

УТВЕРЖДАЮ

Генеральный директор

ООО «НПО «Криста»

_____ Ю. А. Черных

_____.____.20____

**Программный комплекс
Информационно-аналитическая платформа Криста VI
Информационно-аналитическая платформа Krista VI**

Руководство системного программиста

ЛИСТ УТВЕРЖДЕНИЯ

НПОК.00000.ИАПВИ.00.32.1–20180329–1–ЛУ

СОГЛАСОВАНО

Заместитель генерального директора

по производству

_____ Е. Н. Черных

29.03.2018

Представители предприятия-разработчика

Главный конструктор

_____ М. В. Баркова

29.03.2018

Руководитель, ответственный за
разработку документации

_____ С. А. Лапина

29.03.2018

Исполнитель

_____ Н. А. Погодин

29.03.2018

Нормоконтролер

_____ Т.А. Рытова

29.03.2018

2018

УТВЕРЖДЕНО

НПОК.00000. ИАПВІ.00.32.1–20180329–1–ЛУ

Программный комплекс
Информационно-аналитическая платформа Криста ВІ
Информационно-аналитическая платформа Krista ВІ

Руководство системного программиста
НПОК.00000.ИАПВІ.00.32.1–20180329–1

Листов 21

Аннотация

Данный документ представляет собой руководство системного программиста Программного комплекса Информационно-аналитическая платформа Krista ВІ (далее – ПК «Платформа Криста ВІ»). В документе описан порядок установки и настройки ПК «Платформа Криста ВІ» на начало работы, приведены технические требования к аппаратному и системному программному обеспечению.

ООО «НПО «Криста» оставляет за собой право вносить изменения в программное обеспечение без внесения изменений в документацию. Изменения программного обеспечения при выпуске новых версий отражается в сопроводительной документации к версии.

ООО «НПО «Криста» оставляет за собой право вносить поправки и (или) изменения в документацию без прямого или косвенного обязательства уведомлять кого-либо о таких правках и изменениях.

Содержание

| | |
|--|----|
| Перечень рисунков..... | 4 |
| Перечень таблиц..... | 5 |
| Перечень терминов и сокращений..... | 6 |
| 1 Общие сведения о программном комплексе | 7 |
| 1.1 Назначение программного комплекса | 7 |
| 1.2 Перечень функций, выполняемых платформой | 7 |
| 1.3 Требования к программно-аппаратному обеспечению..... | 7 |
| 2 Структура программы..... | 8 |
| 2.1 Слой интеграции данных | 11 |
| 2.2 Слой хранения данных | 12 |
| 2.3 Слой метаданных..... | 14 |
| 2.4 Презентационный слой | 15 |
| 2.5 Сервисы управления доступом..... | 16 |
| 2.6 Модуль машинного обучения..... | 16 |
| 3 Настройка программы | 17 |
| 3.1 Установка ПК «Платформа Криста ВІ» под Windows x64 | 17 |
| 3.2 Установка ПК «Платформа Криста ВІ» под linux..... | 19 |
| 4 Проверка программы | 20 |
| 5 Дополнительные возможности | 21 |
| 6 Сообщения системному программисту | 22 |

Перечень рисунков

| | |
|--|----|
| Рисунок 1 - Архитектура ПК «Платформа Криста ВІ»..... | 9 |
| Рисунок 2 - Архитектура ПК «Платформа Криста ВІ»– схема потоков данных. | 10 |
| Рисунок 3 - Архитектура DataLake | 14 |

Перечень таблиц

| | |
|---|----|
| Таблица 1 – Термины и сокращения | 6 |
| Таблица 2 - Программные компоненты Krista ВІ..... | 11 |

Перечень терминов и сокращений

В настоящем документе применены следующие термины и сокращения с соответствующими определениями, представленные в таблице 1.

