



# СИСТЕМА УПРАВЛЕНИЯ ПРОЦЕССАМИ ОБРАБОТКИ В SPD ONLINE FILTER

Студент

Научный руководитель

к.ф.-м.н.

Научный консультант

к.т.н

А. В. Плотников

Е. Ю. Солдатов

Д. А. Олейник

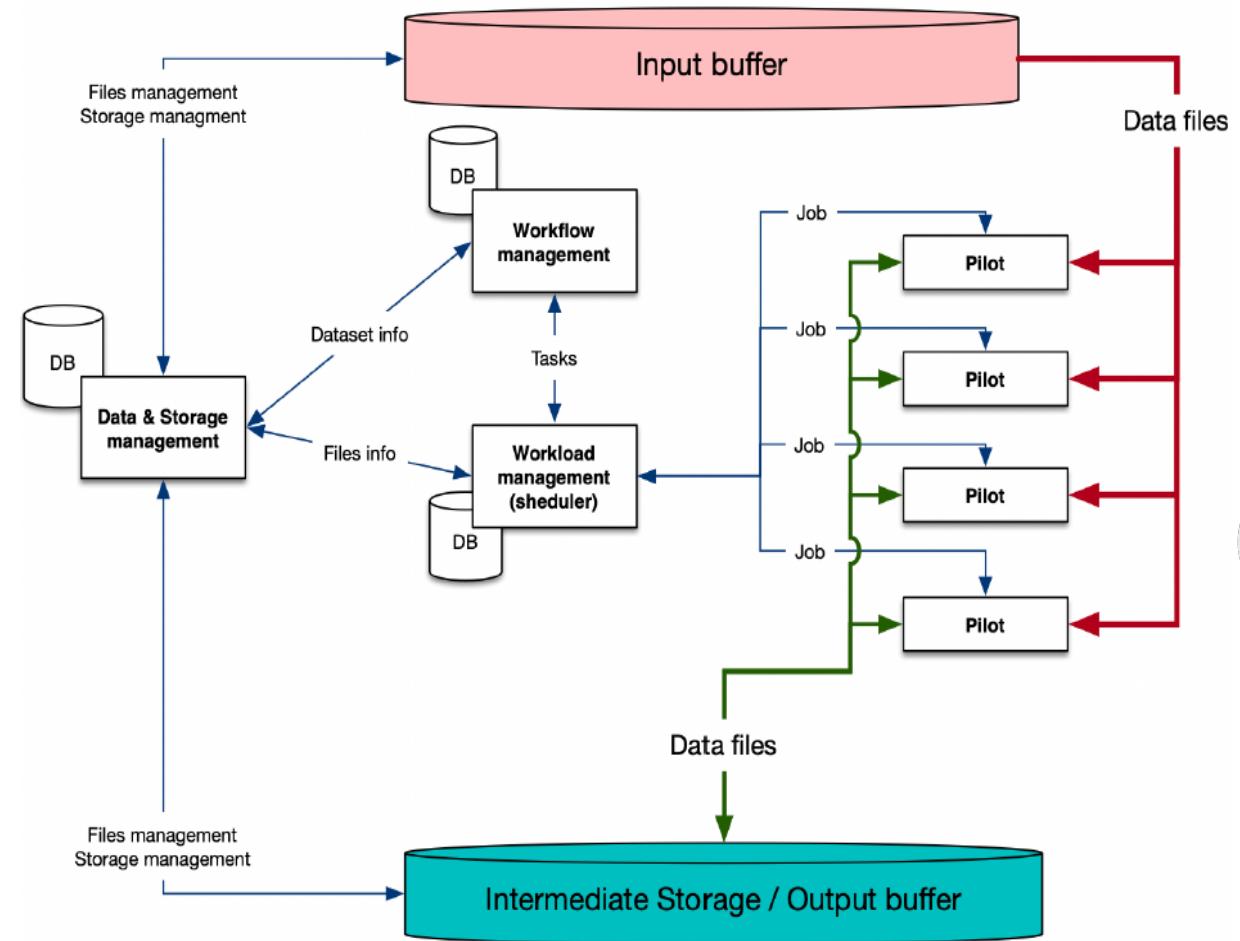


# SPD Online Filter

**SPD Online Filter** — специализированная программно-аппаратная система, предназначенная для предварительной обработки данных эксперимента SPD.

