



СИСТЕМА УПРАВЛЕНИЯ ПРОЦЕССАМИ ОБРАБОТКИ ДЛЯ СПЕЦИАЛИЗИРОВАННОГО ВЫЧИСЛИТЕЛЬНОГО КОМПЛЕКСА SPD ONLINE FILTER

Студент

Научный консультант

А. В. Плотников

Д. А. Олейник

Москва 2024

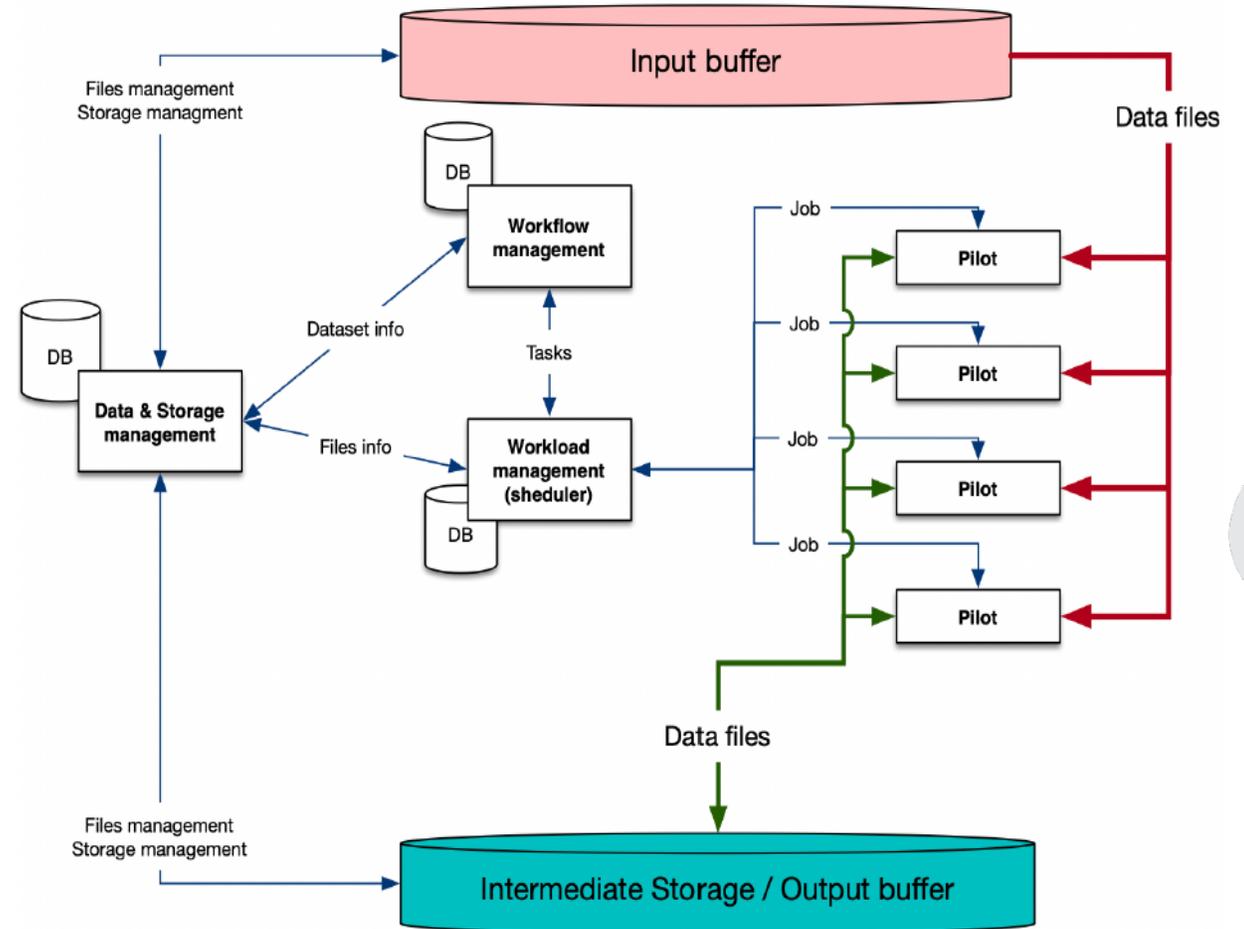


SPD Online Filter

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.