Таблица 1 – Термины и сокращения

| Термин (сокращение) | Определение |
|----------------------------|--|
| Business intelligence (BI) | Процесс получения необходимой информации из массива разнородных исходных данных с использованием информационных технологий (методов и средств) сбора, консолидации и обработки данных, а также средств обеспечения доступа пользователей к полученной информации, в целях анализа и принятия решений. |
| CSV | Текстовый формат представления табличных данных (comma-separated values). |
| ETL | Extract, Transform, Load - один из основных процессов в управлении хранилищами данных, который включает в себя: <ul style="list-style-type: none"> – прием данных из внешних источников; – преобразование и очистку данных для обеспечения соответствия нуждам бизнес-модели; – загрузку данных в хранилище данных. |
| KPI | Ключевые показатели эффективности. |
| ODS | OpenDocument Spreadsheet – это электронные таблицы формата OpenDocument, созданные в программах типа OpenOffice. |
| OLAP | Online analytical processing - аналитическая обработка в реальном времени - технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу. Реализации технологии OLAP являются компонентами программных решений класса Business Intelligence. |
| SOAP | Simple Object Access Protocol — простой протокол доступа к объектам — протокол обмена структурированными сообщениями в распределённой вычислительной среде. Протокол используется для обмена произвольными сообщениями в формате XML между программными компонентами и различными программными комплексами. SOAP может использоваться с любым протоколом прикладного уровня: SMTP, FTP, HTTP, HTTPS и другими. |
| XML | Расширяемый язык разметки (XML Extensible Markup Language) - текстовый формат, предназначенный для хранения структурированных данных и обмена информацией между программами. |
| ПК | Программный комплекс. |
| СУБД | Система управления базами данных. |

1 Общие сведения о программном комплексе

1.1 Назначение программного комплекса

ПК «Платформа Криста ВІ» – инструмент построения информационно-аналитических систем, решения задач бизнес-аналитики и формирования отчетности для принятия решений. ПК «Платформа Криста ВІ» предназначен для сферы управления общественными финансами, социально-экономического развития и сферы управления бизнесом.

1.2 Перечень функций, выполняемых платформой

Основные функции платформы:

- интеграция и взаимодействие с внешними источниками данных;
- хранение и анализ данных;
- предоставление результатов анализа.

Платформа поддерживает полный спектр аналитических возможностей, и за счет простоты использования не требует участия IT-специалистов для подготовки аналитических моделей и сценариев загрузки данных, что позволяет быстро и эффективно решать поставленные задачи.

1.3 Требования к программно-аппаратному обеспечению

Минимальные требования для локальной работы:

- процессор с рабочей частотой не менее 1500 МГц;
- 8 Гб оперативной памяти;
- 10 Гб места на диске.

Операционная система:

- Windows 7 и выше;
- Linux (Red Hat Linux/CentOS/Oracle Linux, Debian/Ubuntu или другие дистрибутивы).

Необходимое программное обеспечение:

- Java Development Kit (JDK) 8;
- сервер PostgreSQL 9.6;
- браузер актуальной версии Mozilla Firefox/Google Chrome/Яндекс.

2 Структура программы

Логическая архитектура ПК «Платформа Криста ВІ» разделена на слои, каждый слой выполняет отдельную функцию:

- слой интеграции данных;
- слой хранения данных;
- слой метаданных;
- презентационный слой;
- средства управления доступом;
- модуль машинного обучения.

Архитектура платформы является масштабируемой и отказоустойчивой. Архитектура ПК «Платформа Криста ВІ» приведена на рисунке 1.

Связи между составными частями ПК «Платформа Криста ВІ» отражены на рисунке 2.

