



Разработка подсистемы управления данными комплекса промежуточного программного обеспечения SPD Online filter

Студент	П. А. Коршунова
Научный руководитель	Е. Ю. Солдатов
Научный консультант	Д. А. Олейник

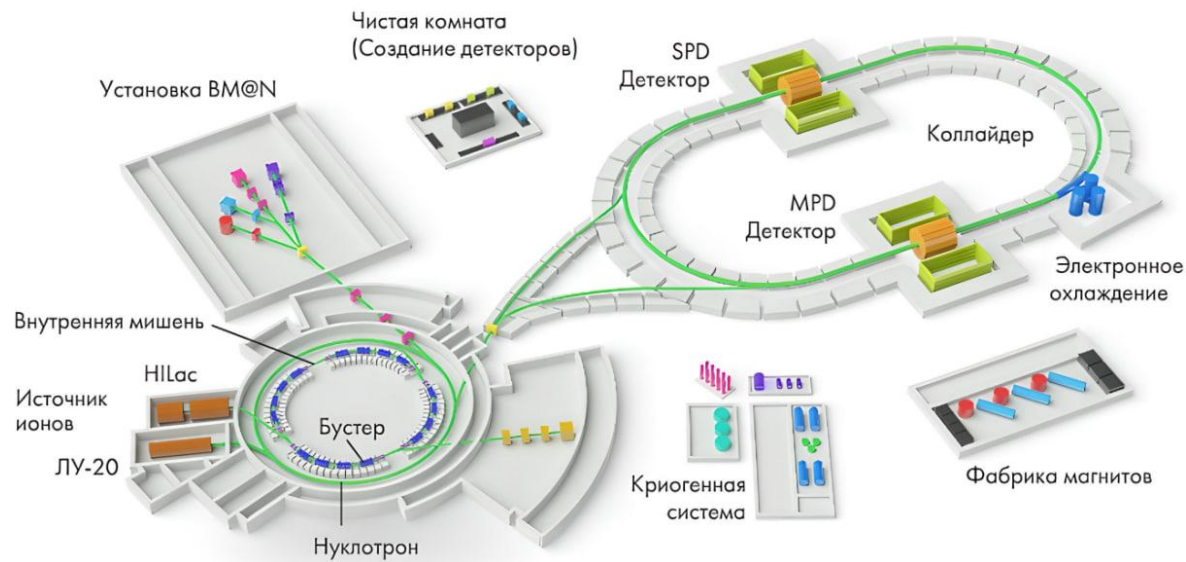


Цель и задачи работы

Цель: концептуальная доработка и реализация системы управления данными, которая является частью вычислительного комплекса SPD Online filter

- ✓ Доработать **взаимодействие** системы управления данными с **другими компонентами** комплекса SPD Online Filter – системой управления процессами обработки и системой управления нагрузкой
- ✓ Реализовать фоновый сервис для **контроля состояния данных** на хранилищах
- ✓ Провести этап **интеграционного тестирования** между системами (выявить неточности в реализованном функционале и доработать его)

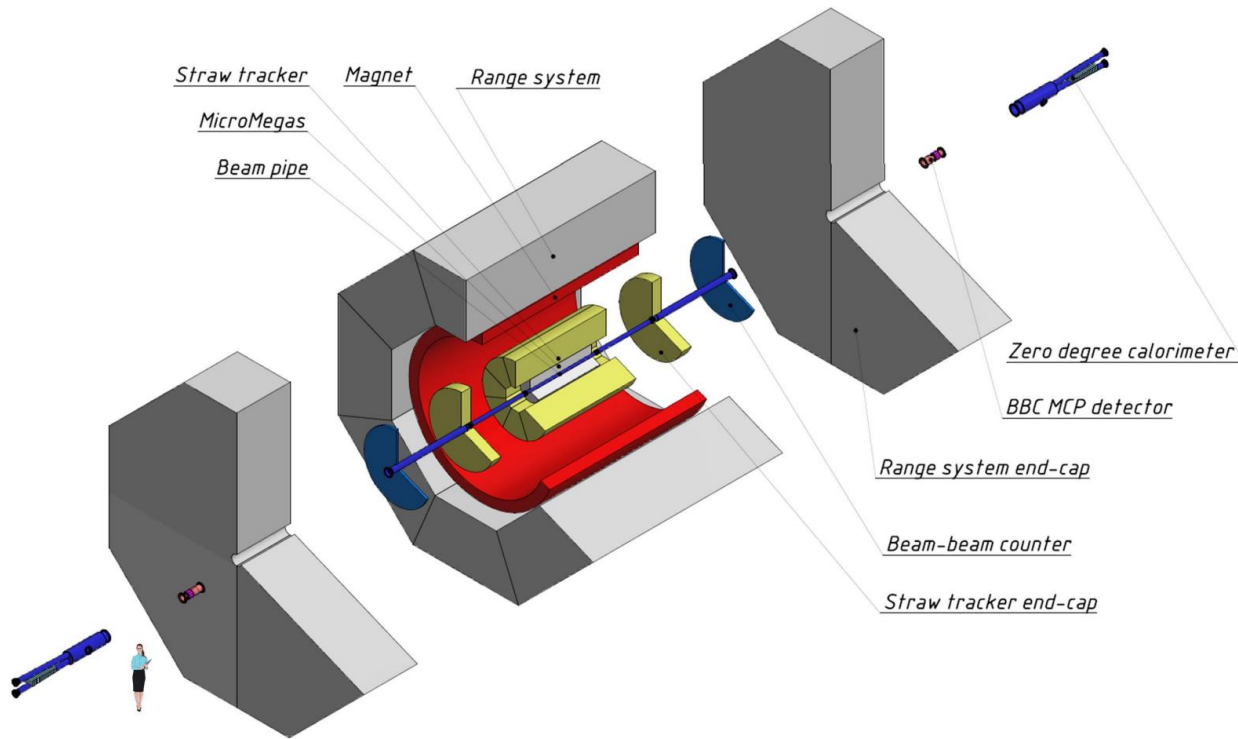
Ускорительный комплекс NICA



Мегасайнс проект NICA нацелен на **воссоздание** и исследование **ядерной материи в экстремальных условиях**, возникавших в природе на ранних стадиях эволюции Вселенной

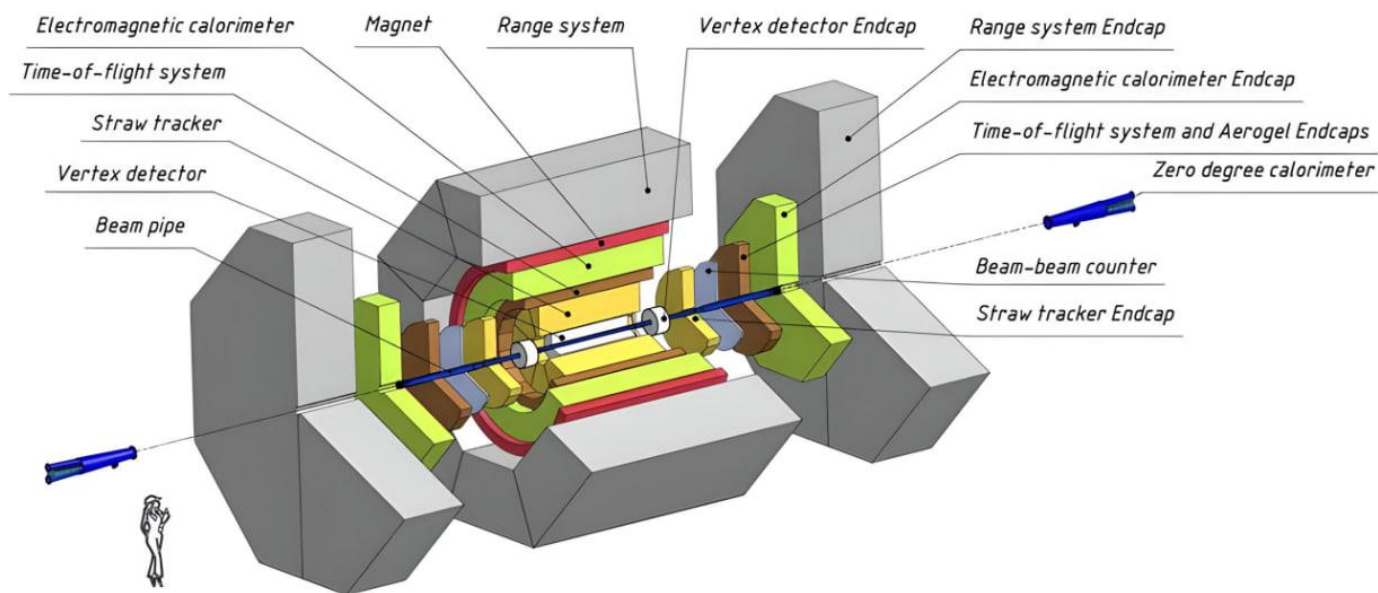
- **MPD** (MultiPurpose Detector)
 - исследование плотной барионной материи на встречных пучках
- **BM@N** (Baryonic Matter at Nuclotron)
 - исследование плотной барионной материи на выведенных пучках нуклотрона
- **SPD** (Spin Physics Detector)
 - изучение спиновой структуры протона и дейтрона и других связанных со спином явлений
 - Поляризованные столкновения
 - pp до 27 ГэВ
 - dd до 13.5 ГэВ
 - pd до 19 ГэВ

SPD: стадия 1

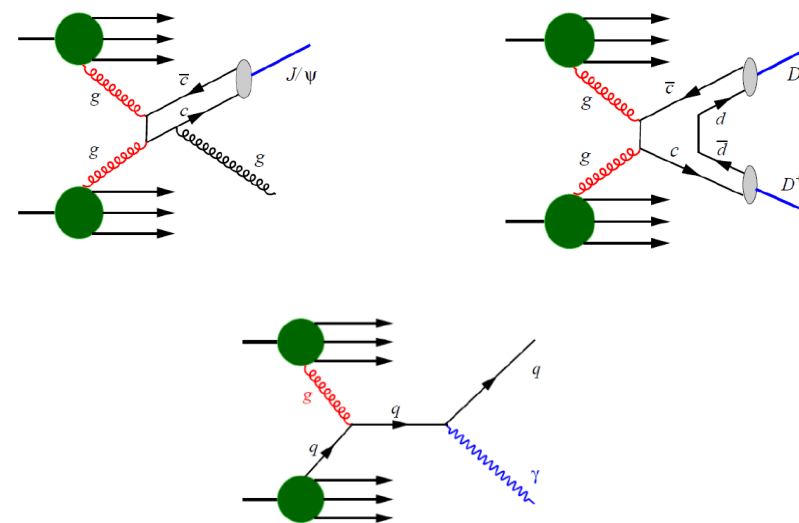


- Исследование **спиновых эффектов** в упругом pp и dd рассеянии и в образовании гиперонов
- Изучение **многокварковых корреляций**
- Изучение реакций с образованием **дибарионов** и странных **гиперядер**
- Образование **чармония** близко к пороговому значению

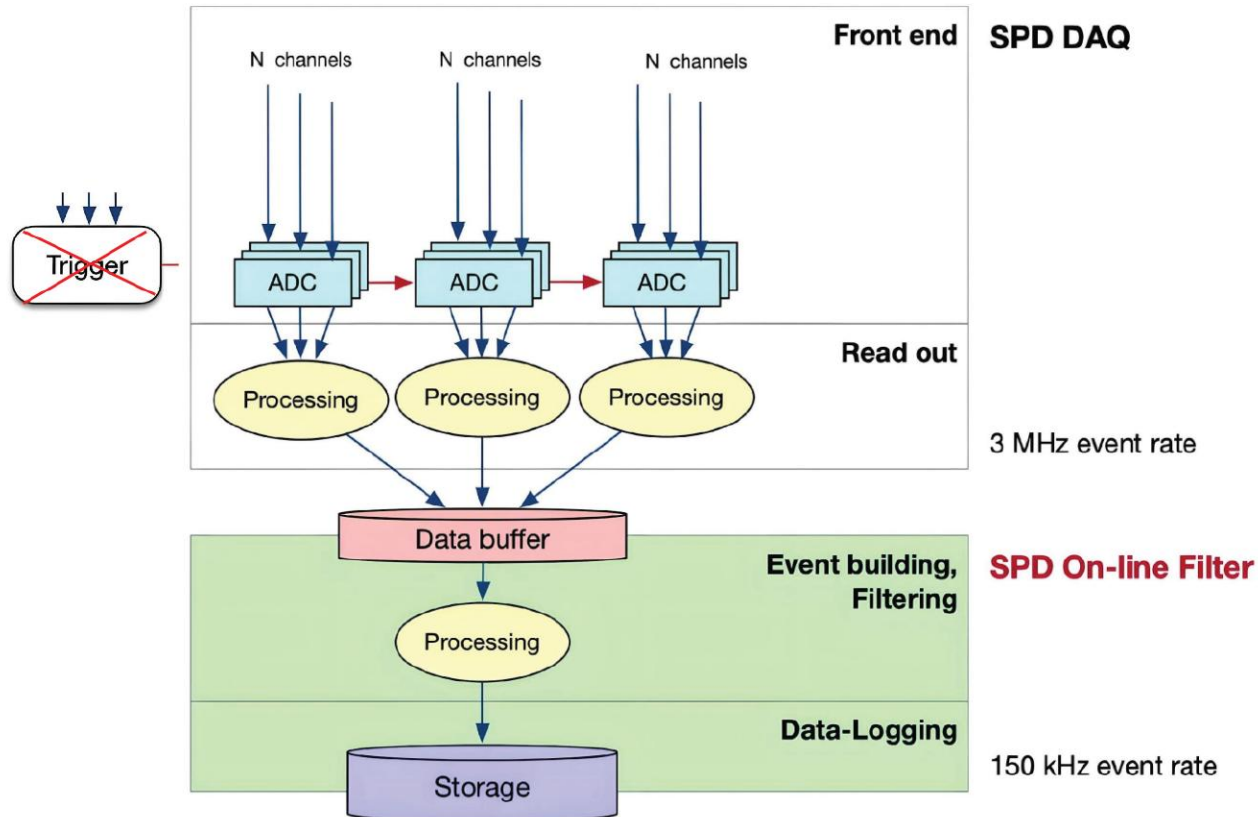
SPD: стадия 2



На второй стадии эксперимента основное внимание уделяется исследованию распределения глюонов в реакциях **рождения чармония, открытого очарования и быстрых фотонов**



