



СИСТЕМА УПРАВЛЕНИЯ ПРОЦЕССАМИ ОБРАБОТКИ ДЛЯ СПЕЦИАЛИЗИРОВАННОГО ВЫЧИСЛИТЕЛЬНОГО КОМПЛЕКСА SPD ONLINE FILTER

Студент

Научный консультант

А. В. Плотников

Д. А. Олейник

Москва

2024



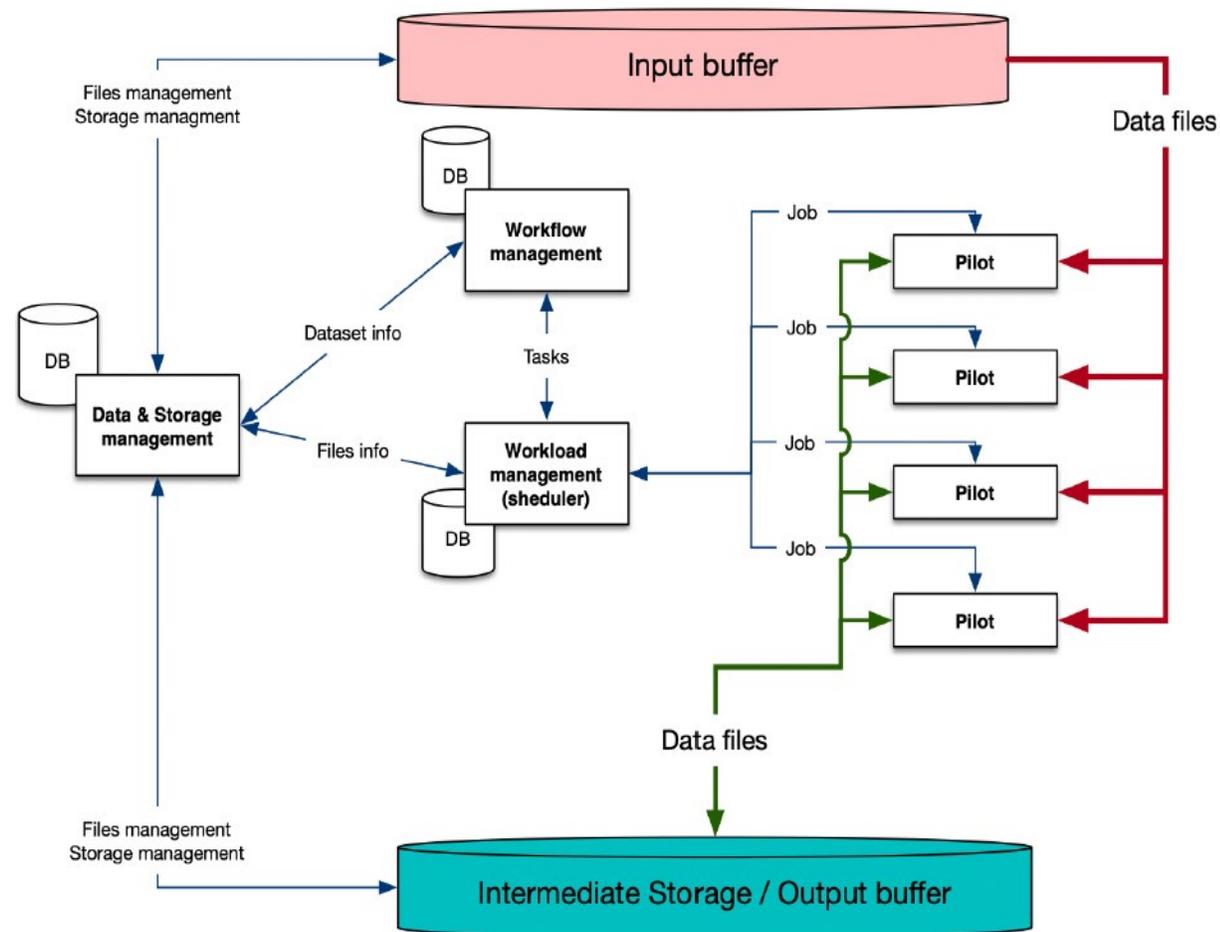
SPD Online Filter

Цель:

Отобрать и сохранить только нужные данные.

Задачи:

- 1) подготовка входных данных;
- 2) частичная реконструкция;
- 3) сравнение данных после частичной реконструкции с данными, полученными после прогонки через обученную модель;
- 4) фильтрация;
- 5) подготовка выходных данных.