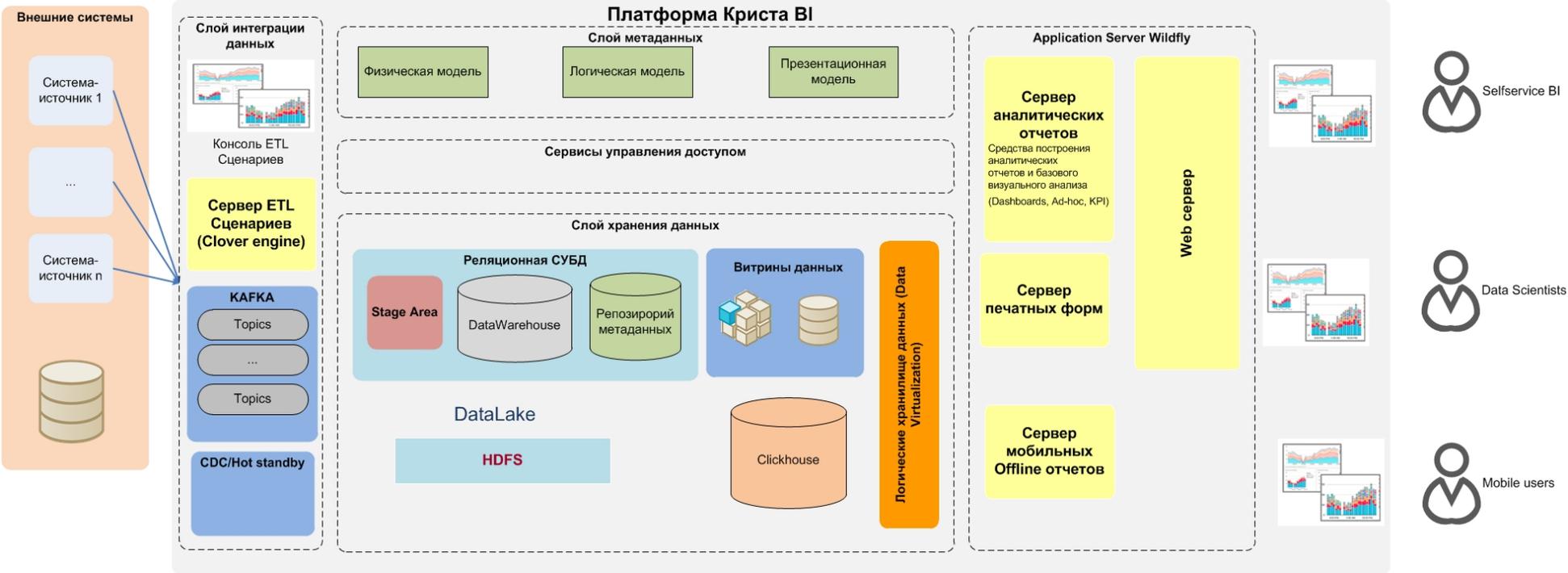


Рисунок 1 - Архитектура ПК «Платформа Криста ВІ»