Особенность сбора данных с детектора



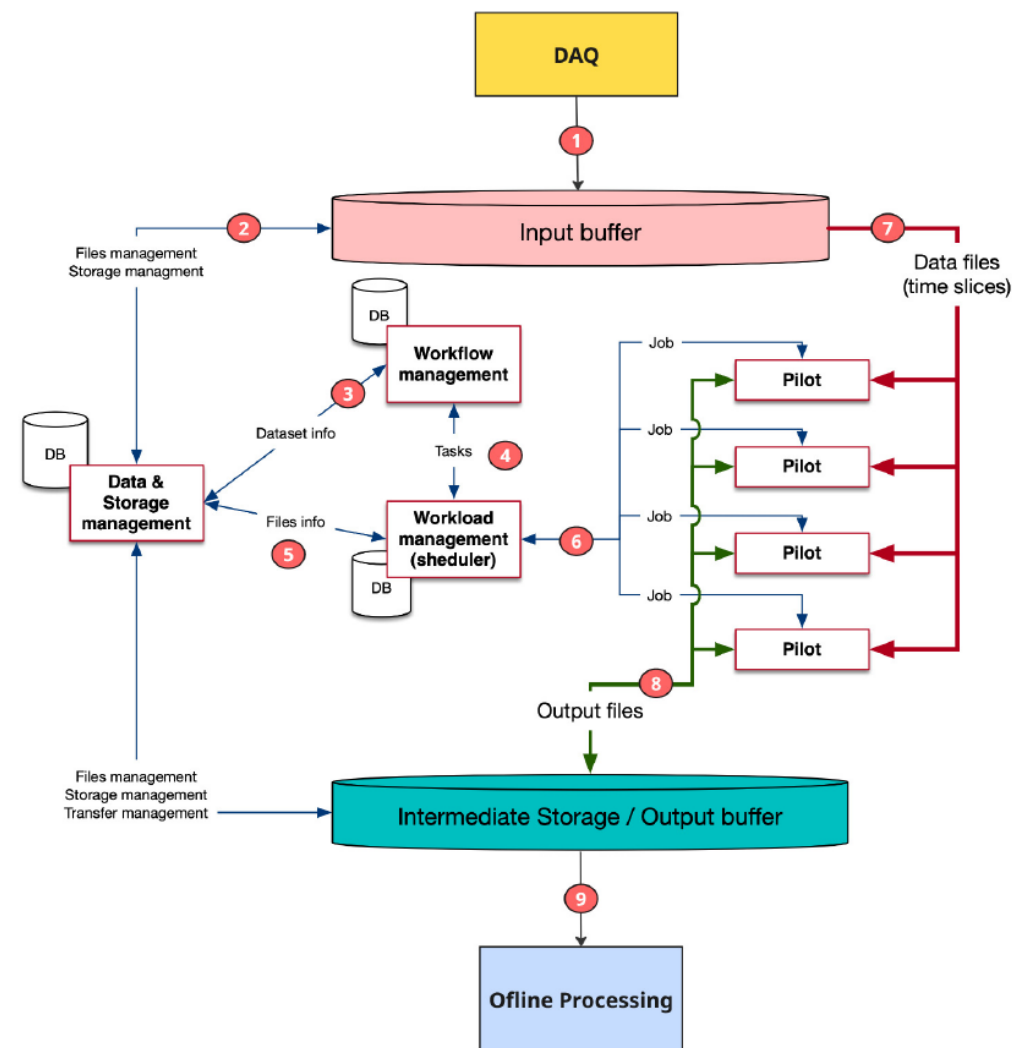
- Невозможен простой выбор физических событий на аппаратном уровне
- Необходимость сбора всего набора генерируемых сигналов с подсистем, объединенных в блоки по времени
- Большой поток данных до 20 ГБ/с (~ 200 ПБ/год)
- Необходимость сократить объем данных для последующего анализа
- Разработка программного триггера – SPD Online filter

Online Filter

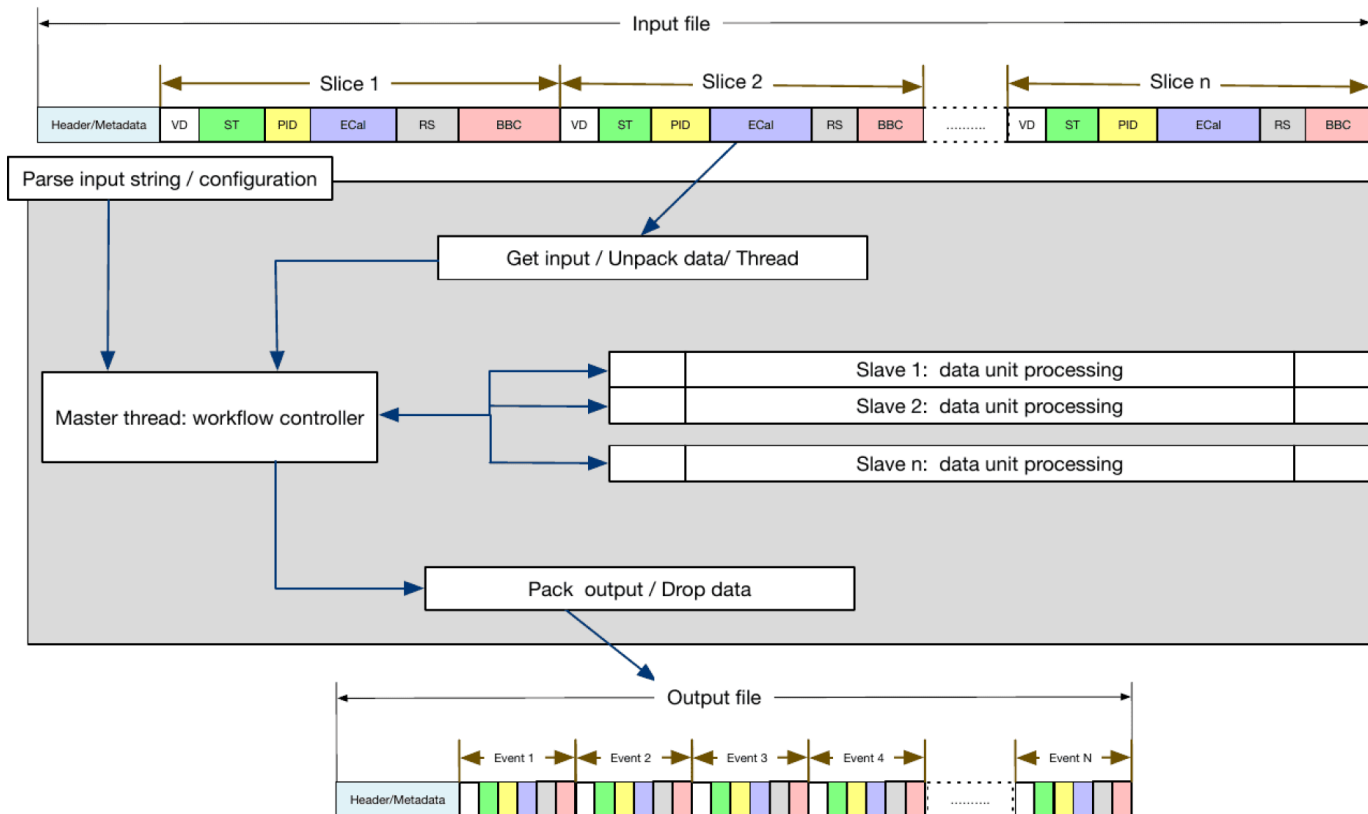
SPD Online Filter - это высокопроизводительная вычислительная система для высокопропускной обработки данных, собранных детектором SPD

Особенностью **высокопропускной обработки** данных является большой объем данных, как первичных, которые необходимо обработать, так и промежуточных, возникающих в процессе обработки

Цель: существенно сократить объем данных для долговременного хранения, последующей обработки и анализа

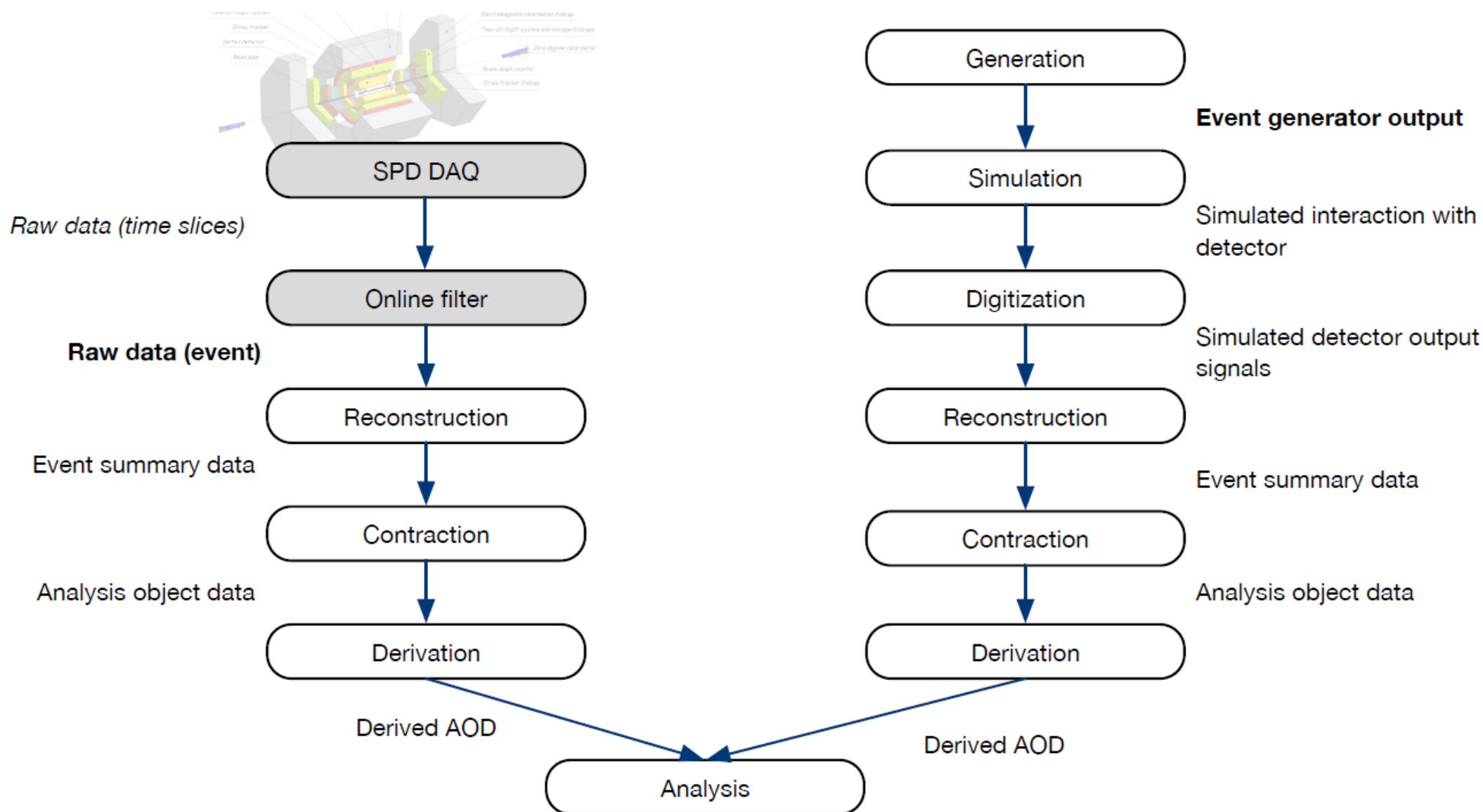


Обработка данных



1. Декодирование данных
2. Частичная реконструкция событий
 - a. Реконструируем **треки** и связываем их с вершинами
 - b. Для каждой вершины определяем **время пересечения пучков**
 - c. Связываем **срабатывания ECal** и **RS** с каждой вершиной
 - d. Присоединяем **несвязанные треки** и **необработанные данные** с других поддетекторов в соответствии со временем пересечения пучков
 - e. Блок информации, связанный с каждой вершиной, называем **событием**
3. Фильтрация событий
4. Подготовка данных для оффлайн обработки

