



KOLMOGOROV

Программное обеспечение «Kolmogorov AI Morphism»

Описание функциональных характеристик

ООО «Дата Сапиенс», kolmogorov.ai

2022

Оглавление

Оглавление	2
1. Аннотация	3
2. Термины и определения	3
3. Основные функции программного обеспечения «Kolmogorov.ai Morphism»	4
Архитектура	4
Создание зерна	4
Создание переменной	5
Регистрация загрузчика	6
Регистрация датасета	8
Экспорт датасета	9
4. Программно-аппаратные требования для установки и функционирования программного обеспечения «Kolmogorov.ai Morphism»	10
Аппаратные требования	10
Требования к операционной системе	10
Требования к установленному программному обеспечению	10
6. Режим функционирования программного обеспечения	10

1. Аннотация

Документ содержит описание функциональных характеристик программного обеспечения «Kolmogorov.ai Morphism» в части ее базового функционала.

2. Термины и определения

Термин	Определение
Программное обеспечение	Kolmogorov.ai Morphism
Зерно	Логическая сущность, объединяющая ключи для связи данных внутри системы
Переменная	Логическая сущность, объединяющая различные алгоритмы ее расчета (версии переменной)
Версия переменной	Вариант реализации переменной
CSV	Текстовый формат, предназначенный для представления табличных данных
SQL	Структурированный язык запросов
ETL	Процесс извлечения (extract), преобразования (transform) и загрузки (load) данных из источника в приемник
Загрузчик	Инструмент загрузки данных в систему
Датасет	Материализованный набор данных в виде таблицы в БД, рассчитанный в разрезе выбранного зерна с использованием версий переменных, согласно конфигурации расчёта
БД	База данных
Stage	Интеграционный слой для загрузки данных в систему

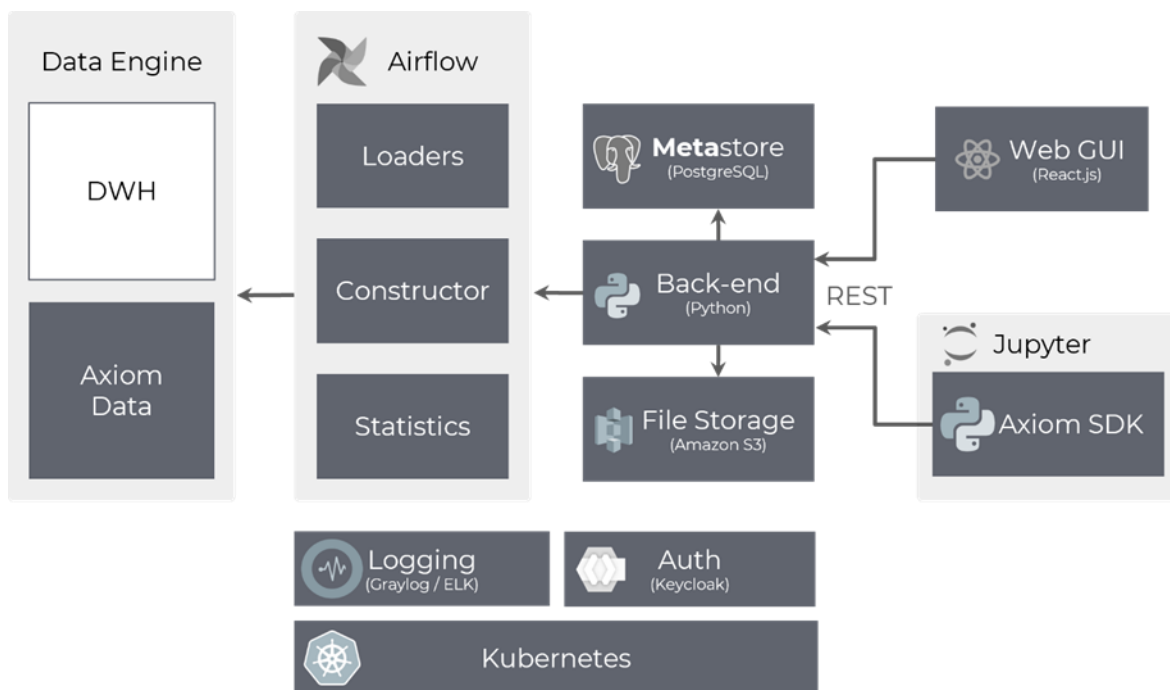
3. Основные функции программного обеспечения «Kolmogorov.ai Morphism»

Программное обеспечение «Kolmogorov.ai Morphism» представляет собой инструмент для управления данными и переменными в задачах анализа данных и продвинутой аналитики — организует каталог всех переменных и упрощает задачу их вывода в промышленную эксплуатацию.