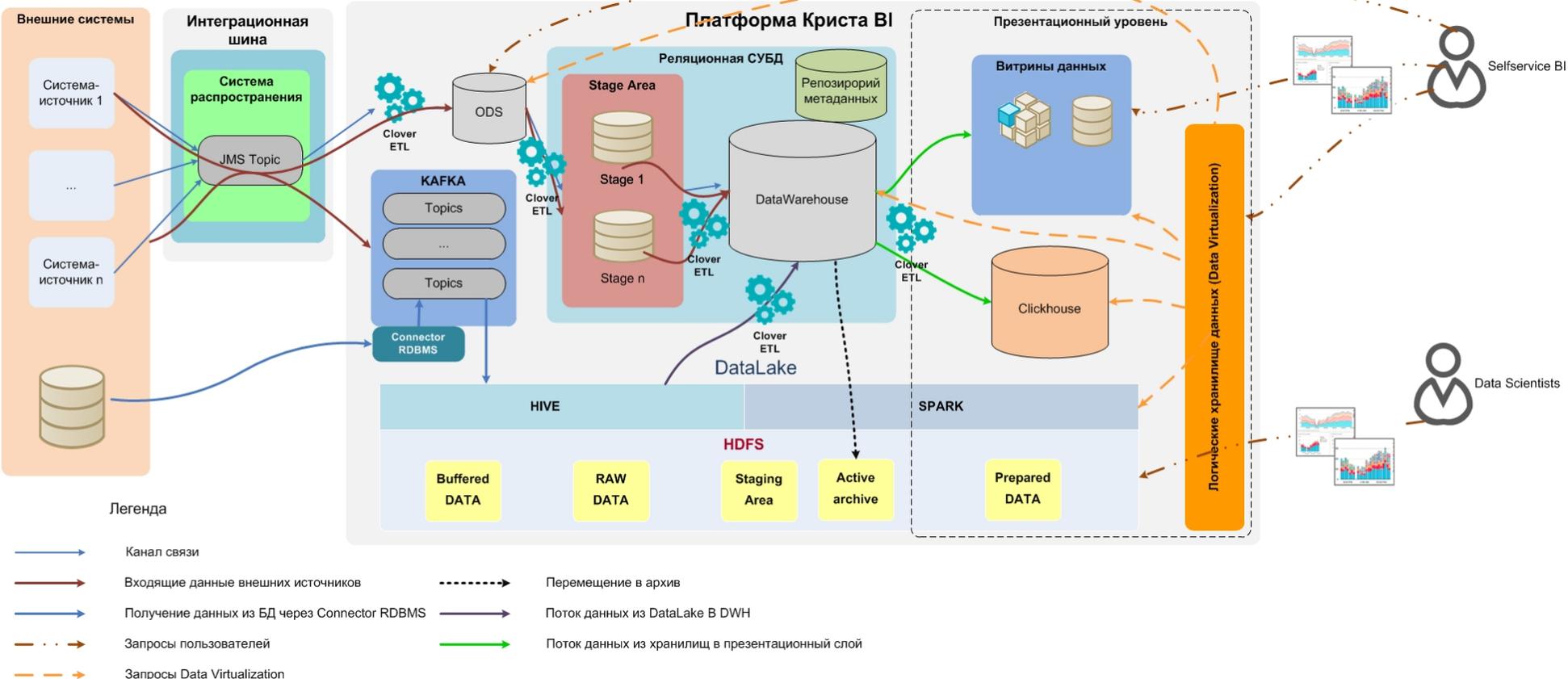


Рисунок 2 - Архитектура ПК «Платформа Криста ВІ»– схема потоков данных.

ПК «Платформа Криста ВІ» включает в себя программные компоненты, которые разворачиваются на сервере приложений Java EE. Компоненты представлены в таблице 2.

Таблица 2 - Программные компоненты Krista ВІ

| Имя программного компонента | Назначение |
|------------------------------------|--|
| bi-server-static.war | Библиотека статического контента |
| bi-server.ear | Ядровой компонент платформы |
| delivery-spammer.war | Модуль, отвечающий за рассылки |
| essbase-datasource-connector.rar | Модуль подключения к источникам данных |
| mdxprinter.war | Модуль, выполняющий выгрузку аналитических отчетов |
| olap4j-datasource-connector.rar | Модуль подключения к источникам данных |
| printer.war | Модуль, выполняющий выгрузку аналитических отчетов |
| publisher-agent-essbase.war | Модуль взаимодействия с OLAP, обновление данных |

2.1 Слой интеграции данных

Слой интеграции обеспечивает взаимодействие с внешними информационными системами (информационные системы органов государственной власти, региональные и муниципальные информационные системы и др.), возможность подключения любых источников данных в различных форматах: Excel, CSV, XML, ODS, получение данных напрямую из реляционных и многомерных баз данных, API-интерфейсов, открытых данных, подключение через SOAP-протокол к веб-сервисам приема данных.

Слой интеграции данных включает в себя:

- ETL и ELT сценарии приема, преобразования и загрузки данных;
- консоль разработчика, обеспечивающую настройку ETL и ELT сценариев;
- интеграционную шину и JMS-брокер для управления потоками данных, очередностью их поступления, например: Apache, Kafka;
- программное обеспечение, обеспечивающее настройку и исполнение ETL и ELT сценариев.

2.2 Слой хранения данных

Слой хранения данных предназначен для консолидации и хранения структурированных данных в хранилище (DataWarehouse).

Слой хранения данных включает в себя:

а) реляционную базу данных;

б) витрины данных:

1) многомерные (OLAP);

2) реляционные.

в) логическое хранилище данных (Data Virtualization) – средство доступа к данным на основе логической метамодел;

г) область DataLake;

д) репозиторий метаданных.

Для обработки больших объемов данных (BigData) ПК «Платформа Криста ВІ» использует технологию DataLake представленную на рисунке 3. В рамках ПК «Платформа Криста ВІ» технология DataLake строится на базе технологий Hadoop, Spark и Hive. Технология Hadoop представляет собой программный фреймворк, позволяющий хранить и обрабатывать данные с помощью компьютерных кластеров, используя парадигму MapReduce. В основе технологии лежит распределённая файловая система HDFS (Hadoop Distributed File System), которая обеспечивает хранение данных Hadoop сразу на нескольких узлах кластера. Таким образом, если один или несколько узлов кластера выходят из строя, то риск потери информации сводится к минимуму и кластер продолжает работу в штатном режиме.