Система реализует многоступенчатую, высокопропускную методику обработки.

Основная, но не единственная, **цель обработки**: существенно уменьшить объём данных для последующего анализа и долговременного хранения.

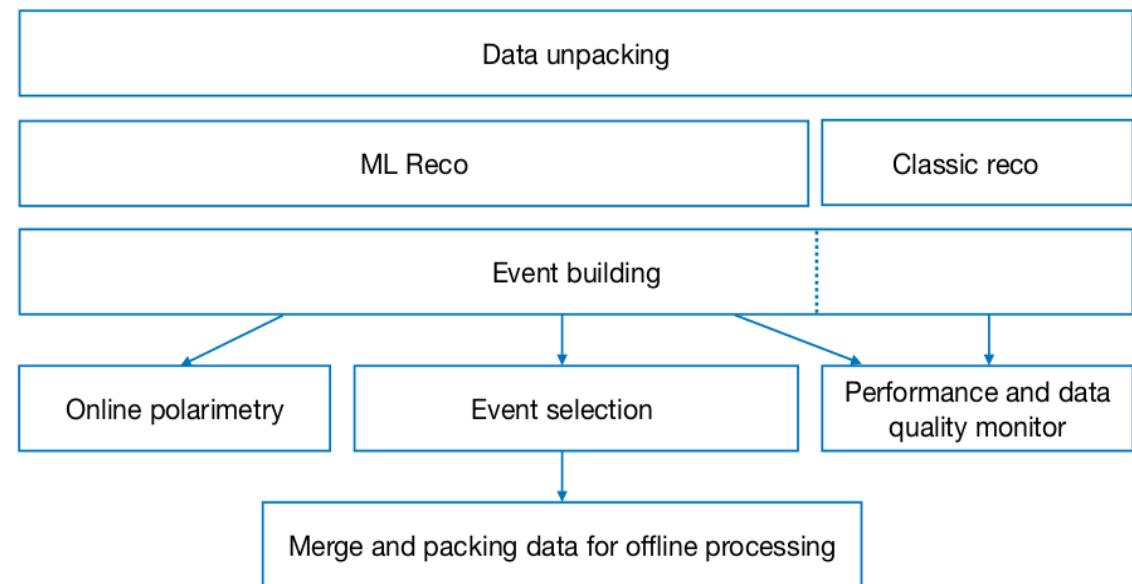


Архитектура Online Filter

# Управление процессами обработки данных

**Многоступенчатая обработка** — набор последовательных этапов обработки данных. При этом на каждом этапе может обрабатываться достаточно большой объем данных.

Каждый шаг, кроме первого, принимает данные, обработанные на предыдущем этапе, и передает результаты следующему, обеспечивая преобразование информации из одного представления в другое.



