
Программное обеспечение «JET GALATEA»

ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК

2022

Оглавление

1	Общие сведения	4
1.1	Обозначение и наименование	4
2	Назначение Jet Galatea	4
3	Условия применения	6
3.1	Конфигурация сервера приложений и управления кластером	6
3.2	Конфигурация сервера хранения данных и вычислений	6
3.3	Конфигурация APM	7
4	Описание задач	7
4.1	Просмотр статистики	7
4.2	Настройка модели данных	7
4.3	Работа с пользовательскими объектами	8
4.4	Обработка и обогащение данных	8
4.5	Ведение служебных справочников	8
4.6	Анализ событий	8
4.7	Настройка механизмов выявления событий	9
4.8	Ведение справочников доступа	Ошибка! Закладка не определена.
4.9	Ведение учётных записей пользователей	Ошибка! Закладка не определена.
4.10	Работа с моделями машинного обучения	9
4.11	Формирование отчетности	Ошибка! Закладка не определена.
5	Входные и выходные данные	10
5.1	Входные данные	10
5.2	Выходные данные	10

Аннотация

В документе описано применение программного обеспечения «Jet Galatea» (далее – JG, ПО).

Краткое содержимое разделов:

- **Общие сведения** – название программного обеспечения и перечень средств его разработки.
- **Назначение Jet Galatea** – назначение ПО и цели его использования, а также возможности, предоставляемые им для решения целевых задач.
- **Условия применения** – требования к программным и техническим средствам ПО.
- **Описание задач** – задачи, выполняемые в Jet Galatea.
- **Входные и выходные данные** – сведения о данных на входе и выходе ПО.

1 Общие сведения

1.1 Обозначение и наименование

Полное наименование: Программное обеспечение «Jet Galatea».

Сокращенное наименование, обозначение: ПО, Jet Galatea, JG.

2 Назначение Jet Galatea

Jet Galatea - решение для обработки данных и принятия решений в автоматическом режиме. Jet Galatea обеспечивает:

- получение, поступающих из различных информационных систем данных в онлайн и оффлайн режимах;
- проведение автоматизированного анализа полученных данных на основании настраиваемой пользователем логики (формирование сегментов, стратегий, сценариев принятия решений);
- формирование результата анализа и взаимодействие с внешними информационными системами, реализующими взаимодействия;
- пользовательский интерфейс для осуществления настроек, проведения качественного анализа данных и определения степени эффективности действующих алгоритмов и их последующей корректировки.
- создания моделей данных анализируемых объектов, разработки рекомендаций на базе технологий искусственного интеллекта и машинного обучения, а также оптимизации бизнес-процессов.

Многофункциональный инструмент настройки JG позволяет без привлечения разработчиков расширять и изменять объекты, настраивать сложные схемы анализа с использованием атрибутов объектов, связанных сущностей и результатов рассчитанных прогнозов. Наличие графического представления полученных прогнозных значений и фактических данных обеспечивает удобный вид обработки информации, полученной от математических моделей.

JG имеет трёхуровневую архитектуру:

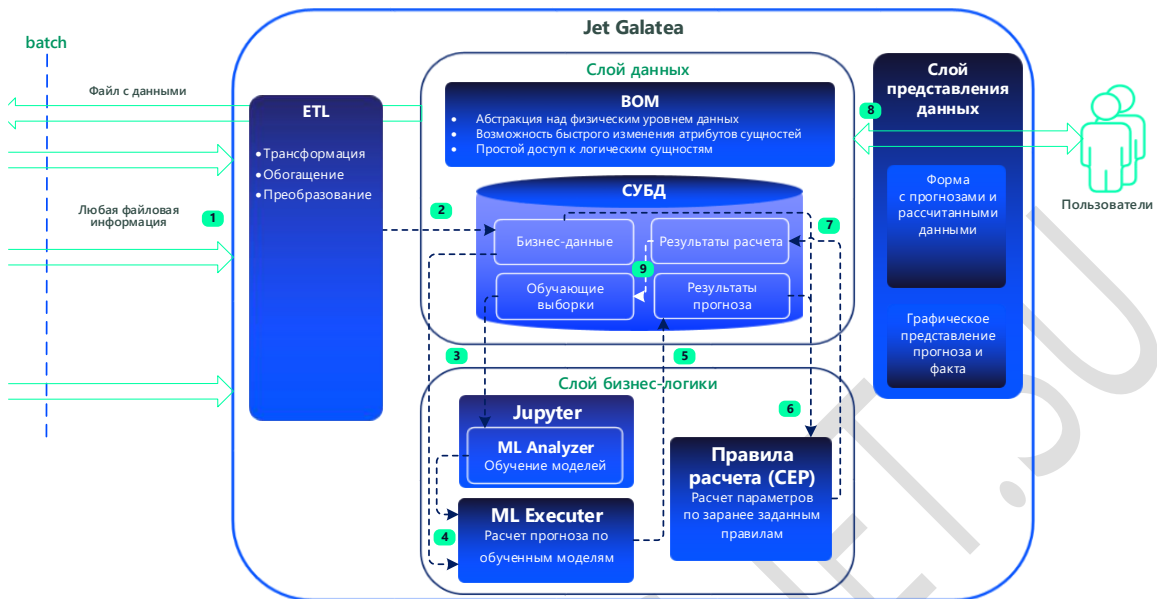
- 1) Клиент – графический интерфейс пользователя.
- 2) Сервер приложений – реализует бизнес-логику, обрабатывает события в ПО и кеширует данные.
- 3) Слой хранения данных, реализованный на базе:
 - реляционной СУБД – используется для хранения метаданных JG и промежуточных результатов подготовки данных;
 - файловой системы – используется для хранения больших объемов информации.

