|  |  |
| --- | --- |
|  |  |

**RT.DataVision**

**Инструкция по установке**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

**Оглавление**

[1 Назначение документа 3](#_Toc165046020)

[2 Общие положения 3](#_Toc165046021)

[2.1 Наименование и условное обозначение 3](#_Toc165046022)

[2.2 Область применения 3](#_Toc165046023)

[2.3 Термины, определения и сокращения 3](#_Toc165046024)

[3 Состав программного обеспечения 4](#_Toc165046025)

[4 Программно-аппаратные требования 5](#_Toc165046026)

[4.1 RT.DataVision 5](#_Toc165046027)

[4.1.1 Системные требования к серверному оборудованию 5](#_Toc165046028)

[4.1.2 Системные требования к клиентскому оборудованию 5](#_Toc165046029)

[4.1.3 Аппаратные требования 5](#_Toc165046030)

[4.1.3.1 Серверное оборудование 5](#_Toc165046031)

[4.1.3.2 Клиентское оборудование 5](#_Toc165046032)

[4.2 Dremio 5](#_Toc165046033)

[4.2.1 Системные требования к серверному оборудованию 5](#_Toc165046034)

[4.2.2 Системные требования к клиентскому оборудованию 5](#_Toc165046035)

[5 Установка RT.DataVision 6](#_Toc165046036)

[5.1 Виртуальная среда Python 6](#_Toc165046037)

[5.2 Установка и инициализация RT.DataVision 6](#_Toc165046038)

[6 Установка Dremio 7](#_Toc165046039)

[7 Настройка RT.DataVision 7](#_Toc165046040)

# Назначение документа

Настоящий документ включает описание по установке системы «RT.DataVision».

# Общие положения

## Наименование и условное обозначение

Наименование системы: RT.DataVision.

## Область применения

RT.DataVision состоит из двух компонентов:

1. RT.DataVision – это современное корпоративное веб-приложение для бизнес-аналитики. Это быстрая, легкая, интуитивно понятная платформа, позволяющая пользователям с любым набором навыков легко исследовать и визуализировать свои данные, от простых круговых диаграмм до высокодетализированных геопространственных диаграмм deck.gl.
2. Dremio, который обеспечивает аналитику самообслуживания для групп обработки данных. Аналитики данных могут исследовать и визуализировать данные с временем ответа на запросы менее секунды, а инженеры данных могут принимать и преобразовывать данные непосредственно в озере данных с полной поддержкой операций DML. Кроме того, аналитики могут объединять данные в озере с данными во внешних базах данных, поэтому им не нужно перемещать данные в объектное хранилище, чтобы извлечь пользу из этих данных. Открытая платформа Dremio Lakehouse, основанная на таких общественных стандартах, как Apache Iceberg и Apache Arrow, позволяет организациям использовать лучшие в своем классе механизмы обработки и устраняет привязку к поставщику.

## Термины, определения и сокращения

В настоящем документе использованы и определены следующие термины и сокращения:

| *Термин/Сокращение* | *Определение* |
| --- | --- |
| Apache | Apache HTTP-сервер – [свободный](https://ru.wikipedia.org/wiki/%D0%A1%D0%B2%D0%BE%D0%B1%D0%BE%D0%B4%D0%BD%D0%BE%D0%B5_%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D0%B5%D1%81%D0%BF%D0%B5%D1%87%D0%B5%D0%BD%D0%B8%D0%B5) [веб-сервер](https://ru.wikipedia.org/wiki/%D0%92%D0%B5%D0%B1-%D1%81%D0%B5%D1%80%D0%B2%D0%B5%D1%80).Apache является [кроссплатформенным ПО](https://ru.wikipedia.org/wiki/%D0%9A%D1%80%D0%BE%D1%81%D1%81%D0%BF%D0%BB%D0%B0%D1%82%D1%84%D0%BE%D1%80%D0%BC%D0%B5%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D0%B5%D1%81%D0%BF%D0%B5%D1%87%D0%B5%D0%BD%D0%B8%D0%B5), поддерживает операционные системы [Linux](https://ru.wikipedia.org/wiki/Linux), [BSD](https://ru.wikipedia.org/wiki/BSD), [Mac OS](https://ru.wikipedia.org/wiki/Mac_OS), [Microsoft Windows](https://ru.wikipedia.org/wiki/Microsoft_Windows), [Novell NetWare](https://ru.wikipedia.org/wiki/Novell_NetWare), [BeOS](https://ru.wikipedia.org/wiki/BeOS).Основными достоинствами Apache считаются надёжность и гибкость конфигурации. Он позволяет подключать внешние модули для предоставления данных, использовать [СУБД](https://ru.wikipedia.org/wiki/%D0%A1%D0%A3%D0%91%D0%94) для [аутентификации](https://ru.wikipedia.org/wiki/%D0%90%D1%83%D1%82%D0%B5%D0%BD%D1%82%D0%B8%D1%84%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F) пользователей, модифицировать сообщения об ошибках и т. д. Поддерживает [IPv4](https://ru.wikipedia.org/wiki/IPv4). |
| deck.gl | Фреймворк для визуализации больших данных на основе WebGL, разработанный uber с открытым исходным кодом. Он имеет характеристики предоставления различных типов слоев визуализации, рендеринга графического процессора, интеграции React и Mapbox GL в сочетании с географическими информационными данными (GPS) и специально используется для крупномасштабного исследования и визуализации баз данных. |
| Gunicorn | [Python](https://ru.bmstu.wiki/Python) [WSGI](https://ru.bmstu.wiki/WSGI_%28Web_Server_Gateway_Interface%29) HTTP сервер. Gunicorn является PreFork моделью, портированой с [Ruby](https://ru.bmstu.wiki/Ruby_%28%D1%8F%D0%B7%D1%8B%D0%BA_%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D1%8F%29) Unicorn проекта. Сервер Gunicorn совместим с рядом веб-фреймворков, просто реализован, не требует много ресурсов сервера и достаточно быстр. |
| Nginx | Программное обеспечение с открытым исходным кодом, которое позволяет создавать веб-сервер. Также его используют как почтовый сервер или обратный прокси-сервер. |
| Python | Высокоуровневый язык программирования общего назначения с [динамической](https://ru.wikipedia.org/wiki/%D0%94%D0%B8%D0%BD%D0%B0%D0%BC%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%82%D0%B8%D0%BF%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F) [строгой](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D1%80%D0%BE%D0%B3%D0%B0%D1%8F_%D1%82%D0%B8%D0%BF%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F) типизацией и автоматическим управлением памятью, ориентированный на повышение производительности разработчика, читаемости кода и его качества, а также на обеспечение переносимости написанных на нём программ. Язык является полностью объектно-ориентированным в том плане, что всё является объектами. |
| SQL | (англ. structured query language, язык структурированных запросов) –декларативный язык программирования, применяемый для создания, модификации и управления данными в реляционной базе данных, управляемой соответствующей [системой управления базами данных](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_%D1%83%D0%BF%D1%80%D0%B0%D0%B2%D0%BB%D0%B5%D0%BD%D0%B8%D1%8F_%D0%B1%D0%B0%D0%B7%D0%B0%D0%BC%D0%B8_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85). |
| Система | Система «RT.DataVision». |
| СУБД | Система управления базами данных. |

# Состав программного обеспечения

В состав программного обеспечения RT.DataVision входит непосредственно исходный код данного приложения, который в процессе установки, в зависимости от используемой операционной системы претерпевает различные изменения.

Для целей связующего компонента между БД, представляющими собой озёра данных, и компонентом RT.DataVision в состав ПО входит дистрибутив Dremio.

Кроме того, в процессе установки RT.DataVision, устанавливается заданная конфигурация Python.

Кроме этого, для работы используется веб-сервер Gunicorn, Nginx, Apache.

# Программно-аппаратные требования

## RT.DataVision

### Системные требования к серверному оборудованию

RT.DataVision устанавливается на операционных системах Linux.