Архитектура Online Filter

Управление процессами обработки данных

Концепция базовой обработки:

- Реконструкция треков и связывание их с вершинами;
- Связывание срабатываний электромагнитного калориметра и пробежной системы с каждой вершиной по времени;
- Определение несвязанных срабатываний детектора;
- Объединение с необработанными данными от других субдетекторов;
- Формирование блоков данных и сохранение частично реконструированных событий.

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

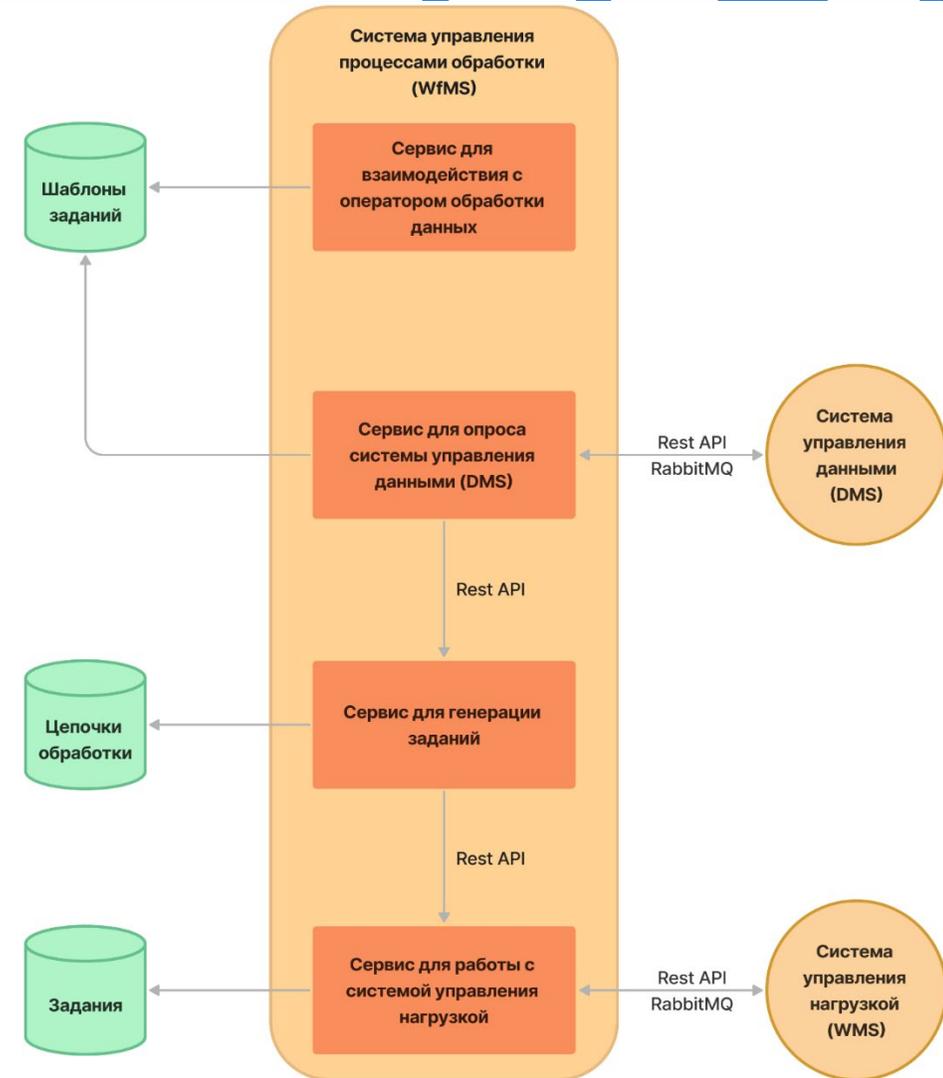
Система управления процессами обработки (WfMS)

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.



Сервисы WfMS

Сервис для взаимодействия с оператором обработки данных:

1. вывод информации о заданиях;
2. изменение приоритетов заданий или их отмена.

Сервис для опроса DMS:

1. получение информации о готовности входных датасетов;
2. формирование выходного датасета и датасета логов;
3. удаление датасета.

Сервис для работы с WMS:

1. отправление заданий на обработку;
2. периодический опрос для отслеживания статуса;
3. изменение приоритета заданий;
4. отмена заданий.

Сервис генерации заданий:

1. определение последовательностей обработки данных;
2. запрос шаблонов;
3. создание заданий по шаблонам для последовательностей.

