# Документация

Использование сервиса осуществляется посредством HTTP запросов к фиксированному URL сервера. Далее приведено описание возможных запросов.

## Организация

* 1. Нода

Информация о подключенных машинах-исполнителях (нодах) хранится в объектах Client.

|  |  |  |
| --- | --- | --- |
| Запрос | Адрес | Комментарий |
| GET | /client | Список всех доступных объектов Client |
| GET | /client/<id> | Объект Client с Id=id |
| POST | /client | Подключение новой ноды |
| DELETE | /client/<id> | Удаление объекта Client с Id=id |

Вывод запроса GET для одной ноды

{

"Id": 3,

"ClientIP": "127.0.0.1",

"ClientPort": "5001",

"Busy": false

}

При добавлении объекта необходимо задать следующие поля:

|  |  |  |  |
| --- | --- | --- | --- |
| Название | Тип данных | Комментарий | Опции |
| ClientIP | STRING | IP ноды в стандартном формате |  |
| ClientPort | STRING | Порт для обращения к ноде |  |

## Данные

* 1. Формат данных

DataFormat задает формат данных, хранящихся в датасете. Также используется при создании задачи (Task) или имплементации (ProgramImplementation) для определения формата входных и выходных данных.

Работа с DataFormat осуществляется с помощью запросов по адресу /data/formats. Доступные запросы:

|  |  |  |
| --- | --- | --- |
| Запрос | Адрес | Комментарий |
| GET | /data/formats | Список всех доступных объектов DataFormat |
| GET | /data/formats/<id> | Объект DataFormat с Id=id |
| POST | /data/formats | Добавление нового объекта |
| DELETE | /data/formats/<id> | Удаление объекта DataFormat с Id=id |

Вывод запроса GET для одного объекта:

{

"Id": 3,

"FormatName": "csv",

"FormatType": "File",

"FormatSchema": null,

"FormatExample": null,

"BaseEntityId": 3

}

При добавлении объекта необходимо задать следующие поля:

|  |  |  |  |
| --- | --- | --- | --- |
| Название | Тип данных | Комментарий | Опции |
| FormatName | STRING | Формат хранения данных |  |
| FormatType | STRING | Enum[File, Table] |  |
| FormatSchema | STRING | Описание schema | Optional |
| Example | STRING | Пример формата данных | Optional |

Пример:

curl <server\_url>/data/formats -X POST -d “Format=CSV”

* 1. Датасет

DataSet определяет расположение файла с данными.

Работа с DataSet осуществляется с помощью запросов по адресу /data/sets. Доступные запросы:

|  |  |  |
| --- | --- | --- |
| Запрос | Адрес | Комментарий |
| GET | /data/sets | Список всех доступных объектов DataSet |
| GET | /data/sets/<id> | Объект DataSet с Id=id |
| POST | /data/sets | Добавление нового объекта |
| DELETE | /data/sets/<id> | Удаление объекта DataSet с Id=id |

Вывод запроса GET для одного датасета:

{

"Id": 10,

"DataFormatId": 3,

"Content": "<schema>://[host]/path",

"BaseEntityId": 26

}

При добавлении объекта необходимо задать следующие поля:

|  |  |  |  |
| --- | --- | --- | --- |
| Название | Тип данных | Комментарий | Опции |
| DataFormatId | INT | Id объекта DataFormat, тип формата хранения данных |  |
| Content | STRING | Ссылка на данные. Connection string в случае таблицы, URL в случае файла. |  |

Пример:

curl <server\_url>/data/sets -X POST -d “DataFormatId=5&Content=<schema>://[host]/path”

### Эксперимент

* 1. Имплементация

Имплементация модели задается в объекте ProgramImplementation.

Работа с ProgramImplementation осуществляется с помощью запросов по адресу /implementations. Доступные запросы:

|  |  |  |
| --- | --- | --- |
| Запрос | Адрес | Комментарий |
| GET | /implementations | Список всех доступных объектов ProgramImplementation |
| GET | /implementations/<id> | Объект ProgramImplementation с Id=id |
| POST | /implementations | Добавление нового объекта ProgramImplementation |
| DELETE | /implementations/<id> | Удаление объекта ProgramImplementation с Id=id |

Вывод запроса GET для одной имплементации:

{

"Id": 19,

"Status": "Active"(“Deprecated”, “Reference”),

"OS": "Win",

"EnvironmentId": 1,

"TaskId": 2,

"InputFormat": 3,

"OutputFormat": 3,

"CommandLineArgs": "--n 55",

"FileURL": "<schema>://<host>/<path>",

"FilePath": "<local\_path\_to\_file>",

"DataProcessing": 0,

"DownloadSuccess": "True",

"ProgramType": "Exec",

"PythonRequirements": null,

"BaseEntityId": 68

}

При добавлении объекта необходимо задать следующие поля:

|  |  |  |  |
| --- | --- | --- | --- |
| Название | Тип данных | Комментарий | Опции |
| OS | STRING | Enum[Win, Linux, MacOS] |  |
| EnvironmentId | INT | Id объекта ExperimentEnvironment |  |
| TaskId | INT | Id объекта Task |  |
| InputFormat | INT | Id объекта DataFormat, формат входных данных |  |
| OutputFormat | INT | Id объекта DataFormat, формат выходных данных |  |
| CommandLineArgs | STRING | аргументы командной строки для запуска модели | Optional |
| FileURL | STRING | URL файла имплементации |  |
| DataProcessing | INT | Способ скачивания входных данных. 0 – скачивание всех данных, число 1-99 – процент от общего количества данных, скачиваемых в одном батче |  |
| ProgramType | STRING | Enum[Python, PythonZip, Exec] – тип программы |  |
| PythonRequirements | STRING | URL файла requirements.txt, если тип программы является Python или PythonZip | Optional |

Пример:

curl <server\_url>/implementations -X POST -d “OS=Win&EnvironmentId=1&TaskId=5&InputFormat=2&OutputFormat=2%DataProcessing=0&ProgramType=Exec&FileURL= file://[host]/path”

* 1. Требования к окружению

Требования к окружению для запуска эксперимента задаются в объекте ExperimentEnvironment. Они описывают минимальные условия, необходимые для успешного запуска эксперимента.

Работа с ExperimentEnvironment осуществляется с помощью запросов по адресу /experiment/enivornments. Доступные запросы:

|  |  |  |
| --- | --- | --- |
| Запрос | Адрес | Комментарий |
| GET | /experiment/environments | Список всех доступных объектов ExperimentEnvironment |
| GET | /experiment/environments /<id> | Объект ExperimentEnvironment с Id=id |
| POST | /experiment/environments | Добавление нового объекта ExperimentEnvironment |
| DELETE | /experiment/environments /<id> | Удаление объекта ExperimentEnvironment с Id=id |

Вывод запроса GET для одного объекта:

{

"Id": 1,

"OS": "Win",

"Processor": "x86",

"Memory": 500,

"HDD": 1000,

"BaseEntityId": 29

}

При добавлении объекта необходимо задать следующие поля:

|  |  |  |  |
| --- | --- | --- | --- |
| Название | Тип данных | Комментарий | Опции |
| OS | STRING | Enum[Win, Linux, MacOS] |  |
| Processor | STRING | тип процессора |  |
| Memory | INT | необходимый объем оперативной памяти в мегабайтах |  |
| HDD | INT | необходимый объем свободного места на диске в мегабайтах |  |

Пример:

curl <server\_url>/experiment/environments -X POST -d “OS=Win&Processor=x86&Memory=200&HDD=500”

* 1. Задача (Task)

Задача, которую выполняет эксперимент, задается в объекте Task.

Работа с Task осуществляется с помощью запросов по адресу /tasks.

Доступные запросы:

|  |  |  |
| --- | --- | --- |
| Запрос | Адрес | Комментарий |
| GET | /tasks | Список всех доступных объектов Task |
| GET | /tasks/<id> | Объект Task с Id=id |
| POST | /tasks | Добавление нового объекта Task |
| DELETE | /tasks/<id> | Удаление объекта Task с Id=id |

Вывод запроса GET для одного объекта:

{

"Id": 2,

"Title": "Test task",

"Input": "File",

"DefaultInputFormat": 3,

"Output": "File",

"DefaultOutputFormat": 3,

"ResultQuality": "none",

"BaseEntityId": 20

}

При добавлении объекта необходимо задать следующие поля:

|  |  |  |  |
| --- | --- | --- | --- |
| Название | Тип данных | Комментарий | Опции |
| Title | STRING | название, описывающее задачу |  |
| Input | STRING | Вид входных данных, Enum[Table, File] |  |
| DefaultInputFormat | INT | Id объекта DataFormat, формат входных данных |  |
| Output | STRING | Вид выходных данных, Enum[Table, File] |  |
| DefaultOutputFormat | INT | Id объекта DataFormat, формат выходных данных |  |
| ResultQuality | STRING | метрика, используемая для определения качества |  |

* 1. Эксперимент

Эксперимент задается в объекте Experiment.

Работа с Experiment осуществляется с помощью запросов по адресу /experiment/experiments.