ПО использует открытые форматы обмена данных. Информационная совместимость со смежными системами обеспечивается на уровнях:

- взаимодействия в пакетном режиме;
- потокового взаимодействия;

- экспорта-импорта документов.

Функциональная структура JG приведена на рисунке ниже.



ПО состоит из следующих функциональных модулей:

- Рабочий стол;
- Фабрика данных;
- Правила анализа;
- Анализ событий;
- Машинное обучение;
- Авторизация.

Модуль **Рабочий стол** предназначен для организации быстрого доступа к статистике и действиям, часто выполняемым во время анализа событий. Модуль включает в себя интерфейс пользователя с диаграммами и графиками. С экранной формы рабочего стола можно перейти к тем или иным инструментам анализа событий.

Модуль **Фабрика данных** предназначен для управления данными и реализует следующие функции:

- настройка модели объектов, основанной на концепции модели бизнес-объектов;
- сбор данных объектов из различных источников: баз данных, файлов различных типов, файлов протоколов серверов, интеграционных компонентов смежных систем и прочих;
- трансформация данных: очистка, обогащение, агрегация, связывание данных.

Модуль **Правила анализа** предназначен для настройки и управления правилами, политиками (группами правил) и моделями выявления.

Модуль **Анализ событий** предназначен для обработки входящих потоков данных и реализует следующие функции:

- настройка и испытание правил и политик выявления событий;

- применение правил и политик выявления к потокам данных;
- реагирование на выявленные события: формирование карточки наступившего события; информирование пользователей; формирование ответа, отправляемого в источник данных; выполнение программных сценариев и прочее.
- Анализ событий и связанных с ними объектов с помощью инструментов анализа.

Модуль **Машинное обучение** предназначен для создания и обучения модели выявления событий и реализует следующие функции:

- создание, обучение и проверка модели выявления;
- создание массивов данных для обучения и проверки модели выявления.

Модуль **Авторизация** предназначен для организации доступа пользователей к функциям и данным **JG** и реализует следующие функции:

- настройка справочников доступа: разрешения, владения, роли пользователей, учётные записи пользователей;
- аутентификация и авторизация пользователя;
- применение настройки прав доступа к действиям пользователей.

3 Условия применения

ПО функционирует на серверах на платформах Intel x86, AMD64.

Минимальная рекомендуемая конфигурация состоит из четырёх серверов:

- Сервер приложений и управления кластером – 1 шт.;
- Сервер хранения данных и вычислений – 3 шт.

3.1 Конфигурация сервера приложений и управления кластером

Минимальные характеристики сервера:

- процессор – 4 ядра с тактовой частотой не менее 2.4 ГГц;
- объём оперативной памяти – 16 ГБ;
- объём жёсткого диска \geq 100 ГБ.

Сервер функционирует под управлением операционной системы Centos 7.6. В качестве сервера приложений используется Embedded Tomcat, в качестве веб-сервера используется Nginx.

3.2 Конфигурация сервера хранения данных и вычислений

Минимальные характеристики сервера:

- процессор – 4 ядра с тактовой частотой не менее 2.4 ГГц;
- объём оперативной памяти – 16 ГБ;
- объём жесткого диска \geq 100 ГБ

Все серверы работают под управлением операционной системы RedOs, Альтлинукс 8СП, Centos, На двух серверах размещена СУБД PostgreSQL 11 в режиме репликации (при необходимости).

На всех трёх серверах также размещены следующие компоненты: Kafka zookeeper

3.3 Конфигурация АРМ

Для управления работой ПО используется веб-интерфейс администратора и интерфейс

Компьютер администратора должен соответствовать характеристикам:

- процессор – не менее 4 ядер с тактовой частотой не менее 2,8 ГГц;
- объём оперативной памяти – не менее 4 ГБ;
- объём жёсткого диска – не менее 128 ГБ;
- разрешение экрана при работе с интерфейсом – не менее 1024x768;
- сетевой адаптер – Ethernet.

На АРМ пользователей должен устанавливаться веб-браузер Mozilla Firefox версии 44.0.2 и выше.

