753
Total requested jobs	8365
Priority	900
Current priority	900
Submit time	12:36 October 2, 2024
Start Time	12:36 October 2, 2024
Time stamp	2:36 October 9, 2024
JEDI parameters	{"nFilesPerJob": 1, "nEventsPerJob": 4000, "noInput": false, "taskName": "2024.27GeV.test-MB.2st.DSSD.reco", "userName": "Artem Petrosyan", "vo": "spd.nica.jinr", "taskPriority": 900, "architecture": "x86_64", "transUses": "A", "transHome": null, "transPath": "http://159.93.221.125:8080/spd_simu_VA_transform.sh", "processingType": "step1", "prodSourceLabel": "managed", "taskType": "test", "workingGroup": "spd.nica.jinr", "cloud": "JINR", "ramCount": 1900, "log": {"dataset": "2024.MC.27GeV.test-minbias.00001.RECO.2.log", "type": "template", "param_type": "log", "token": "DATADISK", "value": "2024.MC.27GeV.test-minbias.00001.RECO.2.${SN}.log.tgz"}, "jobParameters": [{"type": "constant", "value": "singularity run --bind /cvmfs/spd.jinr.ru/production/MC/2024.27GeV.test-MB.2st.DSSD:/prod -H ./:/WORKDIR /cvmfs/spd.jinr.ru/images/spdroot-4.1.6.1.sif spdroot.py -b -q '/prod/reco.C(4000, "}, {"type": "constant", "value": "\\\""}, {"type": "template", "param_type": "input", "value": "${INS}", "dataset": "2024:2024.MC.27GeV.test-minbias.00001.SIMUL.6f25043e-689f-40f7-951f-ba6e0c9f4d14.0.S"}, {"type": "constant", "value": "\\\","}, {"type": "constant", "value": "\\\""}, {"type": "template", "param_type": "input", "value": "${INP}", "dataset": "2024:2024.MC.27GeV.test-minbias.00001.SIMUL.6f25043e-689f-40f7-951f-ba6e0c9f4d14.0.P"}, {"type": "constant", "value": "\\\","}, {"type": "constant", "value": "\\\""}, {"type": "template", "param_type": "output", "token": "DATADISK", "value": "r.2024.MC.27GeV.test-minbias.00001.RECO.2.${SN/P}.root", "dataset": "2024.MC.27GeV.test-minbias.00001.RECO.2.R"}, {"type": "constant", "value": "\\\","}, {"type": "template", "value": "${RNDMSEED}", "param_type": "number"}, {"type": "constant", "value": ")'"}]}

Production Manager Control Panel 3/3



Grouping tier	Field	Description	Example
0	[YEAR]	Main Scope - the year of data production	2050
1	[MC | DATA]	Real data or simulated data	DATA
2	[energy][polarization]		250LT
3	[desc]	Short name of physics aim	minbias
4	[RunNumber]	Run number for DATA, ID for MC	27189
5	[data type]	EVGEN, SIMUL, RECO….	RAW
6	[DatasetUID]	unique ID of the dataset	636763fd78df7d
7	[Version]	for reprocessing	0

- In order to ease metadata catalog navigation, data filtration, identification, etc., a datasets naming convention was proposed in November 2023
- Dataset name example: 2025.MC.250LT.minbias.27189.RAW.636763fd78df7d.0
- Control panel checks input parameters and ensures that dataset is created in accordance with the naming convention

Datasets Naming Convention Change

- In order to make datasets names look cleaner and clearer, we decided to remove UID from the naming convention
- Rucio controls uniqueness of the dataset name in the scope and there is no need to add generated UID to the name
- Rucio can also control SPD naming convention internally, this will be implemented once we have some stable version of the convention

Grouping tier	Field	Description	Example
0	[YEAR]	Main Scope - the year of data production	2050
1	[MC | DATA]	Real data or simulated data	DATA
2	[energy][polarization]		250LT
3	[desc]	Short name of physics aim	minbias
4	[RunNumber]	Run number for DATA, ID for MC	27189
5	[data type]	EVGEN, SIMUL, RECO….	RAW
~~6~~	~~[DatasetUID]~~	~~unique ID of the dataset~~	~~636763fd78df7d~~
7	[Version]	for reprocessing	0

DID
test:2024.MC.27GeV.test-minbias.00001.SIMUL.02ce69e3-cc90-46a6-a85f-aab3c15150a3.0.log
test:2024.MC.27GeV.test-minbias.00001.SIMUL.02ce69e3-cc90-46a6-a85f-aab3c15150a3.0.P
test:2024.MC.27GeV.test-minbias.00001.SIMUL.02ce69e3-cc90-46a6-a85f-aab3c15150a3.0.S
test:2024.MC.27GeV.test-minbias.00001.SIMUL.1.P
test:2024.MC.27GeV.test-minbias.00001.SIMUL.1.S
test:test.2024.MC.27GeV.test-minbias.00001.SIMUL.2.log
test:test.2024.MC.27GeV.test-minbias.00001.SIMUL.2.P
test:test.2024.MC.27GeV.test-minbias.00001.SIMUL.2.S

Remote participants

- Petersburg Nuclear Physics Institute, Gatchina
  - Contact person: Andrey Kiryanov
  - Activity: participation in all S&C activities, including R&Ds, T1, provides resources for mass data production
- Samara State University
  - Contact person: Alexander Baskakov
  - Activity: SSU is in the process of purchasing machines for farm, which will be used to run SPD jobs as T2 center
- Instituter of Nuclear Physics Belarus State University, Minsk
  - Contact person: Dmitry Ermak
  - Activity: FTS transfers, EOS tests, infrastructure use from the view of fully remote person

Manpower

- IAM — Alexander Baranov, MLIT JINR
- CRIC — Alexey Anisenkov, Novosibirsk State University
- Rucio — Alexey Konak, MLIT JINR
- PanDA — Artem Petrosyan, MLIT JINR
- PanDA Pilot — Yuri Basharimov, Gomel State University
- FTS — Artem Petrosyan, MLIT JINR
- ProdSys Control Panel — Nikita Monakov, Moscow Engineering Physics Institute/MLIT JINR
- Monitoring — Fjodor Shuvalov, Saint Petersburg State University

Summary

- The system is in place, as well as manpower for the key subsystems
- Our nearest development plan:
  - Production manager control panel enhancement, we want to allow production managers to define processings in one shot: not at the level of tasks, but at the level of full pipelines
  - Development of the tasks and jobs monitoring
- We are going to run several large processings before the end of the year, the goal is to run 1 billion events production and gather statistics about behavior of the system
- We are looking forward to start working with SAMPO-based payloads, because bug fix cycle of SPDRoot usually takes up to month and does not allow to organize rhythmic production rounds
- Services support and development:
  - Upgrade last remaining services (CRIC, PanDA) to AlmaLinux 9 and change their backend from MySQL to PostgreSQL
  - Continue transition from X.509 to JWT

Thank you for attention!