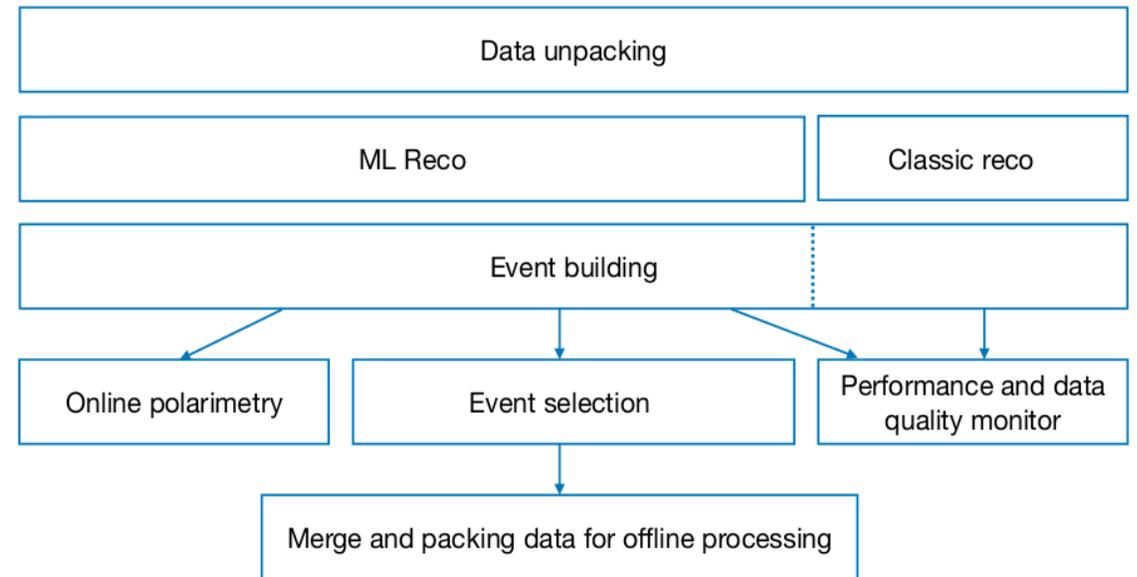


Архитектура Online Filter

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

Обоснование необходимости системы

Система должна одновременно и быстро обрабатывать большой объём данных



Для каждого из шагов обработки нужен свой процесс



При этом каждый из шагов образует промежуточные данные, которые нужно перенаправлять на дальнейшую обработку



Нужна система управления процессами обработки

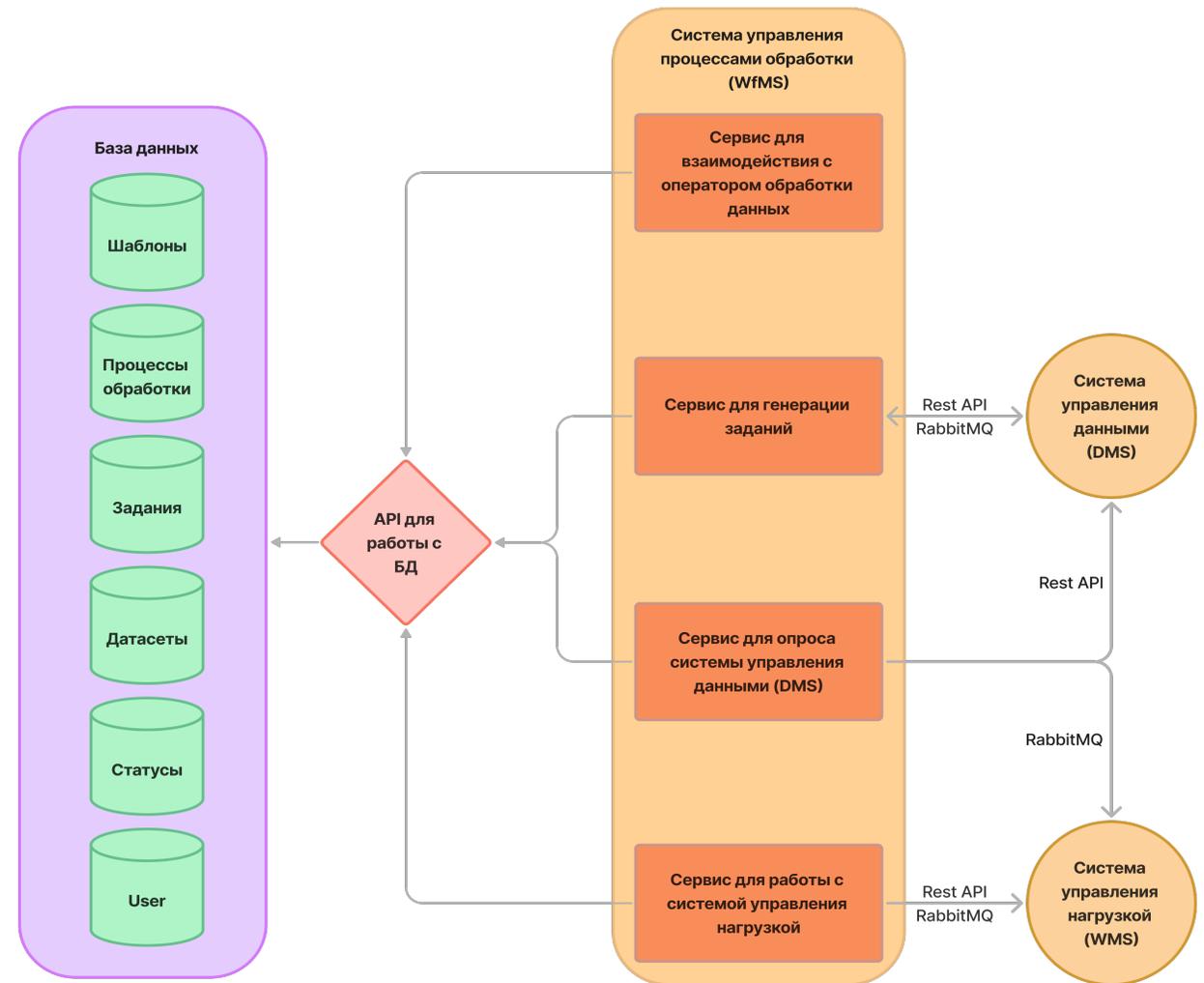
Система управления процессами обработки (WfMS)

