

Система управления данными Rucio для эксперимента SPD

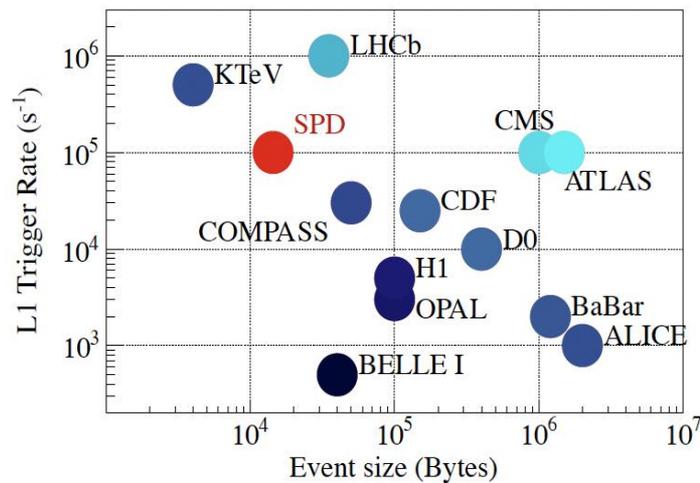
Конак Алексей

Эксперимент SPD как источник данных

SPD Spin Physics Detector - экспериментальная установка для исследования спиновой структуры ядра и других спиновых феноменов.

Особенностью эксперимента является относительно большой объем экспериментальных данных.

Большим объемам данных необходима определенная организация и возможность управления данными.



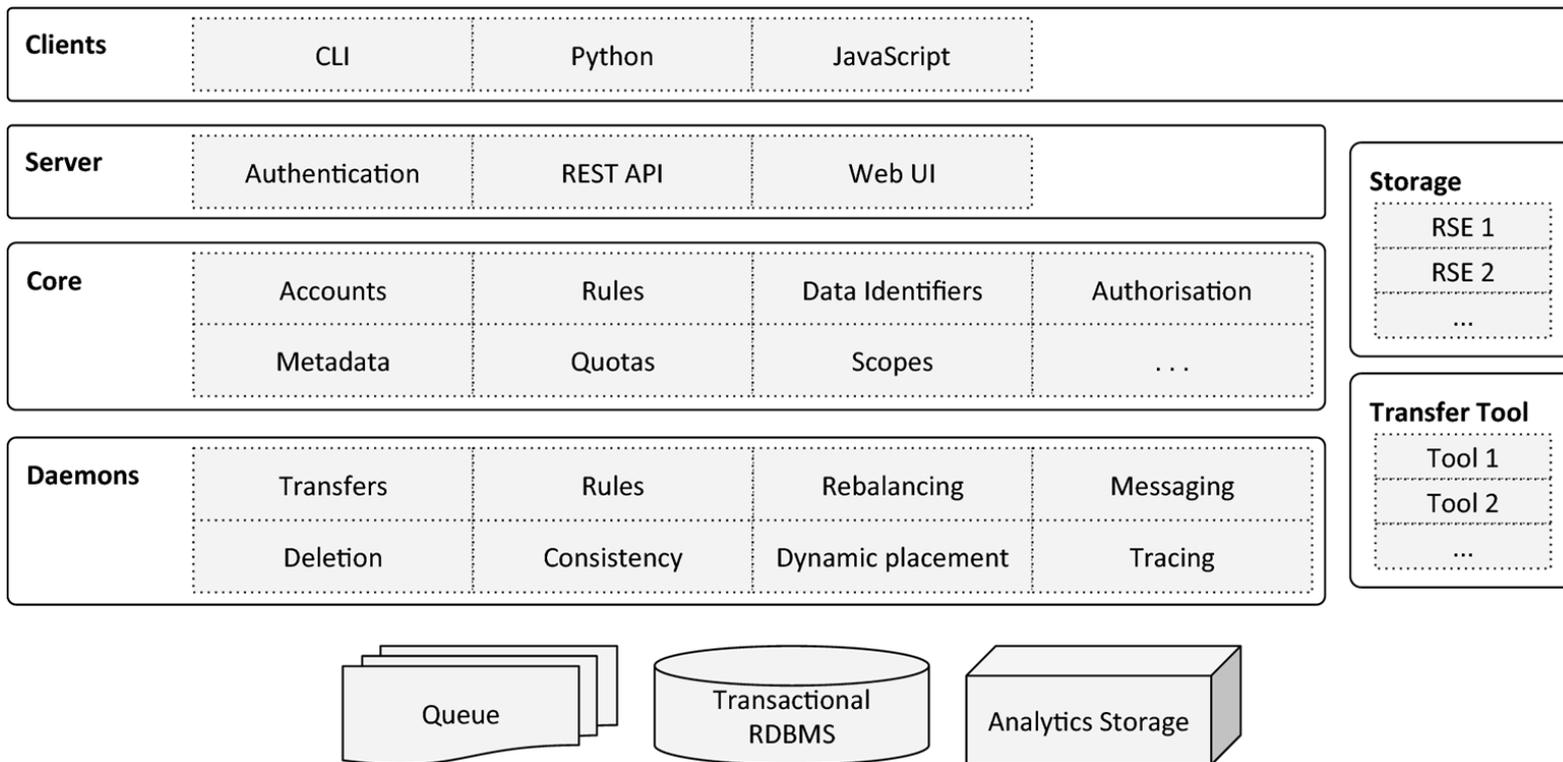
Оценка размеров и количества событий в CDR SPD