Пример последовательных шагов обработки

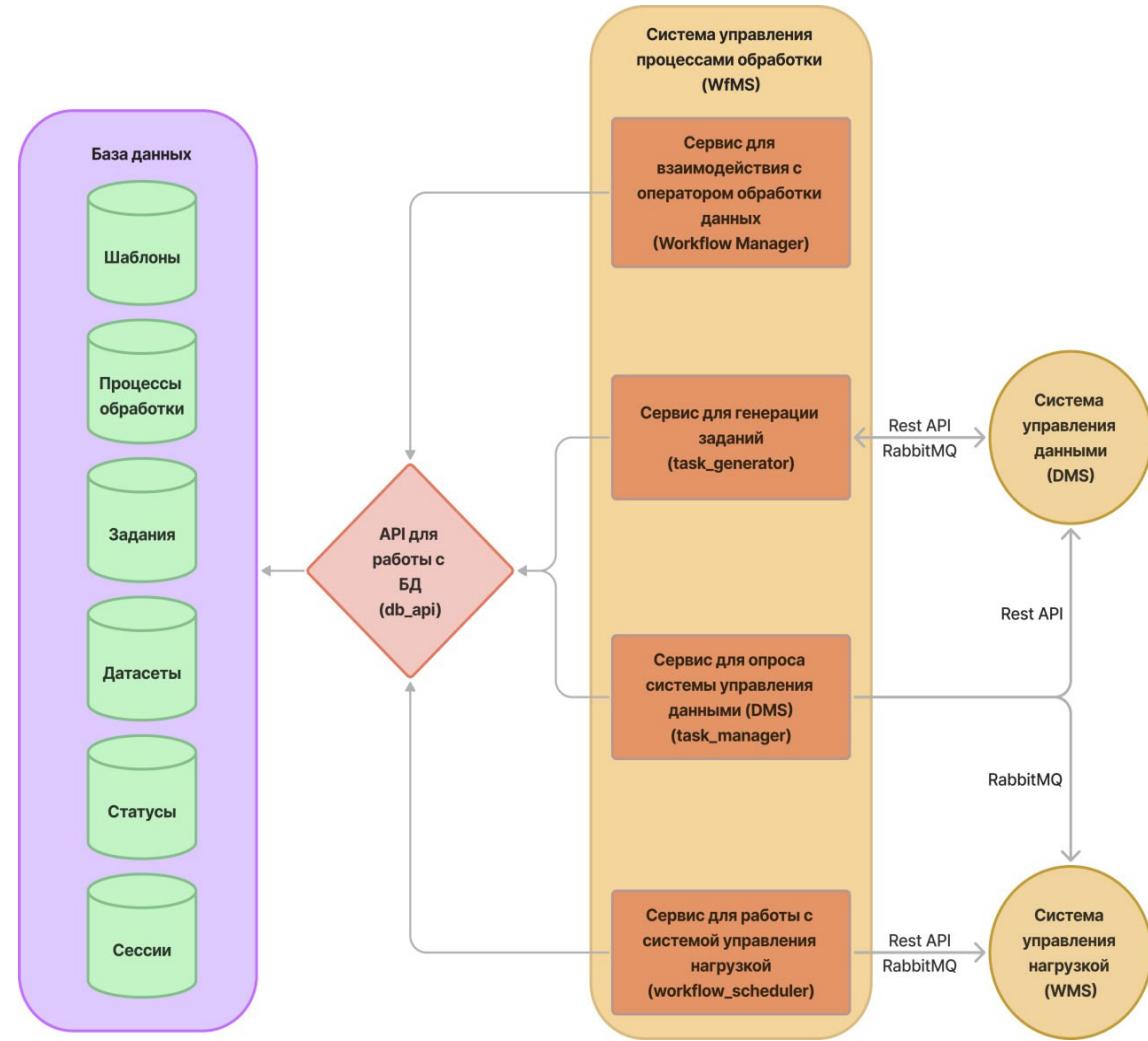
# Система управления процессами обработки (WfMS)

## Основная цель системы:

параллельное и наиболее результативное управление большим количеством процессов обработки данных и контроль статусов выполнения цепочек обработки для физического эксперимента SPD.