Основная цель системы:

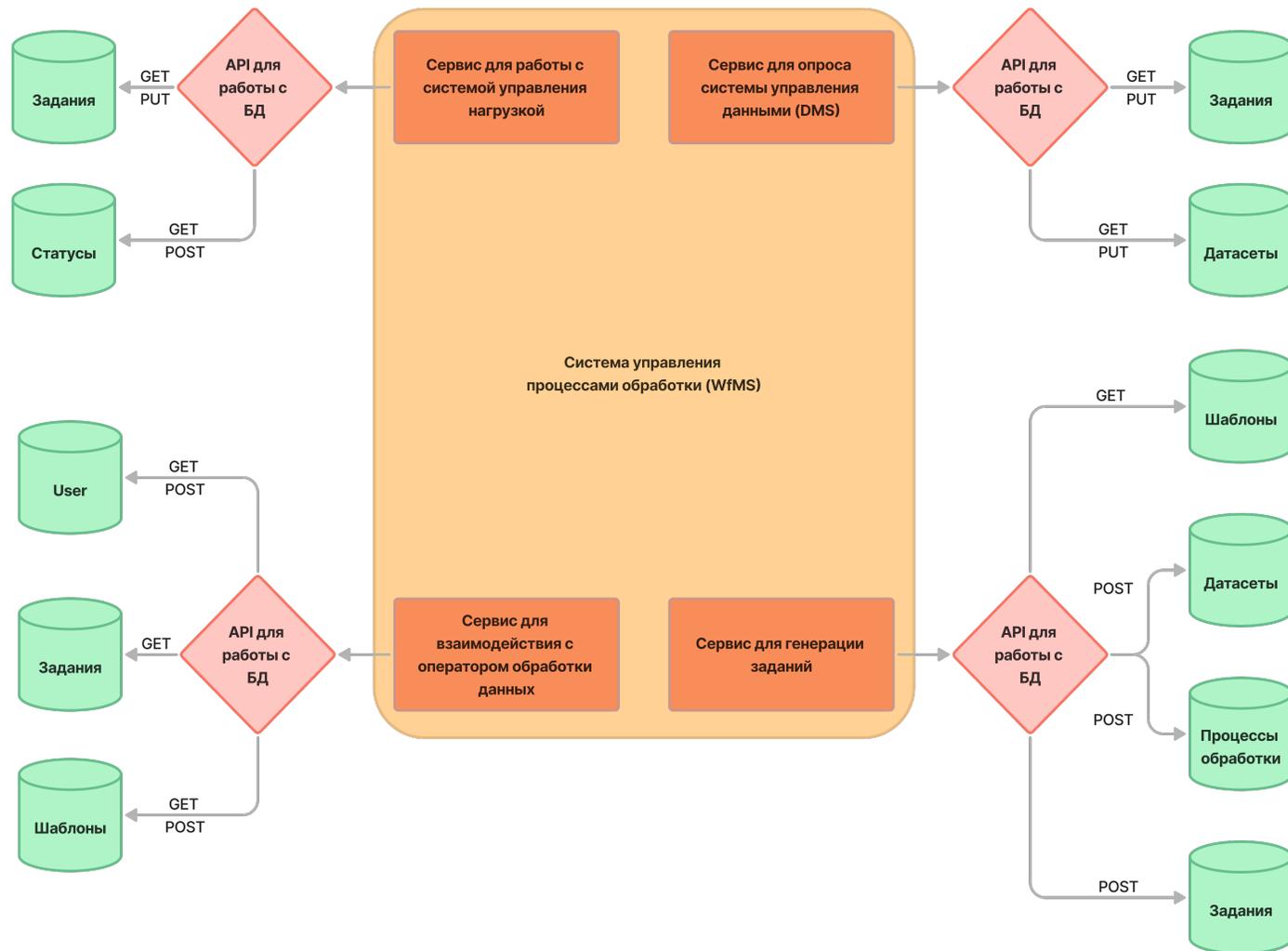
параллельное и наиболее результативное управление большим количеством процессов обработки данных и контроль статусов выполнения цепочек обработки для физического эксперимента SPD.