ПО поддерживает следующие функции:

- Создание зерна;
- Создание переменной;
- Регистрация загрузчика;
- Регистрация датасета;
- Экспорт датасета.

Архитектура



Создание зерна

Функция позволяет пользователю создавать зерна и ключи.

Цель функции: создание сущности для связи данных внутри системы.

Параметры:

№	Параметр	Тип	Проверка
1	Название зерна	Входной	Обязательный
2	Описание зерна	Входной	Необязательный
3	Название ключа зерна	Входной	Обязательный
4	Описание ключа зерна	Входной	Необязательный
5	Тип данных ключа зерна	Входной	Обязательный
6	Зерно	Выходной	-
7	Ключ зерна	Выходной	-

Предусловия: нет.

Постусловия: система должна зарегистрировать зерно и ключи, определить метаданные полей для хранения значений ключей зерна.

Создание переменной

Функция позволяет пользователю зарегистрировать переменную в каталоге.

Цель функции: создание переменной для последующей загрузки данных и использования в конструкторе датасета.

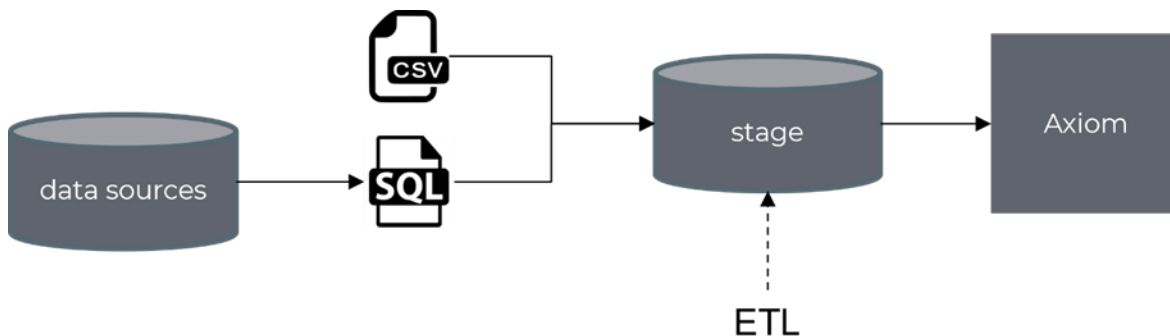
Параметры:

№	Параметр	Тип	Проверка
1	Название переменной	Входной	Обязательный
2	Описание переменной	Входной	Необязательный
3	Переменная	Выходной	-

Предусловия: нет.

Постусловия: система должна зарегистрировать переменную в каталоге.

Регистрация загрузчика



Функция позволяет пользователю зарегистрировать загрузчик. Загрузка во внутренние структуры системы производится через интеграционный слой (stage). Таблицы интеграционного слоя могут являться промежуточными материализациями при загрузке данных из SQL-скрипта и CSV-файла или заполняться внешним ETL.

Цель функции: загрузка данных из внешних источников для использования в конструкторе датасетов.

Параметры:

№	Параметр	Тип	Проверка
1	Название загрузчика	Входной	Обязательный
2	Описание загрузчика	Входной	Необязательный
3	Источник данных	Входной	Обязательный
4	CSV-файл и метаданные: - Разделитель - Признак наличия заголовка	Входной	Если Источник данных = «Файл», то обязательный, иначе не передается
5	SQL-скрипт	Входной	Если Источник данных = «SQL», то обязательный, иначе не передается
6	Таблица stage	Входной	Если Источник данных = «Table», то обязательный, иначе не передается
7	Зерно	Входной	Обязательный

