



Программное обеспечение «JET GALATEA»

ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК



Оглавление

1	Общие сведения	4
1.1	Обозначение и наименование	4
1.2	Инструментальные средства разработки	4
2	Назначение Jet Galatea	4
3	Условия применения	7
3.1	Конфигурация сервера приложений и управления кластером	7
3.2	Конфигурация сервера хранения данных и вычислений	7
3.3	Конфигурация АРМ	7
4	Описание задач	8
4.1	Просмотр статистики	8
4.2	Настройка модели данных	8
4.3	Работа с пользовательскими объектами	9
4.4	Обработка и обогащение данных	9
4.5	Ведение служебных справочников	9
4.6	Анализ событий	9
4.7	Настройка механизмов выявления событий	10
4.8	Ведение справочников доступа	10
4.9	Ведение учётных записей пользователей	11
4.10	Ведение документооборота	11
4.11	Работа с моделями машинного обучения	11
4.12	Формирование отчетности	12
5	Входные и выходные данные	13
5.1	Входные данные	13
5.2	Выходные данные	13



Аннотация

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

Краткое содержание 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**.

1.2 Инструментальные средства разработки

Для создания ПО использовались следующие технологии, платформы, средства разработки, языки программирования, СУБД и хранилища:

- MapReduce, Impala;
- Hadoop, Pentaho DI, Apache Kafka, JVM, Solr, Spark, YARN;
- Hue, OpenJDK;
- Scala, Python, XML, YAML;
- HBase, HDFS, Hive,
- объектно-реляционная система управления базами данных PostgreSQL;
- строго типизированный объектно-ориентированный язык программирования Java;
- универсальный фреймворк для Java-платформы Spring Boot 2;
- инструмент для сборки проекта maven;
- веб-сервис хостинга проектов и их совместной разработки Bitbucket;
- система непрерывной интеграции программного обеспечения Jenkins;
- система мониторинга Prometheus;
- графическая визуализация баз данных Grafana;
- поисковый движок: ELK (ElasticSearch, Logstash, Kibana).

2 Назначение Jet Galatea

Jet Galatea – программное решение для работы с большими потоками данных. Используется для 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.

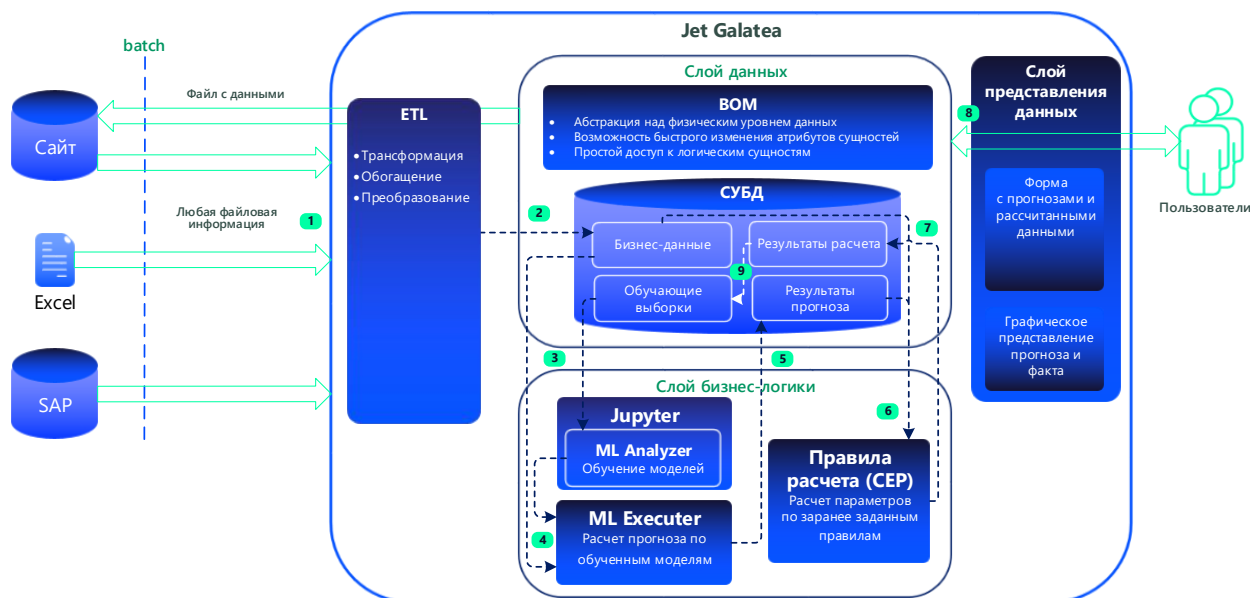
JG имеет трёхуровневую архитектуру:

- 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;
- Документооборот;
- Отчетность;
- 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 (Business Object Model);
- 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 **Документооборот** предназначен для настройки и управления моделями документооборота, статусными моделями событий и объектов и действиями над ними, а также связыванием статусов с действиями над объектами.