Что такое Rucio

Rucio - это открытая система управления научными данными, обеспечивающая организацию и управление данными в распределенной среде хранения. Система разрабатывалась для эксперимента ATLAS на LHC и сейчас используется и для ряда других исследовательских проектов.



Архитектура Rucio



Rucio для эксперимента SPD

Данные современного эксперимента, возникают задолго до начала работы экспериментальной установки. Это данные возникающие в процессе физического моделирования. Объемы и организация этих данных сопоставимы с собираемыми данными.

Для использования Rucio в эксперименте SPD необходимо решить следующие основные задачи:

- развернуть набор необходимых компонент программного комплекса;
- осуществить необходимые предварительную настройку и конфигурирование;
- настроить систему аутентификации и авторизации;
- реализовать первичную организацию данных;
- подключить системы хранения;
- осуществить интеграцию с системой обработки данных.

Развертывание и конфигурирование Rucio

На данный момент в контейнерах развернуты набор необходимых компонент Rucio.

Развертывание реализовано на виртуальной машине из облачной инфраструктуры ОИЯИ с использованием технологии Docker и произведено конфигурирование системы.

```
[konak@vm221-121 dev]$ docker compose ps -a
NAME                IMAGE
dev-activemq-1      docker.io/webcenter/activemq:latest
dev-elasticsearch-1 docker.io/elasticsearch:7.4.0
dev-graphite-1      docker.io/graphiteapp/graphite-statsd
dev-influxdb-1      docker.io/influxdb:latest
dev-rucio-1         docker.io/rucio/rucio-dev:latest-alma9
dev-rucioclient-1   docker.io/rucio/rucio-dev:latest-alma9
dev-ruciodb-1       docker.io/postgres:14
```

```
alex@konak:~$ rucio whoami
status      : ACTIVE
account_type : USER
deleted_at  : None
updated_at  : 2023-11-13T12:25:12
suspended_at : None
account     : konak
email       : konak@jinr.ru
created_at  : 2023-11-13T12:25:12
alex@konak:~$
```

Rucio UI MENU ☰

You are here: Account [konak]

Rucio Version (WebUI / Server): 32.0.0 / 32.0.0

Account info

account	konak
account_type	USER
created_at	2023-11-13T12:25:12
deleted_at	-
email	konak@jinr.ru
status	ACTIVE
suspended_at	-
updated_at	2023-11-13T12:25:12

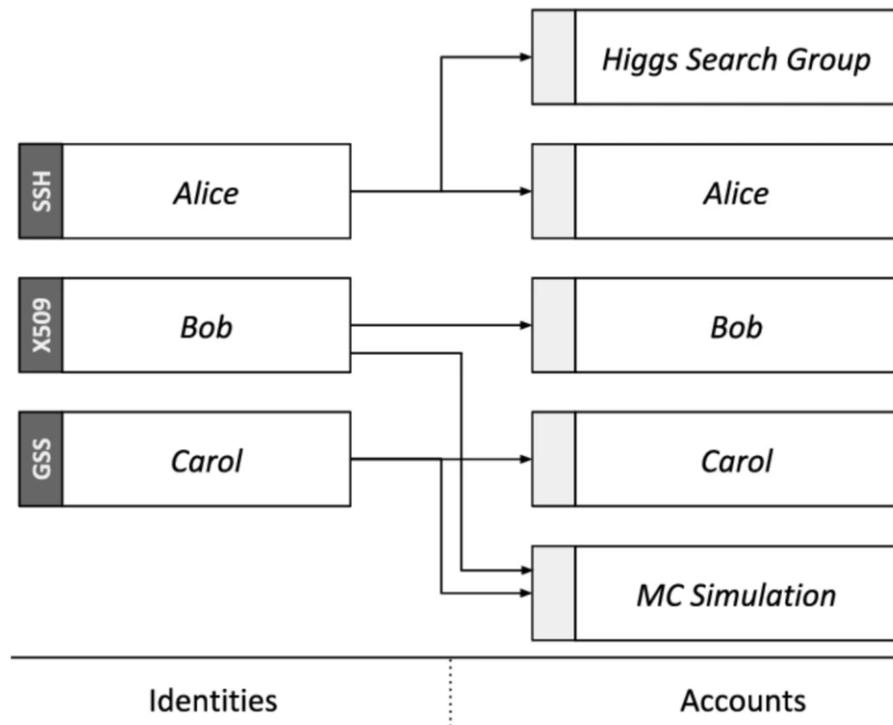
Система аутентификации и авторизации

Готовы к работе:

- сертификаты RDIG X509;
- сертификаты JINR X509.