### Системные требования к клиентскому оборудованию

Поскольку подключение пользователей к RT.DataVision выполняется с помощью веб-приложения, специальных системных требований к клиентскому оборудованию не предъявляется. На клиентском оборудовании могут использоваться различные версии ОС, на которых установлены последние версии поддерживаемых браузеров:

* Mozilla FireFox,
* Google Chrome.

### Аппаратные требования

### Серверное оборудование

Минимальные технические требования для установки серверной части RT.DataVision:

1. CPU (процессор) – четыре ядра.
2. RAM (оперативная память) – не менее 8 Гб.
3. HDD/SSD (жесткий диск/твердотельный накопитель) – не менее 20 Гб свободного пространства (без деления на OS и Data).
4. Сетевой адаптер с высокоскоростным подключением к сети, минимум с одним Гигабитным портом.

### Клиентское оборудование

Специальных технических требований к клиентскому оборудованию не предъявляется.

## Dremio

### Системные требования к серверному оборудованию

Dremio устанавливается на ОС Linux.

Кроме того, требуется 64-битный пакет разработки Java версии 1.8.

Проверьте, установлена ли на вашем экземпляре Java (и какая версия) через панель «Системные настройки».

Загрузите последнюю версию [Java SE](http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html).

### Системные требования к клиентскому оборудованию

Пользовательский интерфейс Dremio лучше всего работает со следующими браузерами:

1. Google Chrome 54+.
2. Mozilla Firefox 50+.

# Установка RT.DataVision

RT.DataVision хранит информацию о подключении к базе данных в своей базе данных метаданных. Для этого мы используем криптографическую библиотеку Python для шифрования паролей подключения. К сожалению, эта библиотека имеет зависимости на уровне ОС.

Для установки RT.DataVision необходимо выполнить приведённые ниже операции.

Следующая команда обеспечит установку необходимых зависимостей:

sudo apt-get install build-essential libssl-dev libffi-dev python-dev python-pip libsasl2-dev libldap2-dev

## Виртуальная [среда](https://superset.apache.org/docs/installation/installing-superset-from-scratch/#python-virtual-environment) Python

Мы настоятельно рекомендуем устанавливать RT.DataVision внутри виртуальной среды. Python поставляется ***virtualenv*** как коробочное решение. Если вы используете [***pyenv***](https://github.com/pyenv/pyenv), вы можете установить [pyenv-virtualenv](https://github.com/pyenv/pyenv-virtualenv). Или вы можете установить его с помощью **pip**:

pip install virtualenv

Вы можете создать и активировать виртуальную среду, используя:

# virtualenv is shipped in Python 3.6+ as venv instead of pyvenv.

# See https://docs.python.org/3.6/library/venv.html

python3 -m venv venv

. venv/bin/activate

Или с ***pyenv-virtualenv***:

# Here we name the virtual env 'superset'

pyenv virtualenv superset

pyenv activate superset

Как только вы активируете виртуальную среду, все пакеты Python, которые вы устанавливаете или удаляете, будут ограничены этой средой. Вы можете выйти из среды, запустив ***deactivate*** в командной строке.

## Установка и инициализация RT.DataVision

Во-первых, начните с установки RT.DataVision:

pip install apache-superset

Затем вам нужно инициализировать базу данных:

superset db upgrade

Завершите установку, выполнив следующие команды:

# Создание пользователя admin

$ export FLASK\_APP=superset

superset fab create-admin

# Загрузка образца данных

superset load\_examples

# Создание ролей и разрешений по умолчанию
superset init

# Запуск на порту 8088, используйте -p для другого порта

superset run -p 8088 --with-threads --reload –debugger

Если все работает, вы сможете перейти ***hostname:port*** в свой браузер (например, локально по умолчанию ***localhost:8088***) и войти в систему, используя имя пользователя и пароль, которые вы создали.

# Установка Dremio

1. Установите Dremio, используя:

sudo yum localinstall dremio-<ВЕРСИЯ>.rpm

1. Запустите Dremio:

sudo service dremio start

1. Теперь вы можете перейти к пользовательскому интерфейсу Dremio по адресу http://localhost:9047.

# Настройка RT.DataVision

После завершения установки RT.DataVision необходимо выполнить его настройку, как это указанно в документе «RT.DataVision. Инструкция администратора».