Доступные запросы:

|  |  |  |
| --- | --- | --- |
| Запрос | Адрес | Комментарий |
| GET | /experiment/experiments | Список всех доступных объектов Experiment |
| GET | /experiment/experiments/<id> | Объект Experiment с Id=id |
| POST | /experiment/experiments | Добавление нового объекта Experiment |

Вывод запроса GET для одного объекта:

{

"Id": 1,

"Title": "Test Experiment",

"TitleShort": "Test Exp",

"Comment": null,

"BaseLine": null,

"RefRes": null,

"EnvironmentId": 1,

"TaskId": 2,

"BaseEntityId": 35,

"DataSets": [

10,

11

]

}

При добавлении объекта необходимо задать следующие поля:

|  |  |  |  |
| --- | --- | --- | --- |
| Название | Тип данных | Комментарий | Опции |
| Title | STRING | название эксперимента, описывающее его |  |
| TitleShort | STRING | краткое название эксперимента |  |
| Comment | STRING | комментарий | Optional |
| DataSets | STRING  (INT,INT,…) | Список чисел, разделенных запятыми, список Id датасетов |  |
| BaseLine | STRING | результат бейзлайна | Optional |
| RefRes | STRING | референсный результат | Optional |
| EnvironmentId | INT | Id объекта ExperimentEnvironment |  |
| TaskId | INT | Id объекта Task |  |

## Запуск эксперимента

Перед запуском эксперимента необходимо создать объект Experiment и ProgramImplementation. Запуск производится с помощью POST запроса по адресу /experiment/run.

|  |  |  |
| --- | --- | --- |
| Запрос | Адрес | Комментарий |
| POST | /experiment/run | Запустить эксперимент, используя выбранную имплементацию |

Необходимо задать следующие поля:

|  |  |  |  |
| --- | --- | --- | --- |
| Название | Тип данных | Комментарий | Опции |
| ExperimentId | INT | Id объекта Experiment |  |
| ProgramImplementationId | INT | Id объекта ProgramImplementation |  |
| OutputString | STRING | Connection String для вывода результата (только если формат вывода имеет тип Table) | Optional |

# Запуск сервера и ноды

### Настройки окружения сервера

Перед запуском сервера необходимо установить параметры окружения:

|  |  |  |
| --- | --- | --- |
| Название | Комментарий | Опции |
| DATABASE\_URL | Connection String для подключения к базе данных SQL |  |
| IMPLEMENTATION\_DIR | Директория для сохранения файлов имплементаций |  |
| FLASK\_APP | Файл server.py |  |

### Запуск сервера

Запуск осуществляется вызовом команды flask run --host <host>

### Настройки окружения ноды

Перед запуском ноды необходимо установить параметры окружения:

|  |  |  |
| --- | --- | --- |
| Название | Комментарий | Опции |
| DATABASE\_URL | Connection String для подключения к базе данных SQL |  |
| IMPLEMENTATION\_DIR | Директория для сохранения файлов имплементаций |  |
| OUTPUT\_DIR | Директория для сохранения результатов запусков |  |
| SERVER\_ADDR | Адрес сервера |  |
| FLASK\_APP | Файл server.py |  |

Запуск клиента аналогично серверу.

### Добавление ноды в базу данных

Далее необходимо добавить информацию о ноде в базу данных с помощью POST запроса.

# Этапы создания и запуска эксперимента

### Загрузка данных

### Создание формата данных

Создать новый объект DataFormat, описывающий формат, в котором хранятся данные. Тип формата может быть файлом или таблицей (File или Table).

### Создание датасета

Создать новый объект DataSet, описывающий используемый датасет. В качестве DataFormatId используется Id созданного на прошлом шаге объекта DataFormat. Если тип формата – файл, то в поле Content необходимо указать URL для скачивания файла. Если тип формата – таблица, то в поле Content необходимо указать Connection String SQL таблицы.

### Создание эксперимента

### Создание требований к окружению

Создать новый объект ExperimentEnvironment, описывающий базовые требования к окружению, необходимые для исполнения имплементации программы.

### Создание задачи (Task)

Создать объект Task, описывающий исполняемую задачу.

### Загрузка файла имплементации

Создать объект ProgramImplementation, описывающий исполняемый файл программы. В InputFormat указать Id созданного объекта DataFormat. В поле EnvironmentId указать Id созданного объекта ExperimentEnvironment.

Доступны два формата – Executable и Python. Если формат – python, то в поле PythonRequirements возможно указать ссылку на файл requirements.txt, содержащий список зависимостей программы.

Исполняемый файл должен поддерживать две обязательные опции «--input» и «--output». В первой подается путь к файлу с данными или connection string, во второй – путь к директории для вывода или connection string выхода.

### Создание эксперимента

Создать объект Experiment, описывающий эксперимент. В поле DataSets указать список разделенных запятой объектов DataSet, на которых по очереди будет запущена программа.

### Запуск эксперимента

С помощью POST запроса по адресу /experiment/run запустить созданный эксперимент с загруженной ранее имплементацией. За статусом исполнения можно следить с помощью GET запросов по этому адресу.