Роль SPD Online filter



Промежуточное ПО

Система управления данными

- регистрация, каталогизация, контроль целостности

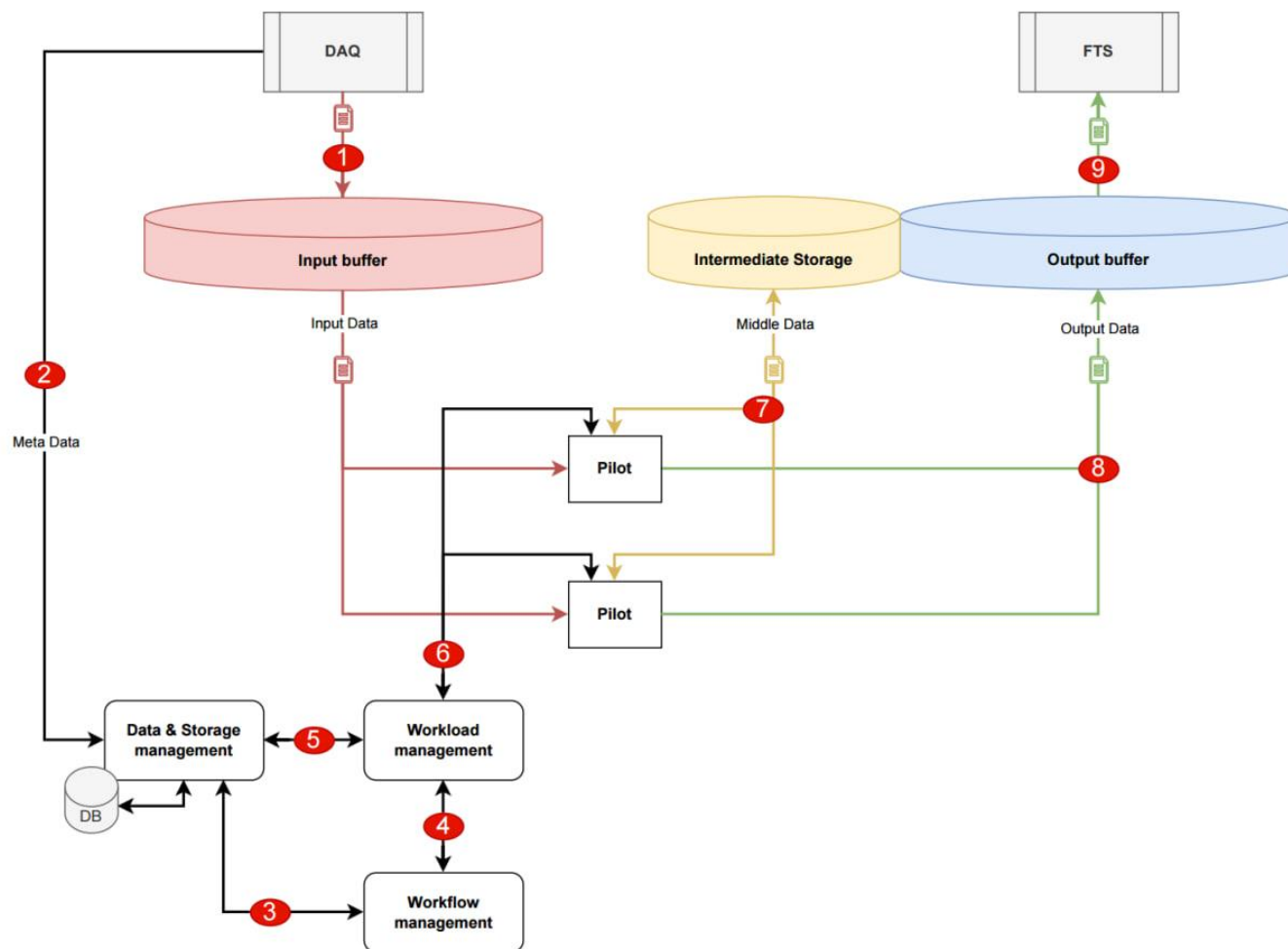
Система управления процессами обработки

- формирование и контроль исполнения этапов обработки данных

Система управления нагрузкой

- реализация этапов обработки (формирование задач, отправка задач пилотам)

Pilot – приложение, работающее на вычислительном узле и исполняющее задачи

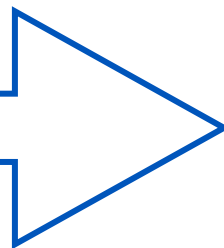


Зачем нужна система управления данными?

Большой поток данных с
детектора
до 20 ГБ/сек

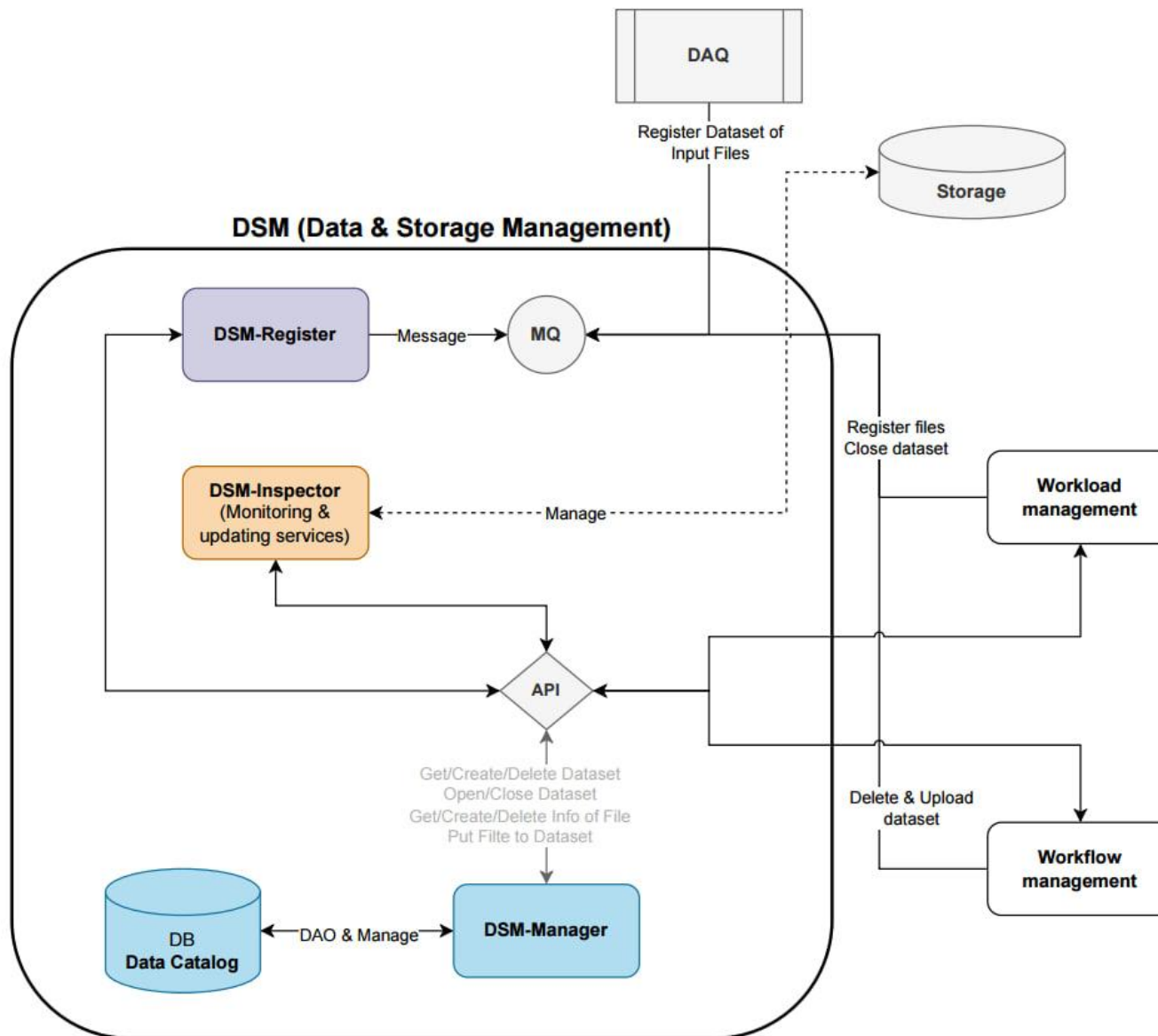


Большой объем
промежуточных данных,
полученных в процессе
обработки



Необходимость разработки
**системы управления
данными**, которая должна
обеспечивать контроль над
организацией, хранением
и целостностью данных

Система управления данными



dsm-register

- сервис, принимающий в асинхронном режиме (через очередь сообщений) заявки на добавление/удаление данных в системе

dsm-manager

- сервис, предоставляющий REST API к каталогу данных (размещение данных в каталоге, обращение к каталогу, изменение данных в каталоге)

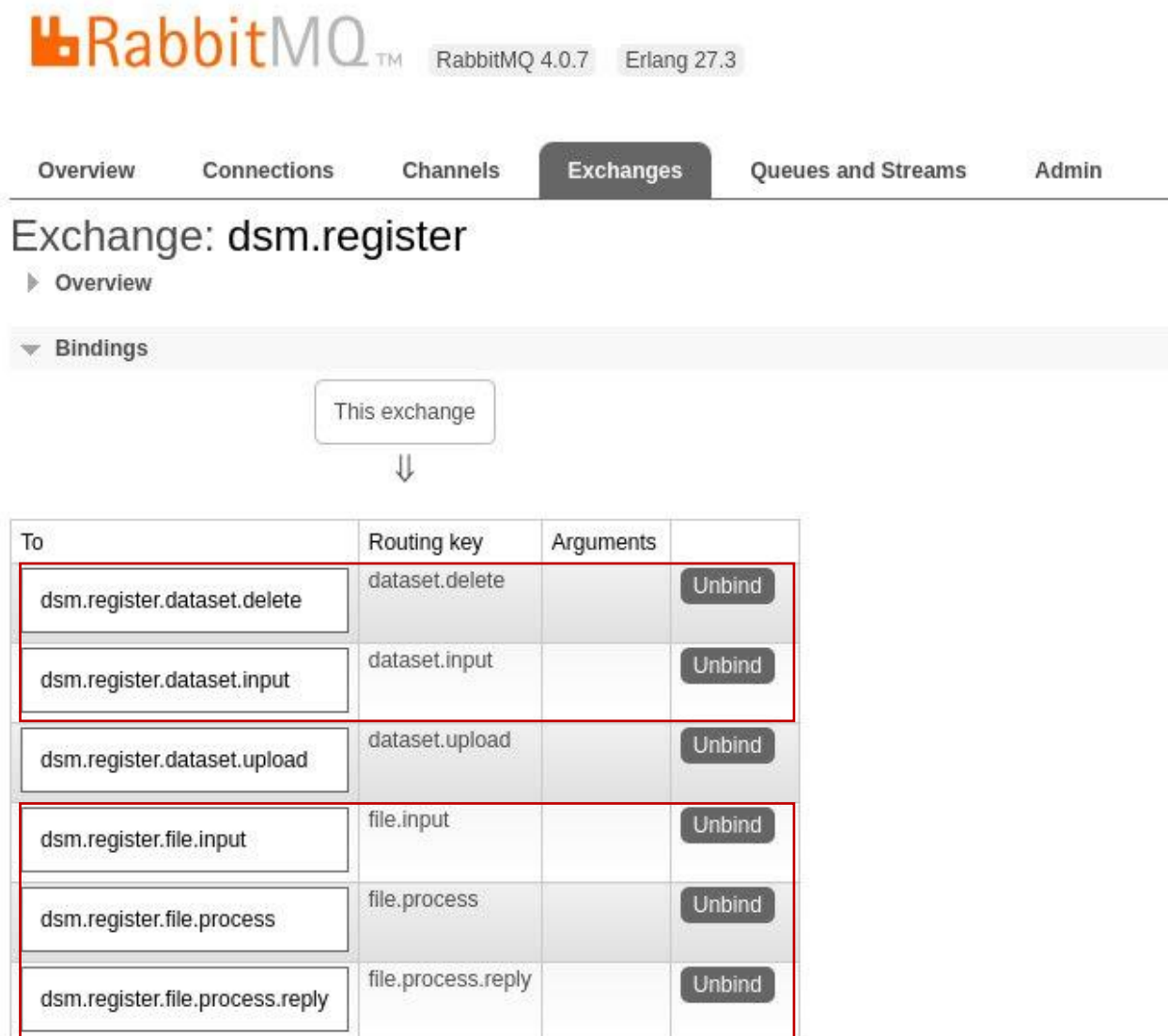
dsm-inspector

- набор фоновых сервисов для мониторинга и контроля состояния данных в хранилище

dsm-register: сконфигурированные очереди

Сервис должен слушать очередь сообщений и обрабатывать заявки на добавление/удаление данных в системе