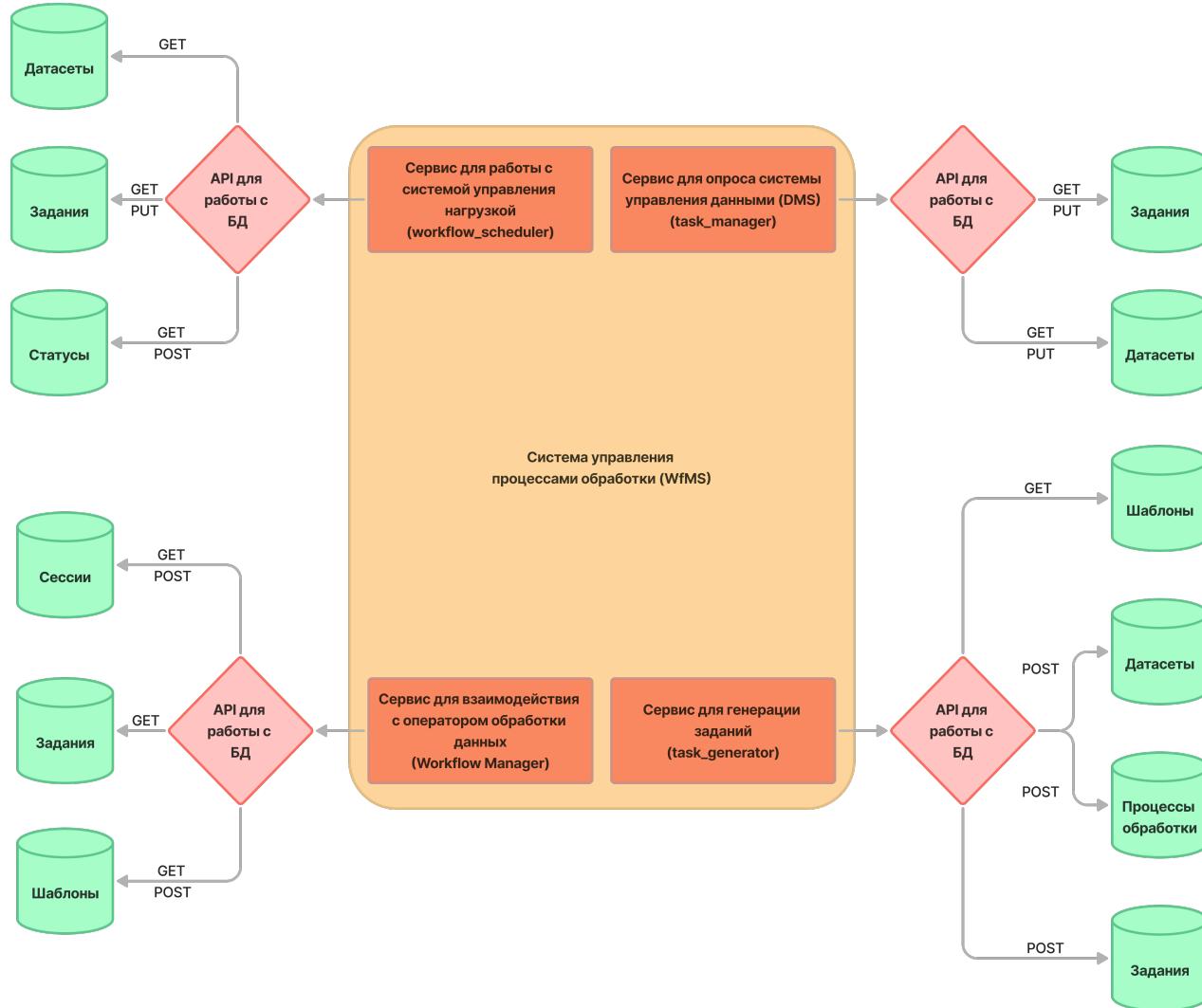
## Основные задачи системы:

- Создание формального описания процесса;
- Контроль за выполнением обработки.



Архитектура WfMS

# Взаимодействия микросервисов WfMS с БД



- База данных PostgreSQL развернута на отдельной виртуальной машине.
- Взаимодействие с базой данных осуществляется через REST API.