В архитектуре DataLake ПК «Платформа Криста ВІ» выделяются следующие концептуальные слои данных:

- временное хранилище неподготовленных данных (Buffered DATA) – слой «сырых» данных, в который размещаются файлы в том виде, в котором они получены из внешних источников без какой-либо обработки. Является промежуточным слоем;

- постоянное хранилище сырых данных (RAW DATA) – слой данных, приведенных в более удобный для обработки формат. В данном слое исключены данные, которые не будут использованы в рамках пользовательских сценариев. Обеспечивается постоянное хранение данных – может служить источником пользовательских данных для опытных специалистов по обработке и анализу больших массивов данных (Data Scientists);
- Staging Area – временный слой хранения данных для последующей их обработки. Данные здесь преобразованы и подготовлены для загрузки их в слой подготовленных данных (Prepared DATA) или напрямую в хранилище (DataWarehouse);
- подготовленные данные (Prepared DATA) – консолидированные, обработанные и очищенные данные, готовые к загрузке в хранилища данных, к формированию витрин данных. Доступны для аналитики с самообслуживанием;
- пользовательские данные («песочница», UserData) – слой данных, доступных и подготовленных пользователями различного уровня компетенции в рамках своих задач анализа и обработки данных, построения собственных срезов данных. Может служить источником данных для слоя Prepared DATA;
- архив (Active archive) – слой данных, которые более неактуальны для основного хранилища данных (DataWarehouse) и витрин данных, но могут быть полезны, например, при ретроспективном анализе.

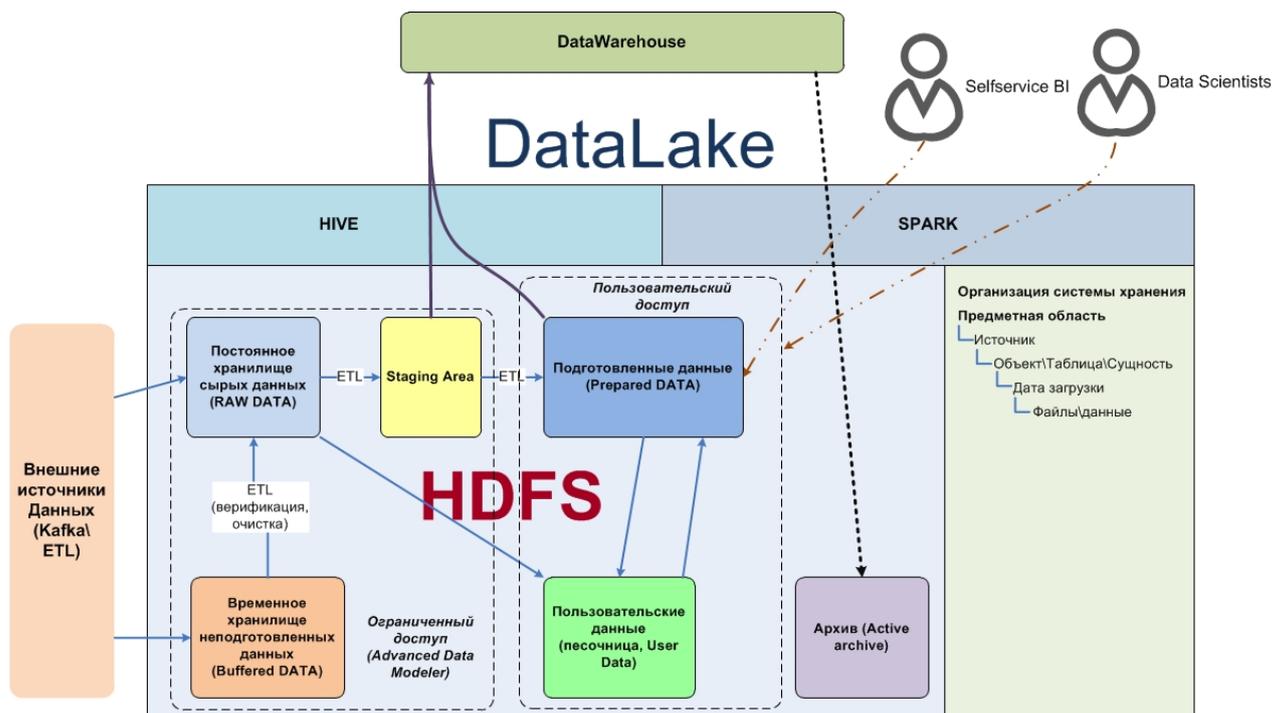


Рисунок 3 - Архитектура DataLake

2.3 Слой метаданных

Слой метаданных включает три основных элемента:

- физическая модель. Описывает физическую структуру данных в хранилищах данных и порядок доступа;
- логическая модель. Описывает логику, структуру и взаимосвязи данных в хранилище данных в терминах мер и измерений;
- презентационная модель. Является производной от логической модели, описывает перечень витрин, атрибутов, доступных конечным пользователям. Презентационная модель переводит информацию в термины предметной области, в понятия мер и изменений. На основе одной логической модели может быть построено несколько презентационных;
- экспертную систему и персонального помощника.

Персональный помощник в составе ПК «Платформа Криста VI» выполняет поиск в базе знаний, реляционной и многомерной базах данных. Помощник использует нейросети и инструменты машинного обучения, понимает голосовой или текстовый вопрос, ответ также может написать или проговорить. Ответ может включать не только текст, но и документы, изображения, ссылки на интерфейсы, сайты, кубы или отчеты. Использование интерактивного персонального помощника позволяет проводить эффективный поиск, учитывать семантические связи и сокращать время обработки и формирования аналитической информации. Персональный помощник работает как чат-бот, на портале и в мобильном приложении.

2.4 Презентационный слой

Презентационный слой обеспечивает доступ пользователей к интерфейсам визуального анализа данных, в том числе применение следующих технологий анализа:

- интерактивная визуализация;
- произвольный анализ с использованием не predetermined заранее запросов (Ad-hoc анализ, OLAP-анализ);
- информационные панели (Dashboard);
- ключевые показатели эффективности KPI, Scorecards (карты показателей), индикация, позволяющая определить «точки внимания», сигналы, в том числе с возможностью перехода на детальные данные;
- витрины данных (Datamart);
- интеллектуальный анализ данных (Data Mining);
- сопоставительный анализ на основе эталонных показателей (Benchmarking);
- Search-based (поисковый интерфейс);
- аналитика с самообслуживанием.
- Презентационный слой включает в себя:

- сервер аналитических отчетов (средства построения аналитических отчетов и базового визуального анализа);
- сервер печатных форм;
- сервер мобильных отчетов;
- Web-сервер.

2.5 Сервисы управления доступом

Сервисы управления доступом обеспечивают аутентификацию, авторизацию пользователей, единую точку входа, интеграцию со сторонними решениями авторизации и аутентификации.

2.6 Модуль машинного обучения

В рамках модуля машинного обучения реализуются следующие модели:

- Логистическая регрессия. Статистический метод, позволяющий предсказать вероятность возникновения некоторого события по значениям множества признаков. Вероятность всегда лежит в диапазоне от 0 до 1. Этот метод применяется для решения задач классификации;
- Решающие деревья. Используемый в статистике и анализе данных метод, в основе которого лежит построение прогнозной модели в виде иерархической, последовательной структуры, где каждому объекту соответствует единственный узел, дающий решение («если ..., то ...»). Цель метода - создать модель, предсказывающую значение целевой переменной на основе нескольких входных. Используются для описания и классификации данных, регрессионного анализа;

- Градиентный бустинг. Метод решения задач повышенной сложности. Метод использует более простые алгоритмы (например, решающие деревья, искусственная нейронная сеть), дополняя их новыми данными, и за счет этого происходит усложнение первоначальной модели, учитываются дополнительные условия и, как результат, повышается качество прогнозирования;
- ИНС (Искусственная нейронная сеть). Математическая модель и ее программная реализации, построенная по принципу сетей нервных клеток живого организма. Используются для решения задач анализа данных, прогнозирования, аппроксимации, при принятии решений и диагностики, оценке рисков;
- ИНС (Искусственная нейронная сеть). Математическая модель и ее программная реализации, построенная по принципу сетей нервных клеток живого организма. Используются для решения задач анализа данных, прогнозирования, аппроксимации, при принятии решений и диагностики, оценке рисков.