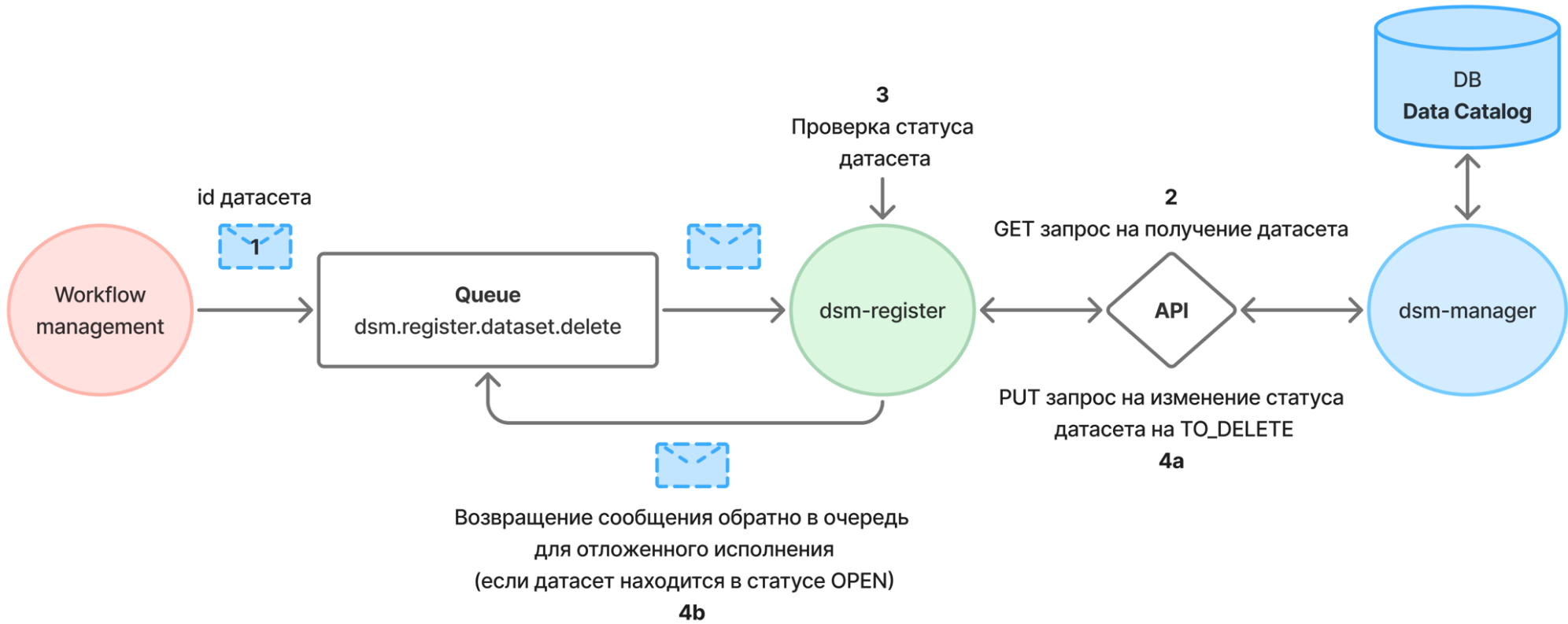
В качестве AMQP-брокера, который осуществляет маршрутизацию и подписку на нужные очереди, используется [RabbitMQ](#)



The screenshot shows the RabbitMQ web interface with the 'dsm.register' exchange selected. The 'Bindings' section is expanded, showing a list of queues bound to the exchange. A diagram indicates that the exchange routes messages to these queues based on their routing keys.

To	Routing key	Arguments	
dsm.register.dataset.delete	dataset.delete		Unbind
dsm.register.dataset.input	dataset.input		Unbind
dsm.register.dataset.upload	dataset.upload		Unbind
dsm.register.file.input	file.input		Unbind
dsm.register.file.process	file.process		Unbind
dsm.register.file.process.reply	file.process.reply		Unbind

dsm.register.dataset.delete



dsm-manager: API к БД

file			^
GET	/api/v1/file/	Get List	✓
POST	/api/v1/file/	Add	✓
GET	/api/v1/file/{file_id}	Get By Id	✓
PUT	/api/v1/file/{file_id}	Update	✓
DELETE	/api/v1/file/{file_id}	Remove	✓
GET	/api/v1/file/file_name/{file_name}	Get By Name	✓
dataset			^
GET	/api/v1/dataset/	Get List	✓
POST	/api/v1/dataset/	Add	✓
GET	/api/v1/dataset/{dataset_id}	Get By Id	✓
PUT	/api/v1/dataset/{dataset_id}	Update	✓
DELETE	/api/v1/dataset/{dataset_id}	Remove	✓
PATCH	/api/v1/dataset/{dataset_id}	Update Status	✓
GET	/api/v1/dataset/name/{dataset_name}	Get By Name	✓

- ✓ Получение содержимого набора
- ✓ Получение списка файлов в определенном статусе



Сервис должен предоставлять REST API к каталогу данных.

В качестве веб-фреймворка используется асинхронный фреймворк **FastAPI**

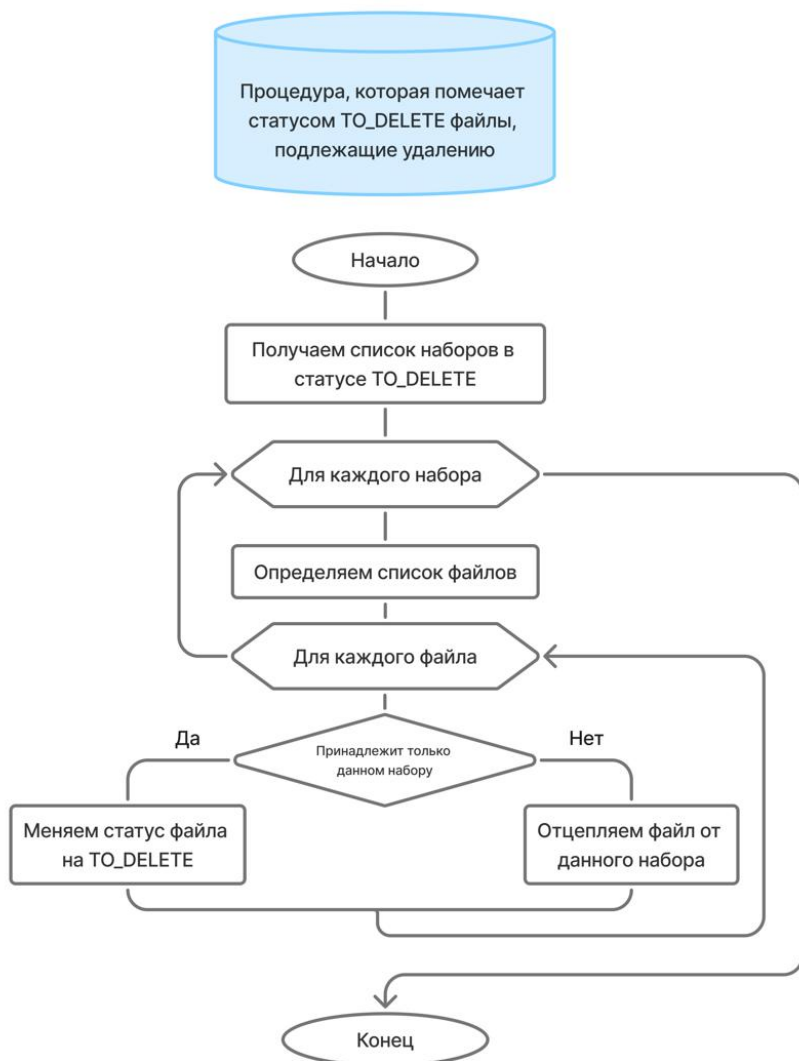
- ✓ Получение списка датасетов, в которых есть определенный файл
- ✓ Получение списка датасетов в определенном статусе

Сервис состоит из набора **фоновых** задач:

- ✓ Удаление файлов на хранилищах
- ✓ Контроль выгрузки файлов
- ✓ Проверка целостности файлов
- ✓ Контроль использования хранилища

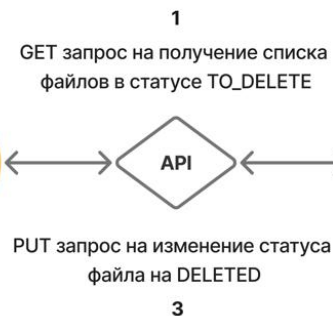
Исполнение заявок на удаление датасетов

1. Определение списка файлов, подлежащих удалению

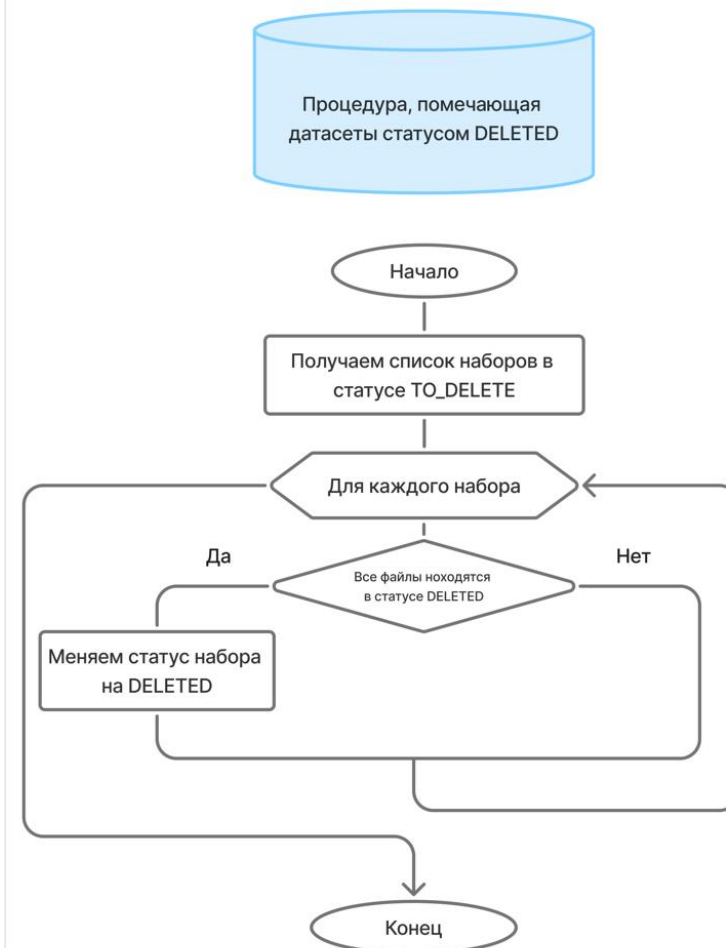


2. Удаление файлов в датасетах

2
Удаление файлов с хранилища



3. Удаление датасетов



SPD DAQ generator



- Используя SPD DAQ data generator было сгенерировано 50 файлов по 2 ГБ (по ~ 7 минут на файл)
- Создан входной датасет с данными файлами (~ 10 секунд)

```
{  
  "name": "SPD_Run1001_Dataset_raw_run_10032025",  
  "metaData": {  
    "run_number": 1,  
    "files": 50  
  },  
  "statusCode": "CLOSED",  
  "id": "03297720-9b97-4cc6-a560-2d5c3ba776a5"  
}
```

- Сгенерировано задание и создан выходной датасет

```
{  
  "name": "SPD_Run1001_Dataset_raw_run_10032025.output.1",  
  "metaData": {  
    "task_id": 1  
  },  
  "statusCode": "OPEN",  
  "id": "5f9c548f-a684-4dbc-be33-90e8d007d6cf"  
}
```

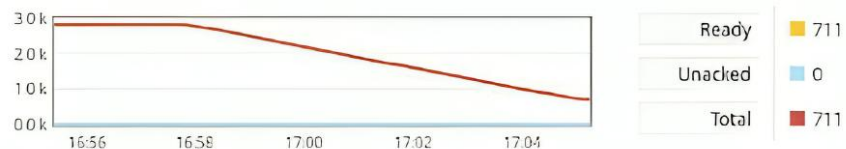
- Файлы были успешно обработаны pilot-ами

Нагрузочное тестирование

Queue pilot-CPU

Overview

Queued messages last ten minutes ?



Message rates last ten minutes ?



100 конкурирующих Pilot-ов обработали ~ 2,100 задач за 7 минут (~15 сек/задача) на стандартной JINR Cloud VM: 12x 1-core Intel Xeon E5-2650