SQLAlchemy ORM

+

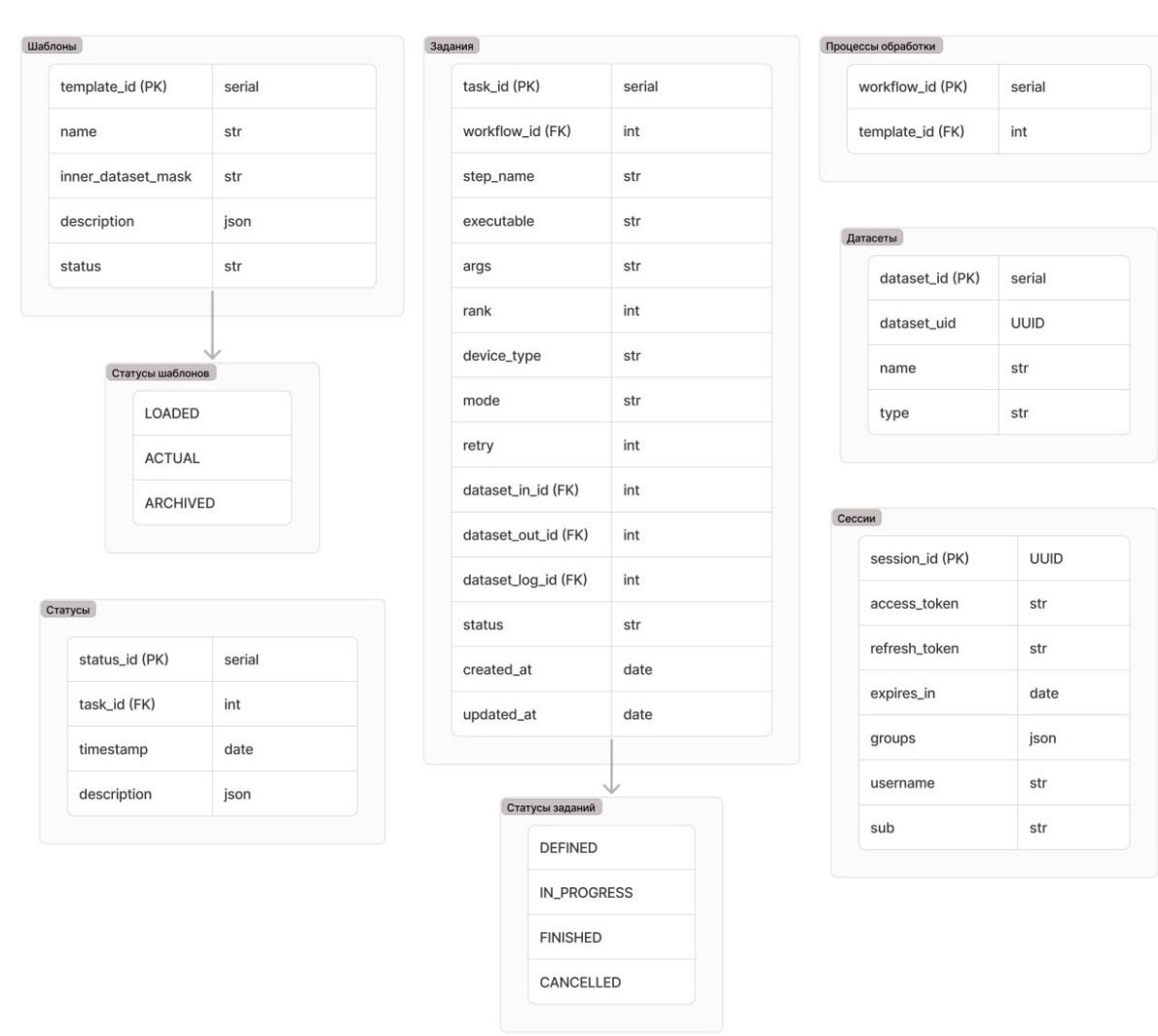
асинхронные сессии и asyncpg

+

миграции Alembic

Схема взаимодействия микросервисов с БД

# Структура базы данных и API к ней



Templates	
GET	/template/all Get All Templates
GET	/template/actual Get Actual Templates
GET	/template/{specific_id} Template Response
POST	/template/create Create Template
PUT	/template/{specific_id}/status Change Template
DELETE	/template/{specific_id}/delete Delete Template

Tasks	
GET	/task/all Get All Tasks
GET	/task/joined Get Joined Tasks
GET	/task/{specific_id} Task Response
GET	/task/status/{status_name} Get Defined Tasks
POST	/task/create Create Task
PUT	/task/{specific_id}/rank Change Rank
PUT	/task/{specific_id}/status Change Status

Datasets	
GET	/dataset/all Get All Datasets
GET	/dataset/{specific_id} Dataset Response
POST	/dataset/create Create Dataset
PUT	/dataset/{specific_id}/uid Change Rank

Workflows	
GET	/workflow/all Get All Workflows
GET	/workflow/{specific_id} Workflow Response
POST	/workflow/create Create Workflow

UserSessions	
GET	/session/{specific_id} Session Response
POST	/session/create Create Session
PUT	/session/{specific_id}/token Update Session
GET	/session/sub/{sub} Get User Session By Sub
DELETE	/session/{specific_id}/delete Delete Session

# Сервис для взаимодействия с оператором обработки данных

Основной функционал сервиса

1. Авторизация пользователей с помощью SPD-IAM;
2. Вывод CWL-шаблонов;
3. Вывод заданий;
4. Создание CWL-шаблонов суперпользователями (ввод в специальное поле / загрузка из файла / клонирование существующих шаблонов);
5. Предварительная валидация и запись в БД CWL-шаблонов;
6. Изменение статусов шаблонов суперпользователями на "ACTUAL" и "ARCHIVED";
7. Удаление суперпользователями шаблонов в статусе "LOADED".