8	Гранулярность данных	Входной	Если источник = «Предагрегат», то необязательный, иначе не передается
9	Режим загрузки	Входной	Если источник = «Готовые переменные», то обязательный, иначе не передается
10	Атрибут источника для даты актуальности или значение даты актуальности	Входной	Обязательный
11	Атрибуты источника для ключей зерна	Входной	Обязательный
12	Атрибуты источника для версий переменных с метаданными приемника: - Переменная из каталога - Существующая версия переменной или метаданные новой версии. Если выбрана новая версия, то передается: - Название версии переменной - Описание версии переменной - Тип данных	Входной	Если приемник = «Сегмент», то не передается, иначе обязательный
13	Загрузчик	Выходной	
14	Версия переменной	Выходной	Только при создании новых версий переменных
15	Сегмент	Выходной	Только если приемник = «Сегмент»
16	Данные в БД	Выходной	

Предусловия:

1. Данные источника подготовлены для загрузки;
- Страница 7 из 10

2. Используемое зерно зарегистрировано в каталоге;
3. Используемые переменные зарегистрированы в каталоге.

Постусловия:

1. Система должна сохранить метаданные по загрузчикам и создать новые версии переменных (если применимо);
2. Система должна загрузить данные из источника во внутренние структуры системы.

Регистрация датасета

Функция позволяет пользователю сконфигурировать и рассчитать датасет для обучения модели.

Цель функции: создание датасета для обучения модели.

Параметры:

№	Параметр	Тип	Проверка
1	Название датасета	Входной	Обязательный
2	Описание датасета	Входной	Необязательный
3	Зерно	Входной	Обязательный
4	Конфигурация переменной датасета: - Версия переменной из каталога - Алиас переменной датасета - Конфигурация преагрегатов - Агрегирующая функция - Условия фильтрации - Формулы расчета	Входной	Обязательный
5	Дата расчета датасета	Входной	Обязательный
6	Сегмент	Входной	Необязательный
7	Фильтрация датасета	Входной	Необязательный

8	Датасет	Выходной	
9	SQL-код расчёта датасета	Выходной	
10	Данные в БД	Выходной	

Предусловия:

1. Зерно и используемые переменные зарегистрированы в каталоге;
2. Данные по используемым версиям переменных загружены во внутренние структуры системы;
3. Используемый сегмент зарегистрирован и загружен.

Постусловия:

1. Система должна сохранить метаданные по датасету;
2. Система должна сохранить файл с SQL-кодом расчета датасета в файловое хранилище;
3. Система должна загрузить данные по датасету согласно настроенной конфигурации;

Экспорт датасета

Функция позволяет пользователю получить SQL-код расчета датасета и данные датасета в формате CSV.

Цель: использование данных и алгоритмов датасета.

Параметры:

№	Параметр	Тип	Проверка
1	Датасет	Входной	Обязательный
2	SQL-код расчёта датасета	Выходной	
3	CSV-файл с данными датасета	Входной	Обязательный

Предусловия: датасет должен быть зарегистрирован и успешно загружен.

Постусловия:

1. Система должна инициировать загрузку CSV-файла с данными датасета;
2. Система должна инициировать загрузку SQL-файла с алгоритмом расчета датасета.

4. Программно-аппаратные требования для установки и функционирования программного обеспечения «Kolmogorov.ai Morphism»

Аппаратные требования

- Оперативная память – не менее 8 Гб
- Частота процессора – не менее 2.3 ГГц
- Количество ядер – не менее 4
- Свободное место на HDD – не менее 100 Гб
- Скорость интернета на входящий и исходящий трафик – не менее 50 Мбит

Требования к операционной системе

Программное обеспечение «Kolmogorov.ai Morphism» устанавливается на кластер серверов под управлением Kubernetes.

Рекомендуемая ОС – Ubuntu 20.04 LTS.

Поддерживаемая ОС – любая *nix-ОС с поддержкой Kubernetes.

Требования к установленному программному обеспечению

На локальную машину пользователя должен быть установлен один из следующих браузеров для доступа к программному обеспечению:

1. Firefox 63.0.1 или выше;
2. Chrome 70.0.3538.67 или выше;
3. Safari 12 или выше;
4. Opera 56.0.3051.99 или выше;
5. Edge 80.0 или выше.

6. Режим функционирования программного обеспечения

В основном режиме функционирования программное обеспечение предполагает:

- выполнение всех функций в полном объеме, за исключением периодов проведения профилактических и других работ, а также устранения возникших нештатных ситуаций;
- непрерывную работу.