Взаимодействие с DMS

Запрос	Тип запроса	Input 1	Input 2	Output	Примечания
Получение входного датасета	Rest API (GET)	Timestamp / None	-	[{ • uid • timestamp • dat_in_uid • file_num }, ...]	Возвращает список параметров датасетов, добавленных в систему позднее timestamp. В случае None возвращает все датасеты.
Создание выходного датасета	Rest API (POST)	Name	meta	UID	
Удаление датасета	RabbitMQ (DELETE)	UID	-	-	Удаление датасета — область ответственности DMS.
Получение информации о хранилище	Rest API (GET)	Type	-	ID, URL	
Получение статуса датасета	Rest API	UID	-	Status	Определение готовности датасета к дальнейшей обработке

- Асинхронность
- Восстановление при сбоях в работе – входные данные для первого этапа остаются до тех пор, пока не появятся выходные данные

Взаимодействие с WMS

Запрос	Тип запроса	Input 1	Input 2	Output
Отправка задания	RabbitMQ (POST)	Task	-	-
Получение статуса обработки задач	Rest API (GET)	Id	-	{ <ul style="list-style-type: none"> • total • running • canceled • killed • failed }
Изменение приоритета задания	Rest API (PUT)	Id	rank	-
Отмена задания	Rest API (PUT)	Id	-	-

Взаимодействие WfMS и WMS

Ключ	Тип	Описание
task_id	int	Уникальный идентификатор задания
executable	str	Исполняемый файл запуска
args	str None	Флаги для исполняемого файла задания
rank	int	Приоритет задания
device_type	str	Определение выбранного вычислительного ресурса, который будет использоваться для выполнения задания (CPU/GPU)
mode	str	Тип задания (Map/Merge)
retry	int	Максимальное число попыток выполнения задач в случае неудачной попытки
dat_in_uid	list[int]	Уникальные идентификаторы входного датасета
dat_out_uid	list[int]	Уникальные идентификаторы выходного датасета
dat_stor_url	str	Унифицированный указатель ресурса выходного хранилища
dat_out_orl	str	Унифицированный указатель ресурса выходного датасета
log_out_uid	int	Уникальный идентификатор выходного датасета логов
log_stor_url	str	Унифицированный указатель ресурса выходного хранилища для хранения логов
log_out_url	str	Унифицированный указатель ресурса выходного датасета логов

Описание задания

Отправка задания в очередь

FastAPI 0.1.0 OAS 3.1

/openapi.json

default

GET / Get Status

POST / Task Response

Responses

Curl

```
curl -X 'POST' \
  'http://0.0.0.0:8000/' \
  -H 'accept: application/json' \
  -H 'Content-Type: application/json' \
  -d '{
    "task_id": 0,
    "executable": "string",
    "args": "string",
    "rank": 1,
    "device_type": "string",
    "mode": "string",
    "retry": 0,
    "dat_in_uid": [
      0
    ],
    "dat_out_uid": [
      0
    ],
    "dat_stor_url": "string",
    "dat_out_url": "string",
    "log_out_uid": 0,
    "log_stor_url": "string",
    "log_out_url": "string"
  }'
```

Request URL

http://0.0.0.0:8000/

Server response

Code Details

200

Response body

```
{
  "task_id": 0,
  "executable": "string",
  "args": "string",
  "rank": 1,
  "device_type": "string",
  "mode": "string",
  "retry": 0,
  "dat_in_uid": [
    0
  ],
  "dat_out_uid": [
    0
  ],
  "dat_stor_url": "string",
  "dat_out_url": "string",
  "log_out_uid": 0,
  "log_stor_url": "string",
  "log_out_url": "string"
}
```

Download

Response headers

```
content-length: 254
content-type: application/json
date: Wed, 05 Jun 2024 13:42:34 GMT
server: uvicorn
```

Get messages

Warning: getting messages from a queue is a destructive action. ?

Ack Mode:

Encoding:

Messages:

Get Message(s)

Message 1

The server reported 0 messages remaining.

Exchange: test_exchange

Routing Key: test_key

Redelivered: 0

Properties

Payload

254 bytes

Encoding: string

```
{"task_id":0,"executable":"string","args":"string","rank":1,"device_type":"string","mode":"string","retry":0,"dat_in_uid":[0],"dat_out_uid":[0],"dat_stor_url":"string","dat_out_url":"string","log_out_uid":0,"log_stor_url":"string","log_out_url":"string"}
```



Спасибо за внимание!