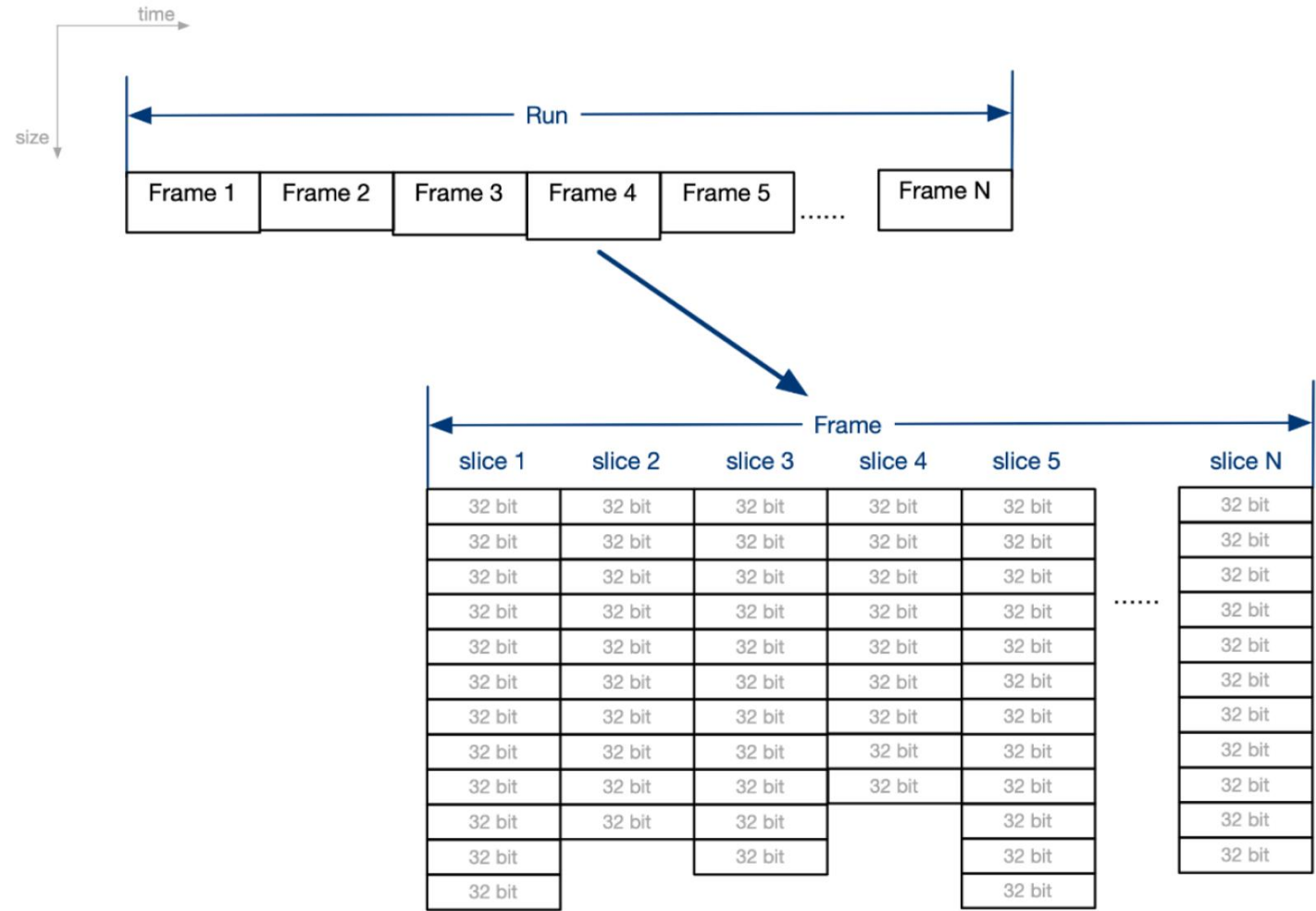
Заключение

- ✓ **dsm-manager** полностью функционален для данного этапа
- ✓ **dsm-register** реализован для данного этапа
 - Прием заявок на регистрацию входных файлов
 - Оповещение WFMS о входных датасетах
 - Прием заявок на регистрацию промежуточных файлов
 - Оповещение WMS о статусе регистрации файла
 - Прием заявок на удаление датасетов
- ✓ **dsm-inspector** реализован на 75%
 - Проверка целостности файлов
 - Исполнение заявок на удаление датасетов
 - Удаление темных файлов
 - Мониторинг заполненности хранилища
 - Мониторинг хранилища на предмет темных файлов
- ✓ Проведено первое **нагрузочное тестирование**, все операции с данными выполнены успешно → разработана **надежная** система

Дополнительные слайды



Организация входных данных



Ран – интервал, когда условия эксперимента неизменны

Фрейм – нумерованный блок данных с сенсоров

Файл – количество данных для обработки одной задачей. В одном файле содержится информация о множестве событий.

Система управления данными: задачи

✓ Регистрация новых данных

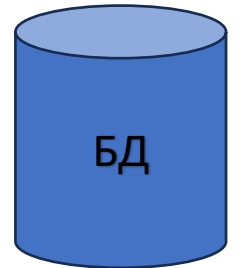
Интерфейс для
приема заявок на
добавление/удаление
данных в системе

✓ Каталогизация

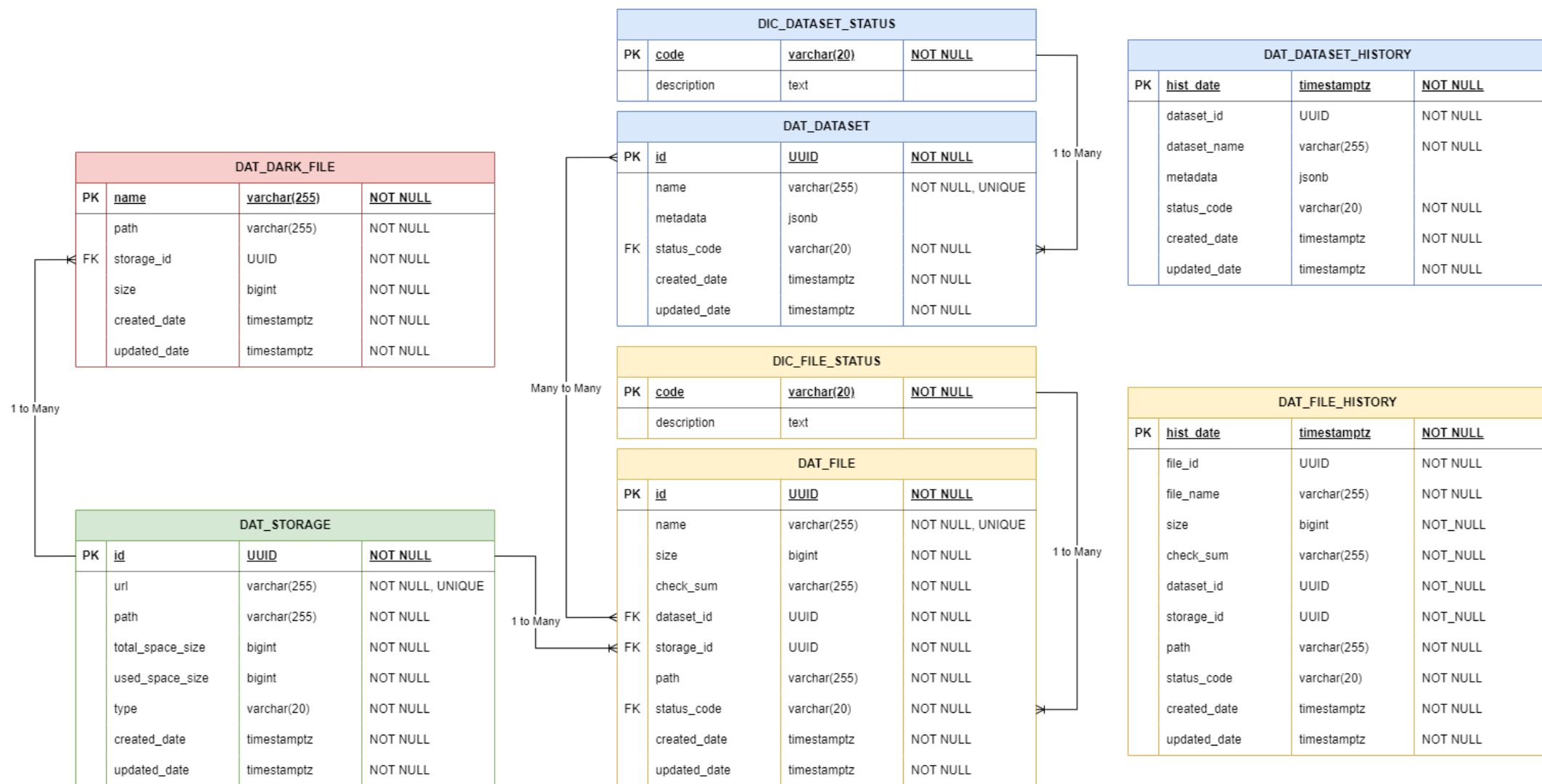
Интерфейс к каталогу
данных

✓ Контроль целостности и выгрузки файлов,
удаление файлов на хранилищах,
мониторинг хранилища

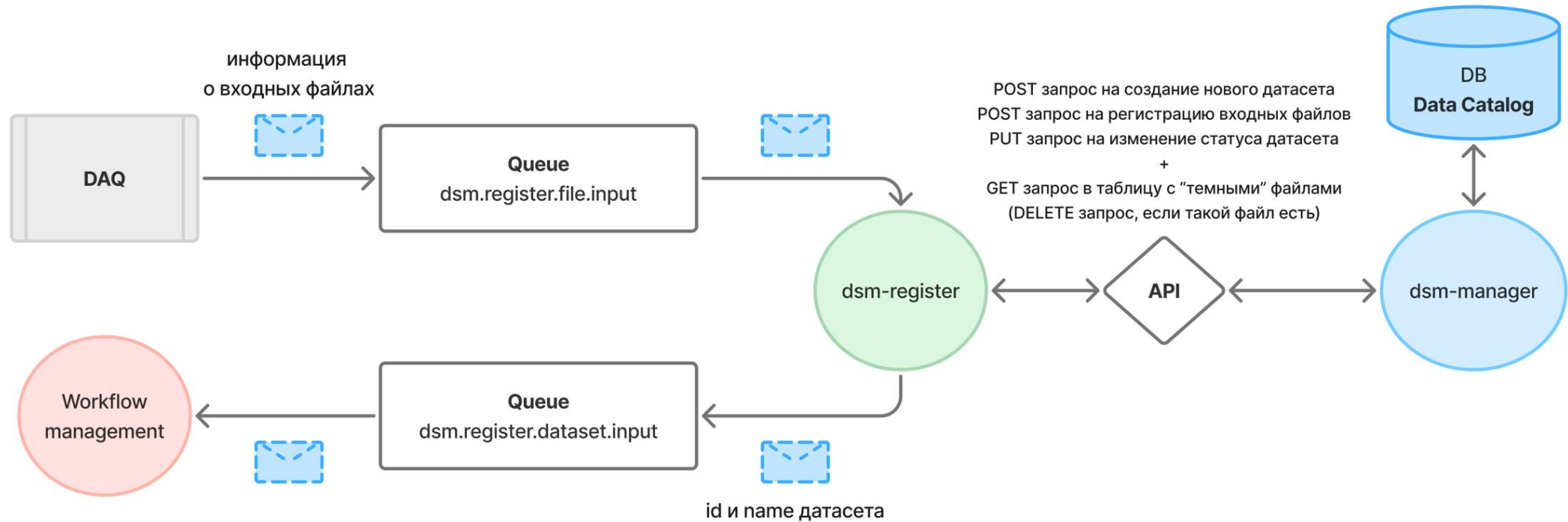
Набор фоновых задач
для обеспечения
согласованности
хранилища и каталога