Основные задачи системы:

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



Взаимодействия микросервисов WfMS с БД

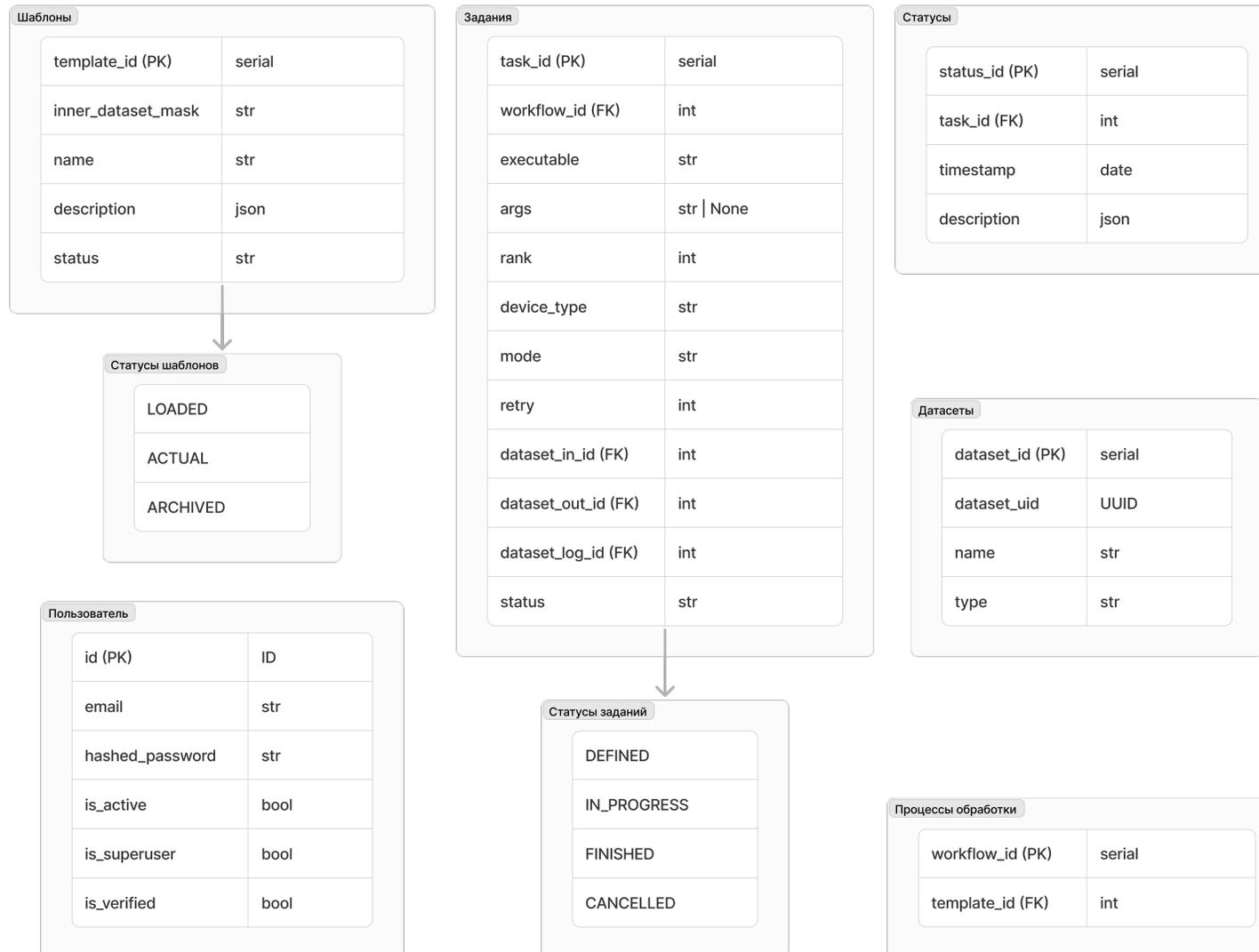


- База данных PostgreSQL развернута на отдельной виртуальной машине.
- Взаимодействие с базой данных осуществляется через REST API.

SQLAlchemy ORM
+
асинхронные сессии и asynprg
+
миграции Alembic

Схема взаимодействия микросервисов с БД

Структура базы данных



API к БД

Templates



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

Datasets



GET	/dataset/all	Get All Datasets	▼
GET	/dataset/{dataset_id}	Dataset Response	▼
POST	/dataset/create	Create Dataset	▼
PUT	/dataset/{dataset_id}/dataset_uid	Change Rank	▼

Workflows



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

Tasks



GET	/task/all	Get All Tasks	▼
GET	/task/{task_id}	Task Response	▼
GET	/task/status/{status_name}	Get Defined Tasks	▼
POST	/task/create	Create Task	▼
PUT	/task/{task_id}/rank	Change Rank	▼
PUT	/task/{task_id}/status	Change Rank	▼
DELETE	/task/{task_id}/delete	Delete Task	▼

API к БД: пользователь

auth		^
POST	/auth/login Auth:Jwt.Login	✓
POST	/auth/logout Auth:Jwt.Logout	🔒 ✓
POST	/auth/register Register:Register	✓
POST	/auth/forgot-password Reset:Forgot Password	✓
POST	/auth/reset-password Reset:Reset Password	✓