3 Настройка программы

3.1 Установка ПК «Платформа Криста ВІ» под Windows x64

Для установки ПК «Платформа Криста ВІ» под Windows x64 необходимо выполнить следующие действия:

- установить JDK 8
(<https://www.oracle.com/technetwork/java/javase/downloads>);
- установить PostgreSQL версии 9.6 (например <https://oscg-downloads.s3.amazonaws.com/packages/PostgreSQL-9.6.10-1-win64-bigsq1.exe>). При установке необходимо задать пароль пользователя postgres;
- распаковать архив KristaBI.zip в папку c:\KristaBI;

- задать путь к папке JDK в файле *c:\KristaBI\wildfly-8.2.1.Final\wf.conf.javahome.bat*, например: *set JAVA_HOME=c:\KristaBI\jdk1.8.0_181;*

- в скрипте *c:\KristaBI\database\create_bi_db.bat* настроить параметры подключения к базе PostgreSQL:

```
set pg_home=c:\PostgreSQL\pg96
```

```
set PGUSER=postgres
```

```
set PGPASSWORD=postgres
```

```
set PGHOST=localhost
```

```
set PGPORT=5432
```

при установке по умолчанию требуется изменить только пароль пользователя postgres PGPASSWORD;

- запустить скрипт: *c:\KristaBI\database\create_bi_db.bat;*
- скрипт создаёт пользователей и базы данных bimeta, bidata;
- выполнить скрипт для создания службы Wildfly KBI: *c:\KristaBI\wildfly-8.2.1.Final\service.bat install;*
- настроить тип запуска службы Wildfly KBI на «Автоматически» и запустить её;
- для подключения к portalу открыть в браузере адрес <http://localhost:8080/static-report/web/portal.html>, для входа использовать имя пользователя и пароль demo/demo;
- имена и пароли пользователей задаются в файле «*c:\KristaBI\wildfly-8.2.1.Final\standalone\configuration\finmon-users.properties*».
- Роли пользователям назначаются в файле «*c:\KristaBI\wildfly-8.2.1.Final\standalone\configuration\finmon-roles.properties*».
- Настройки Wildfly хранятся в файле «*c:\KristaBI\wildfly-8.2.1.Final\standalone\configuration\standalone-full.xml*».

3.2 Установка ПК «Платформа Криста ВІ» под linux

Все настройки выполняются под пользователем root. Для установки ПК «Платформа Криста ВІ» под linux (на примере Oracle Linux 7) необходимо выполнить следующие действия:

- установить JDK 8: `yum install java-1.8.0-openjdk-devel;`
- установить сервер PostgreSQL версии 9.6:
`yum install https://download.postgresql.org/pub/repos/yum/9.6/redhat/rhel-7-x86_64/pgdg-oraclelinux96-9.6-3.noarch.rpm,`
`yum install postgresql96-server/usr/pgsql-9.6/bin/postgresql96-setup initdb;`
- настроить доступ к базе postgres по паролю. Добавить в файл `/var/lib/pgsql/data/pg_hba.conf` строку:
`host all all 0.0.0.0/0 md5;`
- строка должна стоять выше остальных записей:
`host - host all all 127.0.0.1/32 ident u host all all::1/128 ident;`
- настроить автозапуск и запустить службу postgres:
`systemctl enable postgresql-9.6 u systemctl start postgresql-9.6;`
- распаковать архив «KristaBI.zip» в папку `/opt/KristaBI;`
- запустить скрипт `create_bi_db.sh`:
`chmod +x /opt/KristaBI/database/create_bi_db.sh`
`su - postgres -c /opt/KristaBI/database/create_bi_db.sh;`
 скрипт создаёт пользователей и базы данных «bimeta», «bidata»
- задать параметры запуска сервера Wildfly в файле `/opt/KristaBI/wildfly-8.2.1.Final/wildfly.conf`, например:
`export JAVA_HOME="/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-4.b14.el7.x86_64" u export JBOSS_USER=wildfly;`
- запустить скрипт `/opt/KristaBI/init/create_bi_user.sh` для создания пользователя и настройки прав на папку `/opt/KristaBI;`