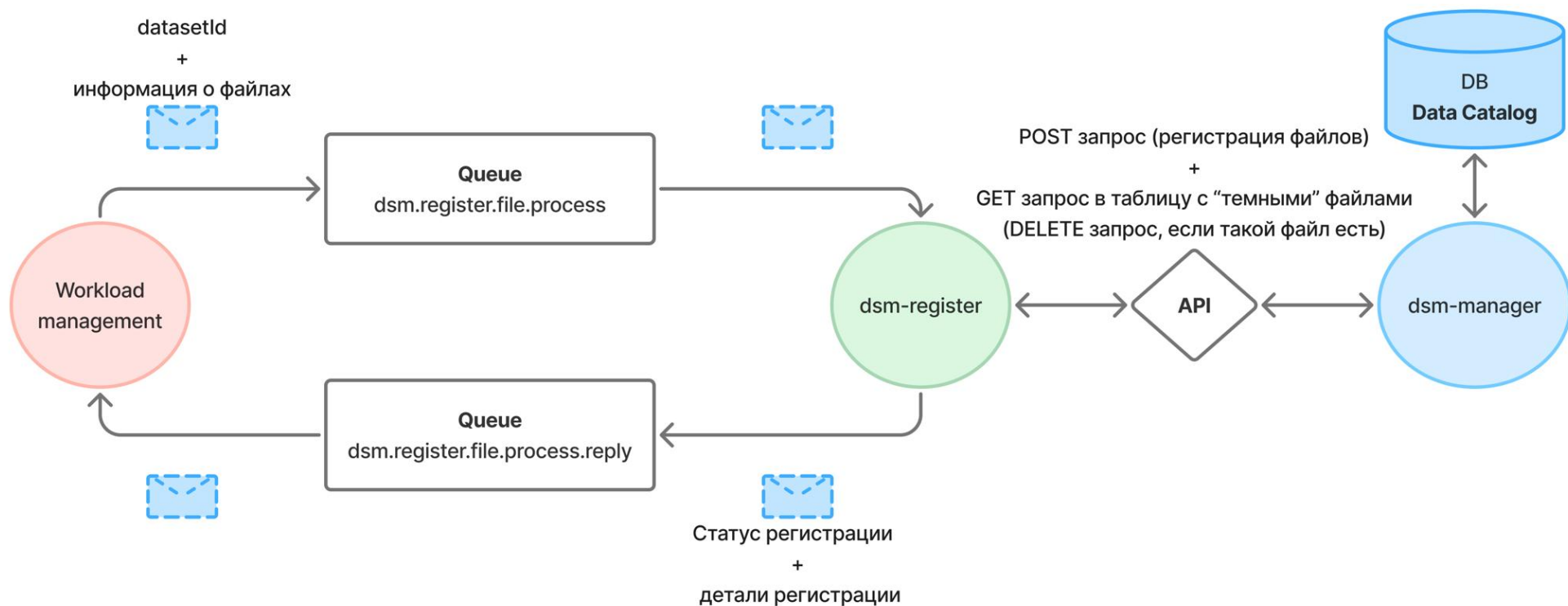
Структура БД



dsm.register.file.input и dsm.register.dataset.input



dsm.register.file.process и dsm.register.file.process.reply



dsm-manager: API к БД

storage ^

GET /api/v1/storage/ Get List

POST /api/v1/storage/ Add

GET /api/v1/storage/{storage_id} Get By Id

PUT /api/v1/storage/{storage_id} Update

DELETE /api/v1/storage/{storage_id} Remove

GET /api/v1/storage/type/{storage_type} Get By Type

dark_file ^

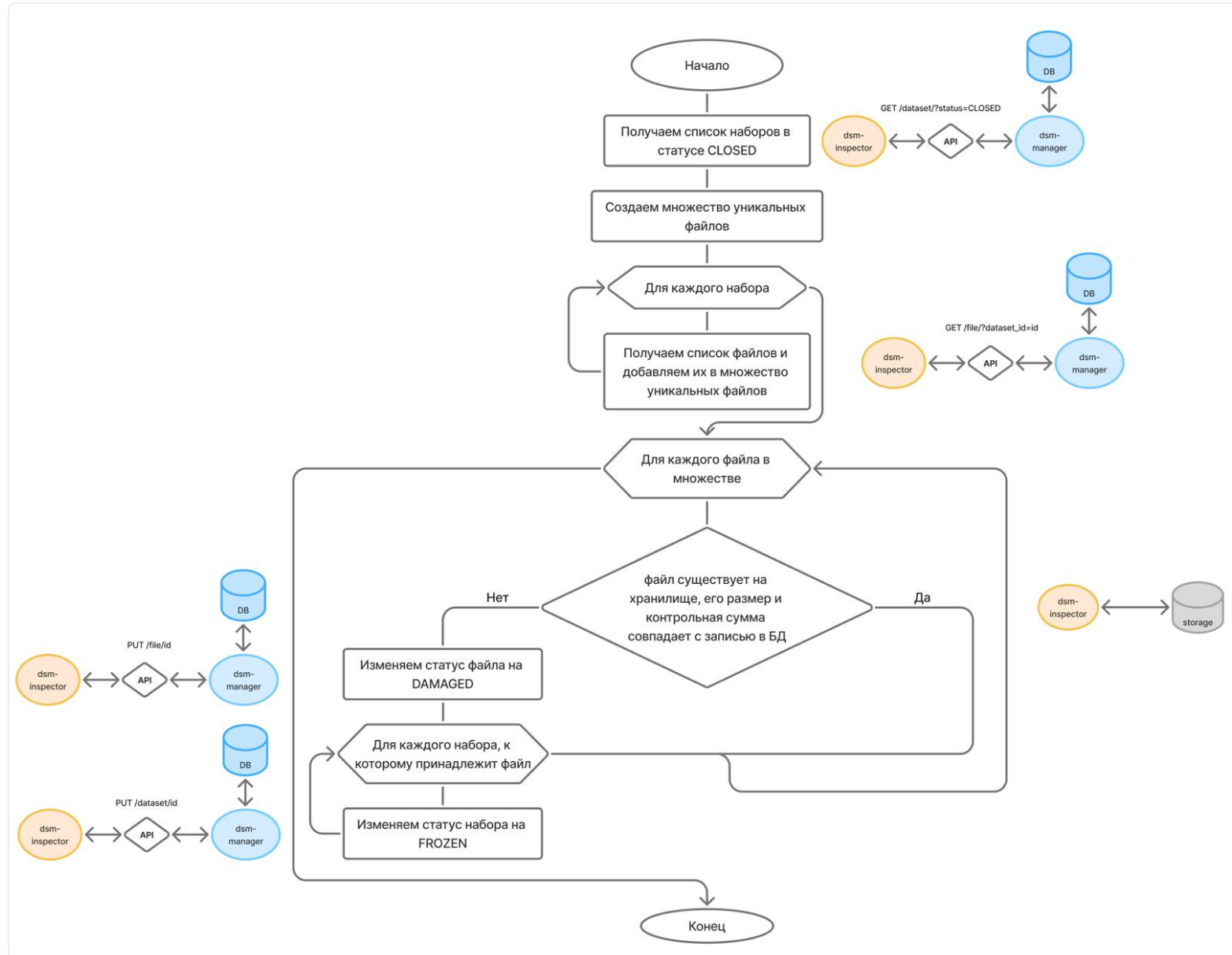
GET /api/v1/dark_file/ Get List

POST /api/v1/dark_file/ Add

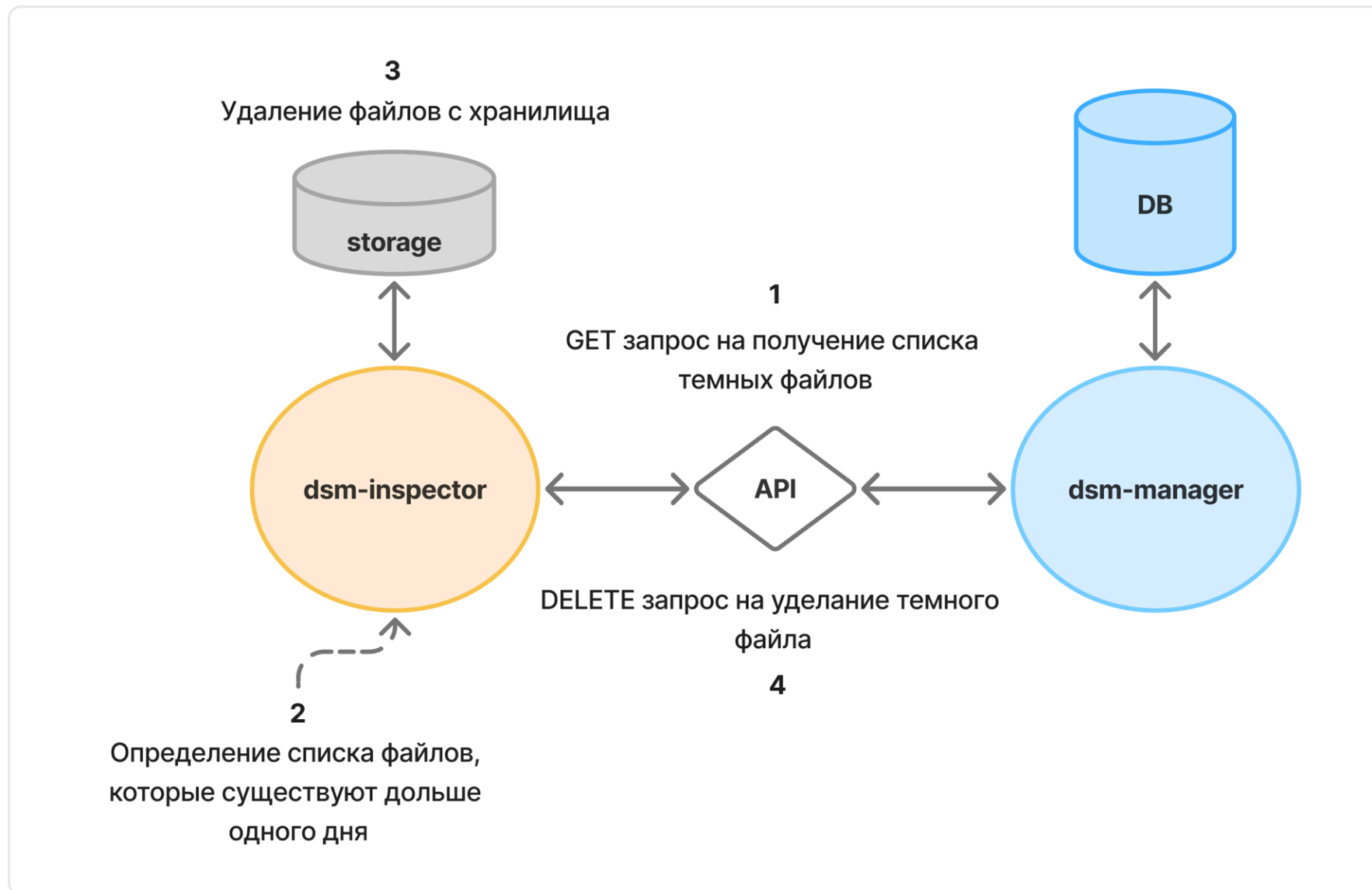
GET /api/v1/dark_file/{file_name} Get By Name