POST	/auth/request-verify-token Verify:Request-Token	✓
POST	/auth/verify Verify:Verify	✓
users		^
GET	/users/me Users:Current User	🔒 ✓
PATCH	/users/me Users:Patch Current User	🔒 ✓
GET	/users/{id} Users:User	🔒 ✓
PATCH	/users/{id} Users:Patch User	🔒 ✓
DELETE	/users/{id} Users>Delete User	🔒 ✓

- FastAPI
- Документация и проверка API — Swagger

- Взаимодействие с пользователем — FastAPI Users
- CORS на сервис для взаимодействия с оператором обработки данных

Сервис для взаимодействия с оператором обработки данных

Основной функционал сервиса

1. Регистрация и авторизация пользователей с разными правами;
2. Вывод CWL-шаблонов;
3. Вывод заданий;
4. Создание CWL-шаблонов суперпользователями;
5. Предварительная валидация и запись в БД CWL-шаблонов;
6. Изменение статусов шаблонов суперпользователями.

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

Template Manager

[Login](#)[Registration](#)

Creation of CWL templates

Create your own CWL here!

[Learn more](#)

© 2024 WfMS

Регистрация и авторизация

Template Manager

Login Registration

Login

 Remember me

[Forgot password?](#) or [Registration](#)

© 2024 WfMS

Template Manager

Login Registration

Registration

© 2024 WfMS

- FastAPI Users
- JWT-token сохраняется в cookie браузера
- Поля is_superuser и is_verified подменяются в backend