## Workflow Manager

Templates Tasks

admin@jinr.ru Logout

id	wflow_id	step	template	exec	args	priority	type	mode	retry	in_ds_name	out_ds_name	log_ds_name	status
2	1	reconstruction	<a href="#">Decoding &amp;Reco</a>	processing program	cable_map	1	CPU	map	5	input.test.4b5 f78b1-2412-4 058-9a7e- f9b09012ec9 d.raw.output.	input.test.4b5 f78b1-2412-4 058-9a7e- f9b09012ec9 d.raw.output.	input.test.4b5 f78b1-2412-4 058-9a7e- f9b09012ec9 d.raw.log.2	DEFINED
1	1	decoding	<a href="#">Decoding &amp;Reco</a>	processing program	cable_map	1	CPU	map	5	input.test.4b5 f78b1-2412-4 058-9a7e- f9b09012ec9 d.raw	input.test.4b5 f78b1-2412-4 058-9a7e- f9b09012ec9 d.raw.output.	input.test.4b5 f78b1-2412-4 058-9a7e- f9b09012ec9 d.raw.log.1	IN_PROGRESS

- Пользовательский интерфейс
- Backend: FastAPI
- Шаблонизатор: Jinja2
- Frontend: Bootstrap (HTML + CSS + JS)

## Workflow Manager

Templates Tasks

admin@jinr.ru Logout

template_id	name	input_dataset_mask	status
2	Decoding&Reco	.test.	ACTUAL
4	Data_Decoding_Clone	.test.	ARCHIVED
1	Data_Decoding	.test.	ARCHIVED
5	RAW data decoding	RAW2024	LOADED

# Создание шаблонов

- Создание шаблонов доступно только суперпользователям (в частности, отсутствует кнопка в интерфейсе)
- Предварительная валидация шаблонов с помощью инструментов cwltools
- Сохранение шаблона в БД

Workflow Manager

Templates Tasks [admin@jinr.ru](#) [Logout](#)