- запустить скрипт `/opt/KristaBI/init/create_bi_service.sh` для настройки и запуска службы `wildfly`;
- запуск сервера выполняется командой:
`service wildfly start;`
- для подключения к portalу открыть в браузере адрес <http://localhost:8080/static-report/web/portal.html>, для входа использовать имя пользователя и пароль `demo/demo`;
- имена и пароли пользователей задаются в файле `/opt/KristaBI/wildfly-8.2.1.Final/standalone/configuration/finmon-users.properties`.
Роли пользователям назначаются в файле `/opt/KristaBI/wildfly-8.2.1.Final/standalone/configuration/finmon-roles.properties`.
Настройки Wildfly хранятся в файле `/opt/KristaBI/wildfly-8.2.1.Final/standalone/configuration/standalone-full.xml`.

4 Проверка программы

Проверка функционирования платформы Krista BI сводится к диагностике компонентов базисного программного обеспечения:

- диагностика реляционных баз данных;
- диагностика многомерных баз данных;
- диагностика платформы Java EE;
- диагностика платформы ETL.

Для того, чтобы дать заключение о работоспособности платформы Krista BI, необходимо:

- а) проверить доступность реляционной базы данных, созданной в п. 3:
 - 1) подключиться к базе данных под пользователем «bimeta»;
 - 2) выполнить запрос «select '1'»;
 - 3) подключиться к базе данных под пользователем «bidata»;

- 4) выполнить запрос «select '1'».

При возникновении ошибок во время подключения или выполнения запроса, требуется обратиться к официальной документации по СУБД, устранить ошибки.

Коды возможных сообщений СУБД можно посмотреть на официальном ресурсе: <https://www.postgresql.org/docs/9.4/static/errcodes-appendix.html>

б) проверить доступность многомерных баз данных. Способы проверки доступности многомерных баз данных будут отличаться в зависимости от выбранной СУБД. Ниже приведены ссылки на официальную документацию по поддерживаемым многомерным СУБД:

- 1) https://docs.oracle.com/cd/E12825_01/epm.111/esb_dbag.pdf;
- 2) <https://mondrian.pentaho.com/documentation/>;
- 3) <https://docs.microsoft.com/ru-ru/sql/analysis-services/analysis-services>;
- 4) <https://clickhouse.yandex/docs>.

в) проверить корректность запуска программных компонент платформы Krista BI. Требуется убедиться, что при запуске программных компонент Krista BI, перечисленных в таблице выше (Таблица 2) не возникло ошибок. Способ проверки зависит от выбранной платформы Java EE. Ниже приведены ссылки на официальную документацию поддерживаемых платформ Java EE:

- 1) <https://docs.jboss.org/author/display/WFLY8/Admin+Guide>;
- 2) <https://docs.oracle.com/en/middleware/middleware.html>.

г) проверить доступность ETL инструментов. Официальные ссылки на поддерживаемые платформой Krista BI ETL инструменты приведены ниже:

- 1) <https://www.cloveretl.com/resources>;
- 2) <https://docs.oracle.com/middleware/12211/odi/docs.htm>.

5 Дополнительные возможности

Дополнительных возможностей не предусмотрено.

6 Сообщения системному программисту

В ходе настройки, проверки и выполнения программы могут выдаваться сообщения. Сообщения могут содержаться в интерфейсе пользователя и в лог-файлах, которые ведутся базисным программным обеспечением. При выводе сообщений в интерфейс пользователя, ПК «Платформа Криста ВІ» перехватывает текст сообщений от базисного программного обеспечения. Таким образом, диагностика сообщений должна проводиться в соответствии с официальной документацией базисного программного обеспечения. Ниже представлен список ресурсов с возможными сообщениями ПК «Платформа Криста ВІ»:

- <https://www.postgresql.org/docs/9.4/static/errcodes-appendix.html>;
- https://docs.oracle.com/cd/E57185_01/ESBEM/launch.html;
- https://docs.oracle.com/cd/E28280_01/core.1111/e10113/chapter_odi_messages.htm#FMERR170;
- [https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2005/ms345164\(v=sql.90\)](https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2005/ms345164(v=sql.90)).