4 Описание задач

Задачи **Jet Galatea**:

- просмотр статистики;
- настройка модели данных;
- работа с пользовательскими объектами: справочники, агрегаты, события, списки, обучающие выборки;
- ведение служебных справочников;
- ведение карточек событий;
- настройка политик, правил и построение прогнозных моделей;
- работа с моделями машинного обучения.
- испытание моделей;
- ведение документооборота в рабочем процессе;
- ведение справочников прав доступа;
- ведение учётных записей пользователей.

4.1 Просмотр статистики

Задачи решаются с помощью функций модуля **Рабочий стол**.

Реализованы следующие функции:

- просмотр статистики событий в графическом виде: круговые и линейные диаграммы;
- переход к списку событий с объектами.

4.2 Настройка модели данных

Задачи решаются с помощью функций модуля **Фабрика данных**.

Реализованы следующие функции:

- просмотр списка объектов;
- создание объекта и настройка его конфигурации;
- добавление, редактирование и удаление полей и атрибутов;
- установка соответствия между атрибутом и полем;
- удаление объекта.

4.3 Работа с пользовательскими объектами

Реализованы следующие функции:

- просмотр экземпляров пользовательского объекта;
- просмотр списка записей экземпляра пользовательского объекта;
- создание, редактирование и удаление записей.

4.4 Обработка и обогащение данных

- Реализованы следующие функции:
- Добавление, редактирование и просмотр в графическом виде ETL-процесса;
- Загрузка данных из систем-источников посредством встроенного ETL-движка.
- Пакетная обработка данных;
- Потокное получение данных;
- Преобразование данных в формат и структуры, определенные в модуле «Фабрика данных»;
- Настройка параметров запуска ETL-процесса по расписанию;
- Запуск и остановка ETL-процесса через интерфейс.

4.5 Ведение служебных справочников

Реализованы следующие функции:

- просмотр записей справочника списков или глобальных переменных;
- просмотр, редактирование и удаление записи.

4.6 Анализ событий

Задачи решаются с помощью функций модуля **Анализ событий**.

Реализованы следующие функции:

- обработка данных, которые поступают из внешних систем, посредством политик выявления;
- **создание** событий на основании результатов работы политик выявления;
- формирование карточек отслеживаемых событий;
- просмотр карточки события;
- работа с отслеживаемыми событиями;
- просмотр комментариев к событию;

- анализ события;
- выполнения действий над событием;
- просмотр списка связанных событий;
- просмотр связанных событий в графическом виде (график);
- просмотр атрибутов выбранного события.

4.7 Настройка механизмов выявления событий

Задачи решаются с помощью функций модуля **Правила анализа**.

Реализованы следующие функции:

- создание, редактирование и удаление правила или группы правил, политики и модели выявления;
- просмотр списка правил, политик и моделей выявления;
- настройка правил с использованием моделей машинного обучения;
- настройка модели выявления с помощью правил и политик;
- настройка матрицы срабатывания политики выявления;
- запуск и остановка выполнения политики выявления;
- просмотр списка объектов выполнения;
- создание, редактирование и удаление объектов выполнения;
- запуск объекта выполнения на проверку политики на хранимых данных.

4.8 Работа с моделями машинного обучения

Задачи решаются с помощью функций модуля **Машинное обучение**.

Реализованы следующие функции:

- Импорт файлов, содержащих данные по готовым моделям в формате PMML;
- Загрузка готовых моделей в виде контейнера, содержащего скрипт работы модели.
- Настройки математических моделей;
- Хранение, совместное использование, управление версиями моделей;
- Запуск моделей в промышленную эксплуатацию;
- Управления моделями;
- Испытание измененных моделей, до ввода их в промышленную эксплуатацию;
- Запуск и остановку обучения модели;
- Настройки действий, выполняемых по результатам анализа с использованием моделей;
- Формирование и сохранение результата анализа.

5 Входные и выходные данные

5.1 Входные данные

На вход **Jet Galatea** поступают следующие входные данные:

- данные систем-источников;
- файлы ETL-процессов, моделей выявления;
- справочные данные (типы, категории отслеживаемых событий, подразделения, должности);
- данные пользователей.

5.2 Выходные данные

Jet Galatea создает следующие выходные данные:

- события;
- объекты, модели данных;
- пользовательские объекты;
- правила, политики и модели выявления;
- статистика отслеживаемых событий в графическом виде.

Термины и сокращения

Термин/сокращение	Описание
CEP	Complex Event Processing, Комплексная обработка событий. В JG – Подсистема правил расчёта
ETL	Extract, Transform, Load – Извлечение, Преобразование, Загрузка – процессы обработки данных
JG	Программное обеспечение «Jet Galatea»
ML	Machine Learning, Машинное обучение
PMML	Predictive Model Markup Language. Язык разметки для прогнозного моделирования
АРМ	Автоматизированное рабочее место
ПО	Программное обеспечение
СУБД	Система управления базами данных