### CWL Template

Template name

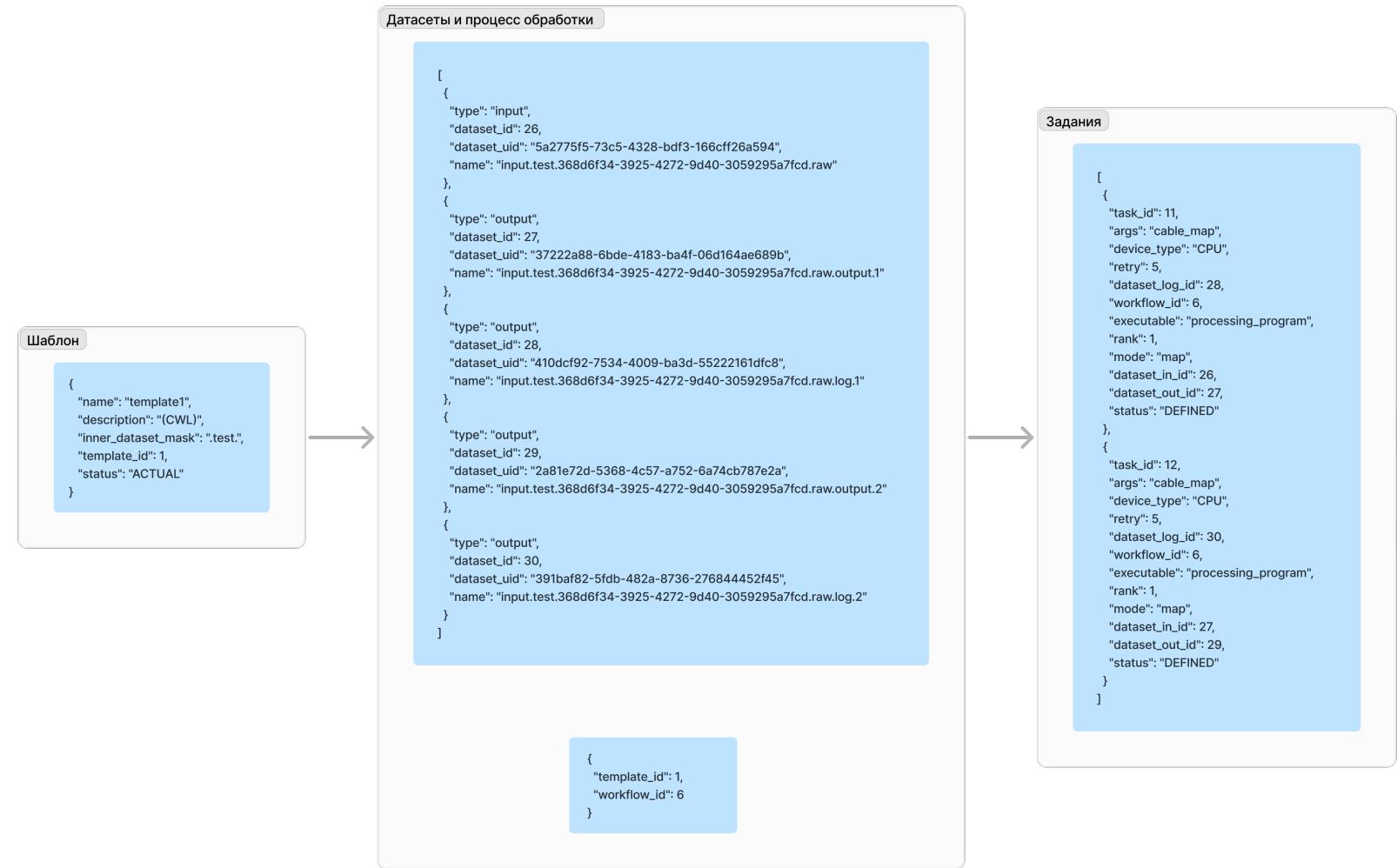
Inner dataset mask

Enter CWL here...

Complete

# Сервис для генерации заданий

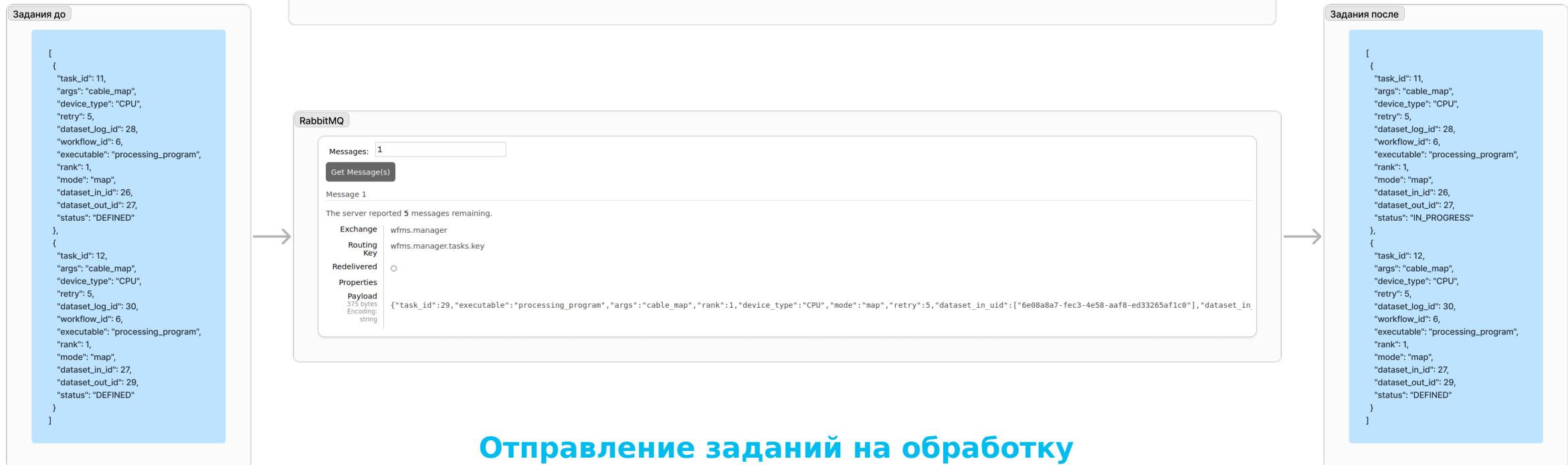
1. Получение зарегистрированных датасетов от DMS из RabbitMQ;
2. Сопоставление датасета по маске имени с нужным шаблоном;
3. Регистрация входного датасета в системе;
4. Создание процесса обработки по шаблону;
5. Создание выходного датасета и датасета логов в системе;
6. Создание заданий.