Вывод шаблонов и заданий

Template Manager

[Templates](#)[Tasks](#)

a@aaa.aaa

[Logout](#)[Create template](#)

template_id	name	inner_dataset_mask	description	status
1	template1	.test.	<pre>{ "steps": { "decoding": { "run": { "class": "CommandLineTool", "baseCommand": "echo", "inputs": { "dataset_name": { "type": "string", "processing_program": { "type": "string", "processing_program_version": { "type": "string", "cable_map": { "type": "File", "input_params": { "type": "File" }, "outputs": { "output_dataset": { "type": "File", "log_dataset": { "type": "File" } } }, "in": { "dataset_name": ".test.", "processing_program": "processing_program", "processing_program_version": "processing_program_version", "cable_map": "cable_map", "input_params": "input_params", "out": { "output_dataset, log_dataset" }, "reconstruction": { "run": { "class": "CommandLineTool", "baseCommand": "echo", "inputs": { "dataset_name": { "type": "string", "processing_program": { "type": "string", "processing_program_version": { "type": "string", "cable_map": { "type": "File", "input_params": { "type": "File" }, "outputs": { "output_dataset": { "type": "File", "log_dataset": { "type": "File" } } }, "in": { "dataset_name": ".test.", "processing_program": "processing_program", "processing_program_version": "processing_program_version", "cable_map": "cable_map", "input_params": "input_params", "out": { "output_dataset, log_dataset" } } } } } } } } } } } } } } } } } } }</pre>	ACTUAL
2	template2	.test.	<pre>{ "steps": { "decoding": { "run": { "class": "CommandLineTool", "baseCommand": "echo", "inputs": { "dataset_name": { "type": "string", "processing_program": { "type": "string", "processing_program_version": { "type": "string", "cable_map": { "type": "File", "input_params": { "type": "File" }, "outputs": { "output_dataset": { "type": "File", "log_dataset": { "type": "File" } } }, "in": { "dataset_name": ".test.", "processing_program": "processing_program", "processing_program_version": "processing_program_version", "cable_map": "cable_map", "input_params": "input_params", "out": { "output_dataset, log_dataset" } } } } } } } } } } }</pre>	ARCHIVED

Template Manager

[Templates](#)[Tasks](#)

a@aaa.aaa

[Logout](#)

task_id	wflow_id	exec	args	rank	device	mode	retry	datas_in_id	datas_out_id	datas_log_id	status
11	6	processing_program	cable_map	1	CPU	map	5	26	27	28	IN_PROGRESS
12	6	processing_program	cable_map	1	CPU	map	5	27	29	30	IN_PROGRESS
13	7	processing_program	cable_map	1	CPU	map	5	31	32	33	IN_PROGRESS
14	7	processing_program	cable_map	1	CPU	map	5	32	34	35	IN_PROGRESS
15	8	processing_program	cable_map	1	CPU	map	5	36	37	38	IN_PROGRESS
16	8	processing_program	cable_map	1	CPU	map	5	37	39	40	IN_PROGRESS
17	9	processing_program	cable_map	1	CPU	map	5	41	42	43	IN_PROGRESS
18	9	processing_program	cable_map	1	CPU	map	5	42	44	45	IN_PROGRESS
19	10	processing_program	cable_map	1	CPU	map	5	46	47	48	IN_PROGRESS
20	10	processing_program	cable_map	1	CPU	map	5	47	49	50	IN_PROGRESS
21	11	processing_program	cable_map	1	CPU	map	5	51	52	53	IN_PROGRESS
22	11	processing_program	cable_map	1	CPU	map	5	52	54	55	IN_PROGRESS
23	12	processing_program	cable_map	1	CPU	map	5	56	57	58	IN_PROGRESS
24	12	processing_program	cable_map	1	CPU	map	5	57	59	60	IN_PROGRESS

- Просмотр шаблонов и заданий доступен всем пользователям, прошедшим процесс авторизации
- Создание шаблонов доступно только суперпользователям (в частности, отсутствует кнопка в интерфейсе)

Создание шаблонов

- Предварительная валидация шаблонов с помощью инструментов cwltools
- Сохранение шаблона в БД

Template Manager

Templates

Tasks

a@aaa.aaa [Logout](#)

CWL Template

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.