DELETE /api/v1/dark_file/{file_name} Remove

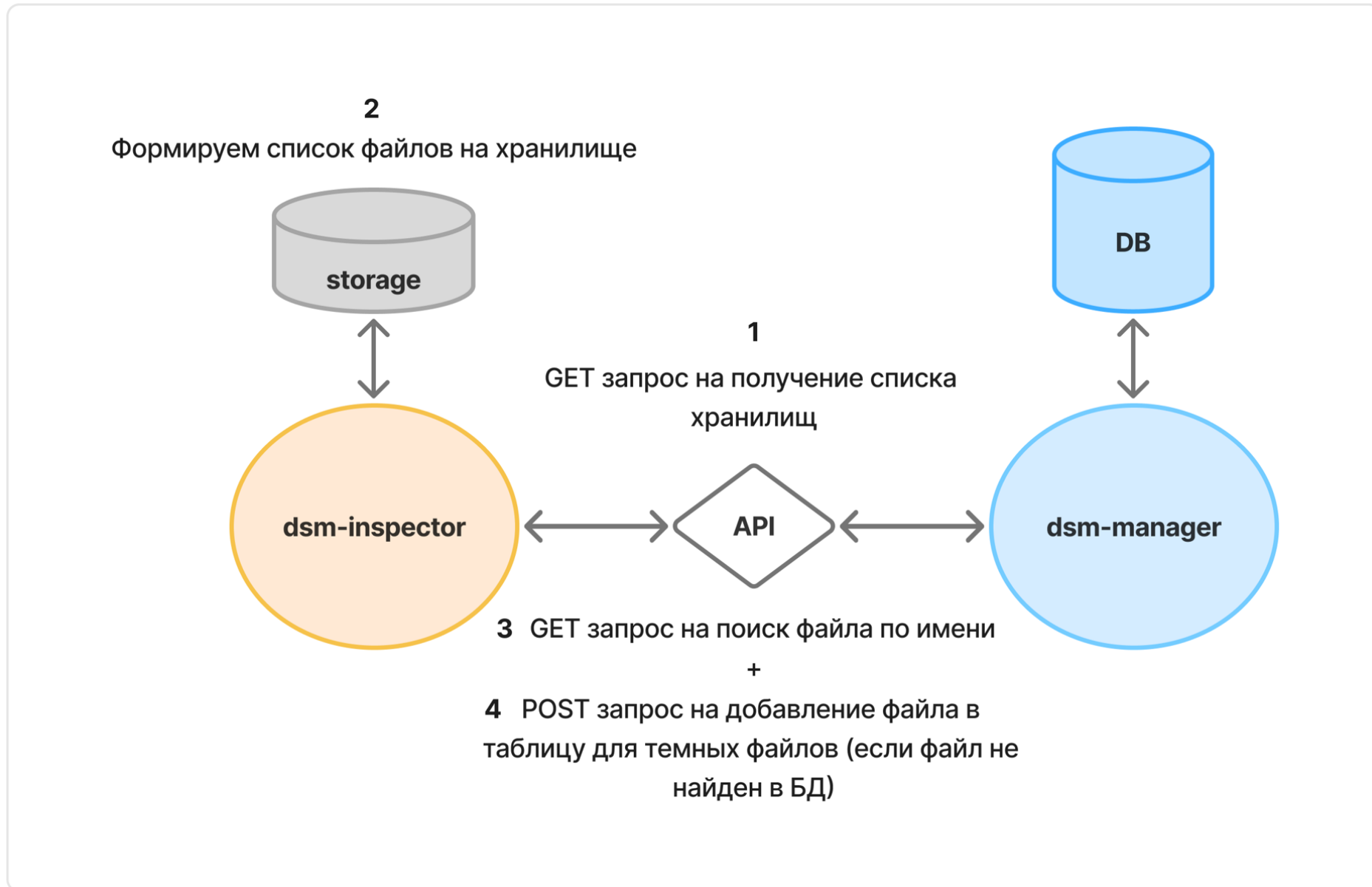
Проверка целостности файлов



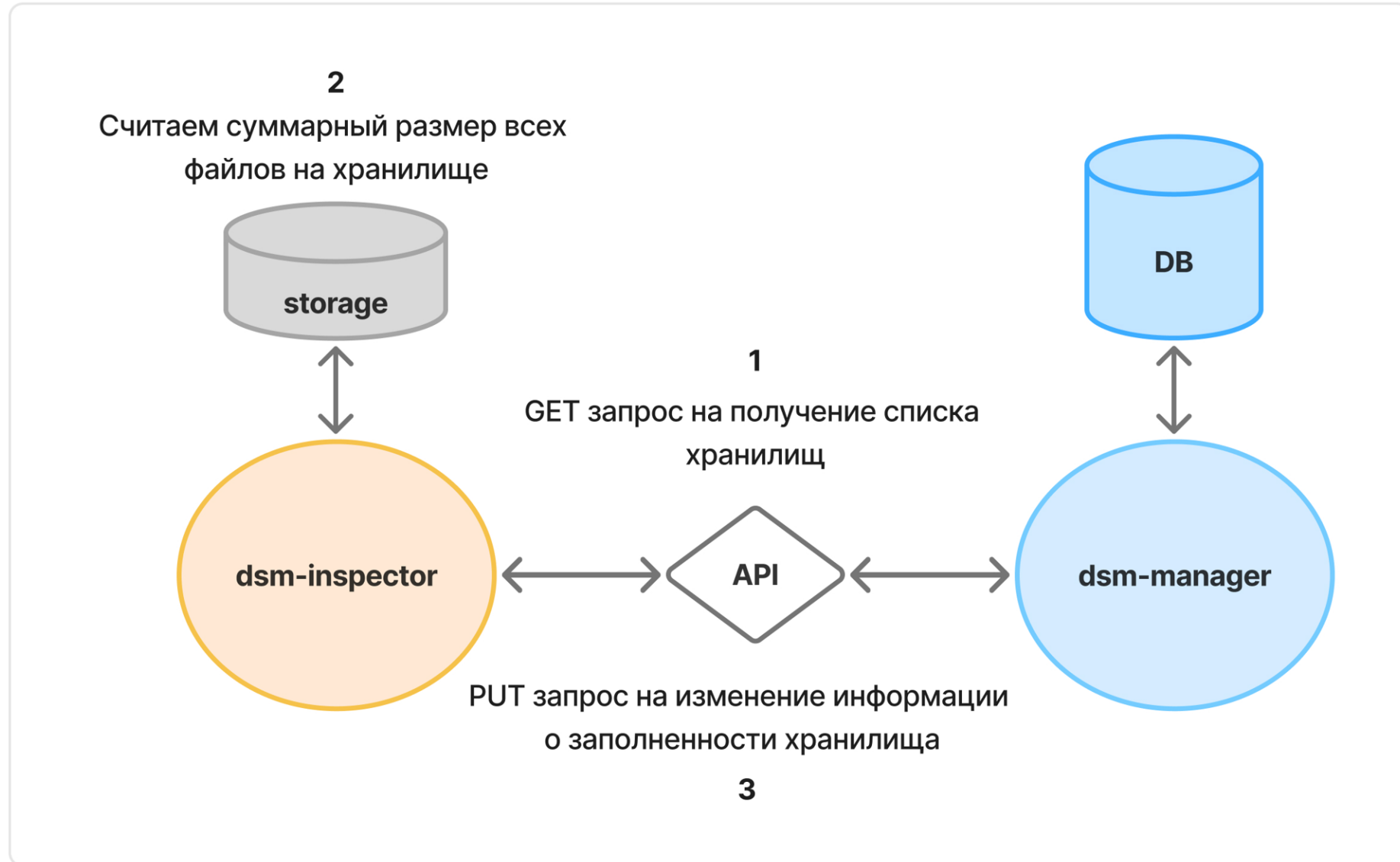
Удаление тёмных файлов



Контроль использования хранилища: мониторинг тёмных файлов



Контроль использования хранилища: мониторинг заполненности



Регистрация входных файлов (1/2)



▼ Publish message

Message will be published to the default exchange with routing key **dsm.register.file.input**, routing it to this queue.

Delivery mode: 1 - Non-persistent ▾

Headers: ? = String ▾

Properties: ? =

Payload:

```
{
  "meta": {
    "runNumber": "1"
  },
  "files": [
    {
      "name": "file_sample",
      "path": "/",
      "size": 100,
      "checksum": "e742438aa8bbf4d034408f07a654308d"
    }
  ]
}
```

Payload encoding: String (default) ▾

Publish message

(a) Сообщение о входном файле.

Responses

Curl

```
curl -X 'GET' \
  'http://localhost:8080/api/v1/file/?dataset_id=b21d78e9-8f4d-4649-96a2-887f20404311' \
  -H 'accept: application/json'
```

Request URL

```
http://localhost:8080/api/v1/file/?dataset_id=b21d78e9-8f4d-4649-96a2-887f20404311
```

Server response

Code	Details
200	Response body <pre>[{ "name": "file_sample", "path": "/", "storageId": "ca8b640d-b768-444f-b478-cdb50480104f", "size": 100, "checksum": "e742438aa8bbf4d034408f07a654308d", "statusCode": "CREATED", "id": "1929896f-089a-41d0-8fa7-8a9d6f13d553", "url": "/data/SPDOF-buffers/input/" }]</pre> Response headers <pre>content-length: 245 content-type: application/json date: Mon, 19 May 2025 14:58:23 GMT server: uvicorn</pre>

(b) Информация о файле в БД.

Регистрация входных файлов (2/2)

```
app-1 | 2025-05-19 14:54:42 INFO: [DSM-REGISTER-I004] [CONSUMER-HANDLER] Received message FilesInputDto(meta={'runNumber': '1'}, files=[FileInputDto(name='file_sample', path='/', size=100, check_sum='e742438aa8bbf4d034408f07a654308d')]) from queue '.dsm.register.file.input' with delivery tag 1 [in /src/app/executer/rabbit/consumers/base/consumer_handler.py:50]
app-1 | 2025-05-19 14:54:42 INFO: Processing msg [in /src/app/executer/services/processor/file_input_processor.py:42]
app-1 | 2025-05-19 14:54:42 INFO: Registered dataset ID=b21d78e9-8f4d-4649-96a2-887f20404311. [in /src/app/executer/services/processor/file_input_processor.py:56]
app-1 | 2025-05-19 14:54:42 INFO: Registered file ID=1929896f-089a-41d0-8fa7-8a9d6f13d553 in dataset ID=b21d78e9-8f4d-4649-96a2-887f20404311. [in /src/app/executer/services/processor/file_input_processor.py:70]
app-1 | 2025-05-19 14:54:43 INFO: Finish registering files. Dataset ID=b21d78e9-8f4d-4649-96a2-887f20404311 CLOSED! [in
```

(с) Лог о завершении регистрации входных файлов и закрытии созданного датасета.

Message 1

The server reported 0 messages remaining.

Exchange	dsm.register
Routing Key	dataset.input
Redelivered	0
Properties	
Payload	
109 bytes	
Encoding: string	{"id": "b21d78e9-8f4d-4649-96a2-887f20404311", "name": "input.test.ce86eb79-e300-40a5-9569-38a6cc33e488.raw"}

(d) Сообщение с информацией о созданном датасете.

Регистрация выходных файлов



▼ Publish message

Message will be published to the default exchange with routing key **dsm.register.file.process**, routing it to this queue.

Delivery mode:

1 - Non-persistent ▼

Headers: ?

=

String ▼

Properties: ?

=

Payload:

```
{  "datasetId": "28f16aec-f6dc-49f5-b052-624b0cf87973",  "files": [    {      "storageId" : "ca8b640d-b768-444f-b478-cdb50480104f",      "path" : "/process/test_file_process",      "size" : 100,      "checksum" : "e742438aa8bbf4d034408f07a654308d"    }  ]}
```

Payload encoding:

String (default) ▼

Publish message

Responses

Curl

```
curl -X 'GET' \
'http://localhost:8080/api/v1/file/?dataset_id=28f16aec-f6dc-49f5-b052-624b0cf87973' \
-H 'accept: application/json'
```

Request URL

```
http://localhost:8080/api/v1/file/?dataset_id=28f16aec-f6dc-49f5-b052-624b0cf87973
```

Server response

Code	Details
------	---------

200	
-----	--

Response body

```
[ {  "name": "test_file_process",  "path": "/process",  "storageId": "ca8b640d-b768-444f-b478-cdb50480104f",  "size": 100,  "checksum": "e742438aa8bbf4d034408f07a654308d",  "statusCode": "CREATED",  "id": "b7732b46-5096-498e-a52f-95e788a343b4",  "url": "/data/SPDOF-buffers/input/"}
```

Message 1

The server reported 0 messages remaining.

Exchange	dsm.register
Routing Key	file.process.reply
Redelivered	0
Properties	
Payload	
80 bytes	
Encoding: string	{"status": "SUCCESS", "details": "Registered file = /process/test_file_process"}

Приём заявок на удаление датасетов (1/2)



	id [PK] uuid	name character varying (255)	meta_data json	status_code character varying (20)
1	05d4c366-6d5a-468b-9d80-b21ae2ba0ce0	input.test.c40a08fe-4a75-4157-a04d-92946d5c0967.raw.output...	{"task_id": 19}	OPEN
2	0859c70b-2c4d-45c8-93e2-a2e804d4f53c	input.test.c40a08fe-4a75-4157-a04d-92946d5c0967.raw.log.1	{"task_id": 19}	OPEN
3	14cd5201-7238-4b86-adb6-79c03866fbcf	input.test.c1471f0f-4b55-45d8-9d66-97c42ff7adff.raw	{"run_number": 96, "files": 10}	CLOSED

(а) Информация о датасете в статусе CLOSED.

▼ Publish message

Message will be published to the default exchange with routing key **dsm.register.dataset.delete**, routing it to this queue.

Delivery mode:

1 - Non-persistent ▼

Headers: ?

=

String ▼

Properties: ?

=

Payload:

```
{
  "id": "14cd5201-7238-4b86-adb6-79c03866fbcf"
}
```

Payload encoding:

String (default) ▼

Publish message

(b) Заявка на удаление датасета в статусе CLOSED.

Приём заявок на удаление датасетов (2/2)



```
app-1 | 2025-01-18 16:26:06 INFO: [DSM-REGISTER-I004] [CONSUMER-HANDLER] Received message DatasetDeleteDto(id=UUID('14cd5201-7238-4b86-adb6-79c03866fbcf')) from queue '.dsm.register.dataset.delete' with delivery tag 56 [in /src/app/executer/rabbit/consumers/base/consumer_handler.py:50]
app-1 | 2025-01-18 16:26:06 INFO: Processing msg. [in /src/app/executer/services/processor/dataset_delete_processor.py:23]
app-1 | 2025-01-18 16:26:06 INFO: Dataset ID=14cd5201-7238-4b86-adb6-79c03866fbcf status changed to TO_DELETE. [in /src/app/executer/services/processor/dataset_delete_processor.py:35]
```

(с) Лог с информацией о получении сообщения и изменения статуса набора на TO_DELETE.

	id [PK] uuid	name character varying (255)	meta_data json	status_code character varying (20)
1	05d4c366-6d5a-468b-9d80-b21ae2ba0ce0	input.test.c40a08fe-4a75-4157-a04d-92946d5c0967.raw.output...	{"task_id": 19}	OPEN
2	0859c70b-2c4d-45c8-93e2-a2e804d4f53c	input.test.c40a08fe-4a75-4157-a04d-92946d5c0967.raw.log.1	{"task_id": 19}	OPEN
3	14cd5201-7238-4b86-adb6-79c03866fbcf	input.test.c1471f0f-4b55-45d8-9d66-97c42ff7adff.raw	{"run_number": 96, "files": 10}	TO_DELETE

(d) Проверка изменения статуса набора в базе данных.

Проверка целостности данных (1/2)



	id [PK] uuid	name character varying (255)	path character varying (255)	storage_id uuid	size integer
1	0fb1e1af-5c02-4d17-87a9-64dbf5e6a17	input.test.a976a020-3de5-44e2-91ee-319e426eda2f.raw	input_40	b3307ad4-f2b3-4f3a-a390-4f4e2762c620	50

(а) Корректная информация о файле в БД.

Curl

```
curl -X 'GET' \
'http://10.220.16.177:8080/api/v1/file/0fb1e1af-5c02-4d17-87a9-64dbf5e6a17' \
-H 'accept: application/json'
```

Request URL

http://10.220.16.177:8080/api/v1/file/0fb1e1af-5c02-4d17-87a9-64dbf5e6a17

Server response

Code

Details

200

Response body

```
{
  "name": "input.test.a976a020-3de5-44e2-91ee-319e426eda2f.raw",
  "path": "input_40",
  "storageId": "b3307ad4-f2b3-4f3a-a390-4f4e2762c620",
  "size": 50,
  "checksum": "c1349c048472b4cebd57669e1558b72a",
  "statusCode": "CREATED",
  "id": "0fb1e1af-5c02-4d17-87a9-64dbf5e6a17"
}
```

Download

(b) Полная информация о выбранном файле в БД.

	id [PK] uuid	name character varying (255)	path character varying (255)	storage_id uuid	size integer
1	0fb1e1af-5c02-4d17-87a9-64dbf5e6a17	input.test.a976a020-3de5-44e2-91ee-319e426eda2f.raw	input_40	b3307ad4-f2b3-4f3a-a390-4f4e2762c620	100

(с) Измененная информация (размер) о файле в БД.

Проверка целостности данных (2/2)

```
integrity-inspector-1 | 2025-01-19 13:31:47 INFO: File /data/SPDOF-buffers/input/input_40/input.test.a976a020-3de5-44e2-91ee-319e426eda2f.raw DAMAGED! [in /src/files_integrity_inspector/file_integrity_inspector.py:77]
integrity-inspector-1 | 2025-01-19 13:31:47 INFO: HTTP Request: PUT http://app:8080/api/v1/dataset/f61828be-64b5-44e8-9d18-a1a22068094d "HTTP/1.1 200 OK" [in /src/.venv/lib/python3.11/site-packages/httpx/_client.py:1038]
integrity-inspector-1 | 2025-01-19 13:31:47 INFO: Dataset ID=f61828be-64b5-44e8-9d18-a1a22068094d FROZEN [in /src/files_integrity_inspector/file_integrity_inspector.py:89]
```

Curl

```
curl -X 'GET' \
  'http://10.220.16.177:8080/api/v1/file/0fb1e1af-5c02-4d17-87a9-64defb5e6a17' \
  -H 'accept: application/json'
```

Request URL

<http://10.220.16.177:8080/api/v1/file/0fb1e1af-5c02-4d17-87a9-64defb5e6a17>

Server response

Code	Details
200	Response body <pre>{ "name": "input.test.a976a020-3de5-44e2-91ee-319e426eda2f.raw", "path": "input_40", "storageId": "b3307ad4-f2b3-4f3a-a390-4f4e2762c620", "size": 100, "checksum": "c1349c048472b4cebd57669e1558b72a", "statusCode": "DAMAGED", "id": "0fb1e1af-5c02-4d17-87a9-64defb5e6a17" }</pre> Download

(a) Информация о выбранном файле.