От шаблона к заданиям

# Сервис для опроса системы управления данными (DMS)

1. Итерация по заданиям в статусе “DEFINED”;
2. Опрос DMS о статусе входного датасета (“CLOSED”);
3. Создание выходных датасетов и датасетов логов в DMS;
4. Отправление задания в RabbitMQ для последующей обработки в WMS;
5. Смена статуса задания на “IN\_PROGRESS”.



# Сервис для работы с системой управления нагрузкой (WMS)

## Основной функционал сервиса

1. Завершение заданий с закрытием обработанных датасетов в DMS;
2. Отмена заданий на основе данных о заполненности выходного хранилища с оповещением WMS;
3. Изменение приоритетов заданий в процессе их выполнения.

## 1 этап

1. Итерация по заданиям в статусе “IN\_PROGRESS”;
2. Опрос WMS о состоянии задания;
3. Запись состояния;
4. Для успешно завершенных заданий:
  - а) Смена статуса выходных датасетов на CLOSED в DMS;
  - б) Смена статуса задания на “FINISHED”;
5. Удаление входного и промежуточных датасетов.

# Логирование, контейнеризация и оркестрация

- Логирование: custom logger

```
task_generator 2024-09-24 20:59:44,380 INFO Logger started
task_generator 2024-09-24 20:59:44,665 INFO Templates successfully initialised.
```

docker-compose

```
services:
  db_api:
    ...
  task_generator:
    ...
  task_manager:
    ...
  workflow_manager
    ...
  workflow_scheduler
    ...
  ...
networks:
  wfms:
```

- Контейнеризация и оркестрация:  
Docker и Docker Compose

# Результаты и дальнейшие планы

## Результаты

- 1) разработан API для работы с базой данных, открывающий доступ для сервисов WfMS к определенному набору функций;
- 2) создан сервис для работы с оператором обработки данных, позволяющий просматривать задания и шаблоны, а также предоставляющий возможность суперпользователям создавать шаблоны и управлять их статусами, произведена интеграция с SPD-IAM;
- 3) разработан сервис для генерации заданий, сопоставляющий датасеты с шаблонами по маске имени и создающий по ним цепочки обработки и задания;
- 4) создан сервис для опроса DMS, опрашивающий его о готовности входного датасета для каждого готового задания, создающий выходные датасеты для таких заданий и отправляющий задания на обработку в WMS;
- 5) произведена контейнеризация всех приложений и настроена их оркестрация с помощью docker-compose;
- 6) внедрен logger, сохраняющий информацию об актуальном состоянии системы в каждый момент времени.

## Дальнейшие планы

- 1) Разработка сервиса для взаимодействия с WMS;
- 2) Переход к полной асинхронности;
- 3) Тестирование системы.



**Спасибо за  
внимание!**

# Дополнительные слайды

Планы на сервис для генерации заданий и сервис для опроса DMS

## Перейти:

1. к полной асинхронности;
2. от скрипта к полноценному демону.



Планы на сервис для взаимодействия с оператором обработки данных

## Добавить:

1. возможность изменения статусов шаблонов суперпользователями;
2. возможность изменения шаблонов в статусе “LOADED”;
3. загрузка шаблона из файла;
4. возможность работать с аккаунтом пользователя. → интеграция с SPD-IAM



Приступить к созданию сервиса для работы с системой управления нагрузкой (WMS).