Шаблон

```
{
  "name": "template1",
  "description": "(CWL)",
  "inner_dataset_mask": ".test",
  "template_id": 1,
  "status": "ACTUAL"
}
```

Датасеты и процесс обработки

```
[
  {
    "type": "input",
    "dataset_id": 26,
    "dataset_uid": "5a2775f5-73c5-4328-bdf3-166cff26a594",
    "name": "input.test.368d6f34-3925-4272-9d40-3059295a7fcd.raw"
  },
  {
    "type": "output",
    "dataset_id": 27,
    "dataset_uid": "37222a88-6bde-4183-ba4f-06d164ae689b",
    "name": "input.test.368d6f34-3925-4272-9d40-3059295a7fcd.raw.output.1"
  },
  {
    "type": "output",
    "dataset_id": 28,
    "dataset_uid": "410dcf92-7534-4009-ba3d-55222161dfc8",
    "name": "input.test.368d6f34-3925-4272-9d40-3059295a7fcd.raw.log.1"
  },
  {
    "type": "output",
    "dataset_id": 29,
    "dataset_uid": "2a81e72d-5368-4c57-a752-6a74cb787e2a",
    "name": "input.test.368d6f34-3925-4272-9d40-3059295a7fcd.raw.output.2"
  },
  {
    "type": "output",
    "dataset_id": 30,
    "dataset_uid": "391baf82-5fdb-482a-8736-276844452f45",
    "name": "input.test.368d6f34-3925-4272-9d40-3059295a7fcd.raw.log.2"
  }
]
```

```
{
  "template_id": 1,
  "workflow_id": 6
}
```

Задания

```
[
  {
    "task_id": 11,
    "args": "cable_map",
    "device_type": "CPU",
    "retry": 5,
    "dataset_log_id": 28,
    "workflow_id": 6,
    "executable": "processing_program",
    "rank": 1,
    "mode": "map",
    "dataset_in_id": 26,
    "dataset_out_id": 27,
    "status": "DEFINED"
  },
  {
    "task_id": 12,
    "args": "cable_map",
    "device_type": "CPU",
    "retry": 5,
    "dataset_log_id": 30,
    "workflow_id": 6,
    "executable": "processing_program",
    "rank": 1,
    "mode": "map",
    "dataset_in_id": 27,
    "dataset_out_id": 29,
    "status": "DEFINED"
  }
]
```

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".

Задания до

```
[
  {
    "task_id": 11,
    "args": "cable_map",
    "device_type": "CPU",
    "retry": 5,
    "dataset_log_id": 28,
    "workflow_id": 6,
    "executable": "processing_program",
    "rank": 1,
    "mode": "map",
    "dataset_in_id": 26,
    "dataset_out_id": 27,
    "status": "DEFINED"
  },
  {
    "task_id": 12,
    "args": "cable_map",
    "device_type": "CPU",
    "retry": 5,
    "dataset_log_id": 30,
    "workflow_id": 6,
    "executable": "processing_program",
    "rank": 1,
    "mode": "map",
    "dataset_in_id": 27,
    "dataset_out_id": 29,
    "status": "DEFINED"
  }
]
```

RabbitMQ

Messages: 1
Get Message(s)

Message 1

The server reported 5 messages remaining.

Exchange	wfms.manager
Routing Key	wfms.manager.tasks.key
Redelivered	<input type="radio"/>
Properties	
Payload	{ "task_id": 29, "executable": "processing_program", "args": "cable_map", "rank": 1, "device_type": "CPU", "mode": "map", "retry": 5, "dataset_in_uid": ["6e08a8a7-fec3-4e58-aaf8-ed33265af1c0"], "dataset_in...
375 bytes	
Encoding:	string

Задания после

```
[
  {
    "task_id": 11,
    "args": "cable_map",
    "device_type": "CPU",
    "retry": 5,
    "dataset_log_id": 28,
    "workflow_id": 6,
    "executable": "processing_program",
    "rank": 1,
    "mode": "map",
    "dataset_in_id": 26,
    "dataset_out_id": 27,
    "status": "IN_PROGRESS"
  },
  {
    "task_id": 12,
    "args": "cable_map",
    "device_type": "CPU",
    "retry": 5,
    "dataset_log_id": 30,
    "workflow_id": 6,
    "executable": "processing_program",
    "rank": 1,
    "mode": "map",
    "dataset_in_id": 27,
    "dataset_out_id": 29,
    "status": "DEFINED"
  }
]
```

Отправление заданий на обработку

Логирование, контейнеризация и оркестрация

- Логирование: кастомный логгер

```
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.
```

- Контейнеризация и оркестрация:
Docker и Docker Compose

docker-compose

```
services:
  db_api:
  ...
  task_generator:
  ...
  task_manager:
  ...
  template_manager
  ...

networks:
  wfms:
```

Планы

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.

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



**Спасибо за
внимание!**