Модуль **Отчетность** предназначен для формирования отчетов, управления шаблонами отчетов, настройке расписания формирования и форматами выгрузок.

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 ≥ 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 ≥ 100 ГБ

Все серверы работают под управлением операционной системы Centos 7.6. 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 APM

Для управления работой ПО используется веб-интерфейс администратора и интерфейс командной строки (CLI).

Компьютер администратора должен соответствовать характеристикам:

- процессор – не менее 4 ядер с тактовой частотой не менее 2,8 ГГц;
- объем оперативной памяти – не менее 4 ГБ;
- объем 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;
- перевод события в различные статусы в соответствии со статусной моделью (модуль «Документооборот»);
- 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 простых и агрегативных правил;
- 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;
- сравнение политик выявления;
- просмотр результатов сравнения политик;
- просмотр списка объектов выполнения;
- создание, редактирование и удаление объектов выполнения;
- запуск объекта выполнения на проверку политики на хранимых данных.

4.8 Ведение справочников доступа

Задачи решаются с помощью функций модуля **Авторизация**.

Реализованы следующие функции:

Инфосистемы Джет

- просмотр, создание, редактирование и удаление узлов дерева разрешений;
- просмотр, создание, редактирование и удаление узлов дерева владений;
- просмотр, создание, редактирование и удаление узлов дерева ролей.

4.9 Ведение учётных записей пользователей

Задачи решаются с помощью функций модуля **Авторизация**.

Реализованы следующие функции:

- просмотр списка учётных записей пользователей;
- создание, редактирование и удаление записи;
- настройка прав доступа;
- блокировка и разблокировка учетной записи;
- настройка доступа к записям в зависимости от присвоенных пользователям ролей и прав доступа.

4.10 Ведение документооборота

Задачи решаются с помощью функций модуля **Документооборот**.

Реализованы следующие функции:

- создание модели документооборота;
- настройка управления процессом документирования;
- связывание статусной модели с выполняемыми действиями в системе;
- использование моделей документооборота при работе со списком событий.

4.11 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.
- Создание модели выявления;
- Настройки математических моделей;
- Создание массивов данных для обучения;
- Обучение модели выявления на основании обучающих выборок;
- Хранение, совместное использование, управление версиями моделей;
- Запуск моделей в промышленную эксплуатацию;

- Принятие решений, которые базируются на моделях машинного обучения;
- Управления моделями;
- Работа с моделями в режиме редактирования непосредственно из интерфейса.
- Испытание измененных моделей, до ввода их в промышленную эксплуатацию;
- Просмотра модели в графическом виде в виде «дерева решений»;
- Запуск моделей в промышленную эксплуатацию;
- Запуск и остановку обучения модели;
- Настройки действий, выполняемых по результатам анализа с использованием моделей;
- Формирование и сохранение результата анализа.

4.12 Формирование отчетности

Задачи решаются с помощью функций модуля **Отчетность**.

Реализованы следующие функции:

- Загрузка шаблонов отчетов;
- Формирование отчетов по шаблону;
- Выгрузка отчета в файлы формата PDF, CSV, XLS, RTF, DOCX
- Настройка формирования отчета по расписанию;
- Отправка отчета по электронной почте в автоматическом режиме

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;
- обучающие выборки;
- видеопоток данных;
- графические изображения.

5.2 Выходные данные

Jet Galatea создает следующие выходные данные:

- зарегистрированные в БД;
- события;
- объекты, модели данных;
- пользовательские объекты;
- правила, политики и модели выявления;
- роли, разрешения, владения и учётные записи пользователей;
- статистика отслеживаемых событий в графическом виде;
- результаты сравнения применяемых политик.

Термины и сокращения

Термин/сокращение	Описание
BOM	Business Object Model, Модель бизнес-объекта. В JG – Подсистема управления данными
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