В процессе:

- IAM;
- JINR SSO.



Аутентификация и авторизация в Rucio SPD

- созданы пользователи;
- добавлены идентификаторы для авторизации;
- реализована аутентификация на основе цифрового сертификата X509.

```
[root@1b9fa2b55ed4 rucio]# rucio-admin account list
konak
panda
root
testgroup1
testgroup2
virthead
```

```
[root@1b9fa2b55ed4 rucio]# rucio-admin account list-identities root
Identity: konak,          type: USERPASS
Identity: CN=Alexey Konak,OU=jinr.ru,OU=users,0=RDIG,C=RU,      type: X509
Identity: CN=Artem Petrosyan,OU=jinr.ru,OU=users,0=RDIG,C=RU,  type: X509
[root@1b9fa2b55ed4 rucio]#
[root@1b9fa2b55ed4 rucio]# rucio-admin account list-identities konak
Identity: konak,          type: USERPASS
Identity: CN=Alexey Konak,OU=jinr.ru,OU=users,0=RDIG,C=RU,      type: X509
```

Первичная организация данных

В эксперименте SPD есть специальная группа отвечающая за формирование правил именования данных.

Результаты работы этой группы реализуются как начальная организация данных в Rucio.

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

data organisation and naming convention

Подключение системы хранения

Описано RSE которое соответствует реальному хранилищу данных Rucio эксперимента SPD на EOS.

extended_attributes	-
hostname	eos.jinr.ru
impl	rucio.rse.protocols.xrootd.Default
port	1094
prefix	/eos/nica/spd/rucio
scheme	xrootd

rse	SPD_PROD_DATADISK
rse_type	DISK
sign_url	-
staging_area	false
third_party_copy_read_protocol	1
third_party_copy_write_protocol	1
verify_checksum	true
volatile	false
write_protocol	1

Особенности работы с Rucio 1/3

Трудности с документацией:

- Некоторые концепции и терминология понятно объясняются новичку, а некоторые довольно неясны пока сам не познакомишься с внутренним устройством системы;
- Отсутствуют технические вопросы - приходится обращаться к исходному коду.

Как пример - не описано как связывать RSE и хранилище, решение этого вопроса было найдено в support чате разработчиков.

Особенности работы с Rucio 2/3

Проблемы с контейнерами:

- Конфигурационные параметры специфичны для инфраструктуры CERN/ATLAS;
- Необходимы дополнительные действия для настройки окружения для его правильного функционирования такие как выбор протоколов, добавление корневых сертификатов и обеспечение их доверенности и т.д.

Особенности работы с Rucio 3/3

Проблемы в использовании:

- Некоторые интересные моменты при использовании CLI связанные с историей развития проекта.

Добавление идентификатора для аккаунта выполняется командой `rucio-admin identity add ...` , а посмотреть список идентификаторов можно командой `rucio-admin account list-identities ...`

- объекты призраки - после удаления RSE он еще существует в БД, возможности полного его удаления или восстановления не имеются, невозможно получить список удаленных RSE.

Результаты

Выступление на GRID'2023 в рамках секционного доклада, тема доклада - "Scientific data management system for the SPD experiment";

Одна публикация - А. Конак, А. Петросян. Scientific data management system for the SPD experiment - принята к публикации в журнале "Физика элементарных частиц и атомного ядра";

На данный момент система управления данными Rucio версии 32.0.0 на базе AlmaLinux 9 функционирует в контейнерах на виртуальной машине vm221-121.jinr.ru облачной инфраструктуры ОИЯИ. Произведено начальное конфигурирование, реализованы аккаунты и идентификаторы для входа. Реализована аутентификация в систему на основе цифровых сертификатов X509.

Задачи на следующий год

- ввод системы в режим опытной эксплуатации;
- интеграция системы с информационным каталогом вычислительных ресурсов CRIC;
- интеграция системы управления данными с хранилищами данных;
- интеграция с сервисом передачи данных FTS;
- тестирование взаимодействия с системой управления нагрузкой PanDA в рамках обработки задач;
- описание модели данных эксперимента SPD в терминах Rucio;
- управление хранилищем данных SPD на EOS, CTA и хранилищами данных членов коллаборации;
- реализация модели аутентификации на основе технологии JSON Web Token.

Спасибо за внимание!