Curl

```
curl -X 'GET' \
  'http://10.220.16.177:8080/api/v1/dataset/?file_id=0fb1e1af-5c02-4d17-87a9-64defb5e6a17' \
  -H 'accept: application/json'
```

Request URL

http://10.220.16.177:8080/api/v1/dataset/?file_id=0fb1e1af-5c02-4d17-87a9-64defb5e6a17

Server response

Code	Details
200	Response body <pre>[{ "name": "input.test.b255570d-33bf-4dc3-9e6e-718df9a1a8ef.raw", "metaData": { "run_number": 49, "files": 10 }, "statusCode": "FROZEN", "id": "f61828be-64b5-44e8-9d18-a1a22068094d" }]</pre> Download

(b) Информация о соответствующем наборе.

Исполнение заявок на удаление датасетов (1/3)



Запрос История запросов

```
1 SELECT dat_file.id, dat_file.name, dat_file.status_code, dat_file.path FROM dat_dataset
2 JOIN dat_association_file_dataset fd on fd.dataset_id = dat_dataset.id
3 JOIN dat_file on fd.file_id = dat_file.id
4 WHERE dat_dataset.status_code = 'TO_DELETE'
```

Data Output Сообщения Notifications



	id [PK] uuid	name character varying (255)	status_code character varying (20)	path character varying (255)
1	c81794fe-c57c-485c-89c7-0ace02f81883	input.test.41dad14a-e58c-4cd0-979a-16ee3e1cdb03.r...	CREATED	input_4
2	aa20e90d-6892-4d1a-aad1-1f3c5a7e4dc2	input.test.08850bb9-ae56-460e-875e-486f166c7533.r...	CREATED	
3	b41ef691-3a5d-4d17-90d9-715a01e109a7	input.test.297eb0ea-c6e2-4261-acdf-b0d6f1c87482.ra...	CREATED	
4	baa79005-1115-4bef-a82c-3b342c2213f0	input.test.8e607ad1-c88a-4e95-ba61-b1d539282662.r...	CREATED	
5	6fa4d091-f983-4e21-92b8-ce68ea020732	input.test.d8a5c878-1ebe-4828-84b7-b37056bf6ac5.r...	CREATED	
6	4255668b-26f2-4be7-98d0-64d82832c3ac	input.test.faa4e27f-ae2a-49d8-8934-5bfaa3102709.raw	CREATED	
7	ec5ba1bf-b30b-47dc-8bc7-c76689bb58e1	input.test.8c04f7d8-2f26-4cc3-a81e-1b0dc796239e.ra...	CREATED	
8	1e87276a-0462-4964-803e-7b6f96ea018d	input.test.3ea2562e-aa73-4c5e-8911-576db19d7fc5.r...	CREATED	
9	8e238086-b493-459e-a567-627e00ab08e4	input.test.0aec3382-efe0-4986-8237-f07bc545248c.ra...	CREATED	
10	760ecdb0-6f41-4515-b880-c7c21890ed81	input.test.451393ef-5ecb-4ed8-afd4-454acecc92d6.raw	CREATED	

Запрос История запросов

```
1 SELECT dat_file.id, dat_file.name, dat_file.status_code, dat_file.path FROM dat_dataset
2 JOIN dat_association_file_dataset fd on fd.dataset_id = dat_dataset.id
3 JOIN dat_file on fd.file_id = dat_file.id
4 WHERE dat_dataset.status_code = 'TO_DELETE'
```

Data Output Сообщения Notifications



	id [PK] uuid	name character varying (255)	status_code character varying (20)	path character varying (255)
1	c81794fe-c57c-485c-89c7-0ace02f81883	input.test.41dad14a-e58c-4cd0-979a-16ee3e1cdb03.r...	TO_DELETE	input_4
2	aa20e90d-6892-4d1a-aad1-1f3c5a7e4dc2	input.test.08850bb9-ae56-460e-875e-486f166c7533.r...	TO_DELETE	input_4
3	b41ef691-3a5d-4d17-90d9-715a01e109a7	input.test.297eb0ea-c6e2-4261-acdf-b0d6f1c87482.ra...	TO_DELETE	input_4
4	baa79005-1115-4bef-a82c-3b342c2213f0	input.test.8e607ad1-c88a-4e95-ba61-b1d539282662.r...	TO_DELETE	input_4
5	6fa4d091-f983-4e21-92b8-ce68ea020732	input.test.d8a5c878-1ebe-4828-84b7-b37056bf6ac5.r...	TO_DELETE	input_4
6	4255668b-26f2-4be7-98d0-64d82832c3ac	input.test.faa4e27f-ae2a-49d8-8934-5bfaa3102709.raw	TO_DELETE	input_4
7	ec5ba1bf-b30b-47dc-8bc7-c76689bb58e1	input.test.8c04f7d8-2f26-4cc3-a81e-1b0dc796239e.ra...	TO_DELETE	input_4
8	1e87276a-0462-4964-803e-7b6f96ea018d	input.test.3ea2562e-aa73-4c5e-8911-576db19d7fc5.r...	TO_DELETE	input_4
9	8e238086-b493-459e-a567-627e00ab08e4	input.test.0aec3382-efe0-4986-8237-f07bc545248c.ra...	TO_DELETE	input_4
10	760ecdb0-6f41-4515-b880-c7c21890ed81	input.test.451393ef-5ecb-4ed8-afd4-454acecc92d6.raw	TO_DELETE	input_4

Исполнение заявок на удаление датасетов (2/3)



```
deleting-inspector-1 | 2025-01-19 14:17:13 INFO: Foarm file list with status TO_DELETE [in /src/files_deleting_inspector/file_delete_inspector.py:32]
deleting-inspector-1 | 2025-01-19 14:17:13 INFO: HTTP Request: GET http://app:8080/api/v1/file/?file_status=TO_DELETE "HTTP/1.1 200 OK" [in /src/.venv/lib/python3.11/site-packages/httpx/_client.py:1038]
deleting-inspector-1 | 2025-01-19 14:17:13 INFO: HTTP Request: GET http://app:8080/api/v1/storage/b3307ad4-f2b3-4f3a-a390-4f4e2762c620 "HTTP/1.1 200 OK" [in /src/.venv/lib/python3.11/site-packages/httpx/_client.py:1038]
deleting-inspector-1 | 2025-01-19 14:17:13 INFO: File /data/SPD0F-buffers/input/input_4/input.test.8c04f7d8-2f26-4cc3-a81e-1b0dc796239e.raw DELETED! [in /src/files_deleting_inspector/file_delete_inspector.py:44]
```

(a) Лог с информацией об удалении файлов, находящихся в статусе TO_DELETE.

Запрос История запросов

```
1 SELECT dat_file.id, dat_file.name, dat_file.status_code, dat_file.path FROM dat_dataset
2 JOIN dat_association_file_dataset fd on fd.dataset_id = dat_dataset.id
3 JOIN dat_file on fd.file_id = dat_file.id
4 WHERE dat_dataset.status_code = 'TO_DELETE'
```

Data Output Сообщения Notifications

	id [PK] uuid	name character varying (255)	status_code character varying (20)	path character varying (255)
1	c81794fe-c57c-485c-89c7-0ace02f81883	input.test.41dad14a-e58c-4cd0-979a-16ee3e1cdb03.r...	DELETED	input_4
2	aa20e90d-6892-4d1a-aad1-1f3c5a7e4dc2	input.test.08850bb9-ae56-460e-875e-486f166c7533.r...	DELETED	input_4
3	b41ef691-3a5d-4d17-90d9-715a01e109a7	input.test.297eb0ea-c6e2-4261-acdf-b0d6f1c87482.ra...	DELETED	input_4
4	baa79005-1115-4bef-a82c-3b342c2213f0	input.test.8e607ad1-c88a-4e95-ba61-b1d539282662.r...	DELETED	input_4
5	6fa4d091-f983-4e21-92b8-ce68ea020732	input.test.d8a5c878-1ebe-4828-84b7-b37056bf6ac5.r...	DELETED	input_4
6	4255668b-26f2-4be7-98d0-64d82832c3ac	input.test.faa4e27f-ae2a-49d8-8934-5bfaa3102709.raw	DELETED	input_4
7	ec5ba1bf-b30b-47dc-8bc7-c76689bb58e1	input.test.8c04f7d8-2f26-4cc3-a81e-1b0dc796239e.ra...	DELETED	input_4
8	1e87276a-0462-4964-803e-7b6f96ea018d	input.test.3ea2562e-aa73-4c5e-8911-576db19d7fc5.r...	DELETED	input_4
9	8e238086-b493-459e-a567-627e00ab08e4	input.test.0aec3382-efe0-4986-8237-f07bc545248c.ra...	DELETED	input_4
10	760ecdb0-6f41-4515-b880-c7c21890ed81	input.test.451393ef-5ecb-4ed8-afd4-454acecc92d6.raw	DELETED	input_4

(b) Измененная информация об удаленных файлах.

Исполнение заявок на удаление датасетов (3/3)



	id [PK] uuid	name character varying (255)	meta_data json	status_code character varying (20)
1	bf04dfe1-db61-4195-b584-45de6ea8f04d	input.test.d95edab2-8f40-4b67-9fd7-01565512228f.raw	{"run_number": 3, "files": 10}	TO_DELETE

(a) Информация о наборе, к которому привязаны удаленные файлы.

	id [PK] uuid	name character varying (255)	meta_data json	status_code character varying (20)
1	bf04dfe1-db61-4195-b584-45de6ea8f04d	input.test.d95edab2-8f40-4b67-9fd7-01565512228f.raw	{"run_number": 3, "files": 10}	DELETED

(b) Обновленная информация о наборе.

Мониторинг хранилища на предмет тёмных файлов



```
monitoring-inspector-1 | 2025-05-19 22:45:42 INFO: HTTP Request: GET http://app:8080/api/v1/storage/ "HTTP/1.1 200 OK" [in /src/.venv/lib/python3.11/site-packages/httpx/_client.py:1038]
monitoring-inspector-1 | 2025-05-19 22:45:42 INFO: HTTP Request: GET http://app:8080/api/v1/file/file_name/test_dark "HTTP/1.1 404 Not Found" [in /src/.venv/lib/python3.11/site-packages/httpx/_client.py:1038]
monitoring-inspector-1 | 2025-05-19 22:45:42 INFO: HTTP Request: GET http://app:8080/api/v1/dark_file/test_dark "HTTP/1.1 404 Not Found" [in /src/.venv/lib/python3.11/site-packages/httpx/_client.py:1038]
monitoring-inspector-1 | 2025-05-19 22:45:43 INFO: HTTP Request: POST http://app:8080/api/v1/dark_file/ "HTTP/1.1 201 Created" [in /src/.venv/lib/python3.11/site-packages/httpx/_client.py:1038]
monitoring-inspector-1 | 2025-05-19 22:45:43 INFO: File test_dark add into table for dark files. [in /src/monitoring_inspector/i
```

(а) Лог с информацией об обнаружении «тёмного» файла.

Responses

Curl

```
curl -X 'GET' \
'http://localhost:8080/api/v1/dark_file/' \
-H 'accept: application/json'
```

Request URL

```
http://localhost:8080/api/v1/dark_file/
```

Server response

Code	Details
200	Response body <pre>[{ "name": "test_dark", "path": "input", "storageId": "ca8b640d-b768-444f-b478-cdb50480104f", "size": 135, "createdDate": "2025-05-19T22:45:23.155309+00:00", "url": "/data/SPDOF-buffers/input/" }]</pre> Download

(b) Информация о «тёмном» файле в БД.

Мониторинг заполненности хранилища



Responses

Curl

```
curl -X 'GET' \
  'http://localhost:8080/api/v1/storage/type/input' \
  -H 'accept: application/json'
```

Request URL

```
http://localhost:8080/api/v1/storage/type/input
```

Server response

Code	Details
200	Response body <pre>{ "url": "/data/SPDOF-buffers/input", "path": "/", "totalSpaceSize": 10000000, "usedSpaceSize": 135, "type": "input", "id": "ca8b640d-b768-444f-b478-cdb50480104f" }</pre> Download

(а) Информация о входном хранилище.

```
monitoring-inspector-1 | 2025-05-20 11:54:10 INFO: HTTP Request: GET http://app:8080/api/v1/storage/ "HTTP/1.1 200 OK" [in /src/.venv/lib/python3.11/site-packages/httpx/_client.py:1038]
monitoring-inspector-1 | 2025-05-20 11:54:10 INFO: HTTP Request: PUT http://app:8080/api/v1/storage/ca8b640d-b768-444f-b478-cdb50480104f "HTTP/1.1 200 OK" [in /src/.venv/lib/python3.11/site-packages/httpx/_client.py:1038]
monitoring-inspector-1 | 2025-05-20 11:54:10 INFO: input storage uses 370 bytes. [in /src/monitoring_inspector/inspectors/storag
```

(б) Лог с информацией об обновлении данных о хранилище.

Curl

```
curl -X 'GET' \
  'http://localhost:8080/api/v1/storage/type/input' \
  -H 'accept: application/json'
```

Request URL

```
http://localhost:8080/api/v1/storage/type/input
```

Server response

Code	Details
200	Response body <pre>{ "url": "/data/SPDOF-buffers/input", "path": "/", "totalSpaceSize": 10000000, "usedSpaceSize": 370, "type": "input", "id": "ca8b640d-b768-444f-b478-cdb50480104f" }</pre>