

Платформа Данных Селена

ИНФОРМАЦИЯ ДЛЯ УСТАНОВКИ

Версия программы 1.0

Москва

2024

АННОТАЦИЯ

Платформа Данных Селена представляет собой решение по Построению централизованного, аналитического хранилища данных, которое обеспечивает многомерный и высокопараллельный анализ данных в реальном времени.

Установка и конфигурация программного обеспечения осуществляется только сертифицированными инженерами.

По вопросам установки и обновления программного обеспечения обращаться по адресу: info@selenalakehouse.ru.

ТРЕБОВАНИЯ К ОБОРУДОВАНИЮ И ПРОГРАММНОМУ ОБЕСПЕЧЕНИЮ

Поддерживаемые операционные системы:

- AstraLinux
- RedOS 7.3

Поддерживаемые сторонние хранилище:

- S3
- HDFS

Минимальные требования к вычислительным ресурсам для Мастер Серверов:

- Количество серверов: 3
- Количество ядер процессора: 4 CPU с поддержкой инструкций AVX2
- Размер оперативной памяти: 16 GB RAM
- Размер свободного дискового пространства: 100 GB SSD
- Количество и характеристики внешних разъёмов: 1 порт Ethernet 10 Гбит/с

Минимальные требования к вычислительным ресурсам для Вычислительных Серверов:

- Количество серверов: 3
- Количество ядер процессора: 8 CPU с поддержкой инструкций AVX2
- Размер оперативной памяти: 64 GB RAM
- Размер свободного дискового пространства: 500 GB SSD
- Количество и характеристики внешних разъёмов: 1 порт Ethernet 10 Гбит/с

Минимальные требования к вычислительным ресурсам для сервера метакаталога:

- Количество серверов: 3
- Количество ядер процессора: 2 CPU
- Размер оперативной памяти: 4 GB RAM
- Размер свободного дискового пространства: 100 GB SSD
- Количество и характеристики внешних разъёмов: 1 порт Ethernet 10 Гбит/с

Минимальные требования к вычислительным ресурсам для сервера «менеджера управления »:

- Количество серверов: 1
- Количество ядер процессора: 2 CPU
- Размер оперативной памяти: 4 GB RAM
- Размер свободного дискового пространства: 100 GB SSD
- Количество и характеристики внешних разъёмов: 1 порт Ethernet 10 Гбит/с

Минимальные требования к вычислительным ресурсам для сервера Загрузки и Оркестрации Данных:

- Количество серверов: 1
- Количество ядер процессора: 8 CPU
- Размер оперативной памяти: 16 GB RAM
- Размер свободного дискового пространства: 200 GB SSD
- Количество и характеристики внешних разъёмов: 1 порт Ethernet 10 Гбит/с

Минимальные требования к вычислительным ресурсам для сервера хранения Данных:

- Количество серверов: 4
- Количество ядер процессора: 4 CPU
- Размер оперативной памяти: 16 GB RAM
- Размер свободного дискового пространства: 500 GB SSD
- Количество и характеристики внешних разъёмов: 1 порт Ethernet 10 Гбит/с

Минимальные требования к вычислительным ресурсам для серверов Мониторинга:

- Количество серверов: 1
- Количество ядер процессора: 8 CPU
- Размер оперативной памяти: 32 GB RAM
- Размер свободного дискового пространства: 200 GB SSD
- Количество и характеристики внешних разъёмов: 1 порт Ethernet 10 Гбит/с

Рекомендуемые обозреватели:

- Google Chrome версии 100 и выше
- Mozilla FireFox версии 100 и выше
- Любой браузер на Chromium версии 98 и выше

Установка и конфигурация программного обеспечения осуществляется только сертифицированными инженерами. По вопросам установки программного обеспечения обращаться по адресу: info@selena-lakehouse.ru

УСТАНОВКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Подготовительные действия

Для установки компонентов системы необходимо выполнить ряд подготовительных действий:

- Подготовить серверы для установки платформы хранения данных Селена. Аппаратные требования указаны в главе «Требования к оборудованию и программному обеспечению» настоящего документа;
- Скачать и выложить дистрибутив на серверы для установки. За предоставлением дистрибутива, просьба обращаться по адресу: info@selena-lakehouse.ru;
- Скачать лицензию. За предоставлением лицензии, просьба обращаться по адресу: info@selena-lakehouse.ru.

Основными компонентами системы являются:

- Менеджер управления кластером (Cluster Manager);
- Мастер-серверы;
- Вычислительные серверы;
- Сервер Метакаatalogа;
- Внешнее хранилище S3 Compatible или HDFS (в данной инструкции будет рассмотрена установка платформы с участием S3 Minio);
- Модуль загрузки и оркестрации данными (ETL);
- Модуль AI-Copilot.

Установка Менеджера управления кластером

Для установки Менеджера управления кластером, необходимо:

1. На сервере, где планируется установка Менеджера управления кластером убедиться, что установлен `docker` и `docker compose`.

Пример его установки:

```
sudo apt install docker.io docker-compose
```

```
sudo usermod -aG docker $USER
```

```
exec su - $USER
```

2. Перейти в директорию, где выложен дистрибутив.

3. Загрузить образы контейнеров в Docker.

```
docker load -i ./admin-back.tar
```

```
docker load -i ./admin-front.tar
```

4. Разархивировать selena-1-2-docker-compose.zip, выполнив:

```
Unzip selena-1-2-docker-compose.zip
```

5. Перейти в директорию selena-1-2-docker-compose.

```
cd selena-1-2-docker-compose
```

6. Открыть на редактирование файл docker-compose-standalone.yaml.

```
vi docker-compose-standalone.yaml
```

6.1 Изменить IP-адрес сервера в параметрах (указав вместо localhost, hostname или ip-адрес сервера):

```
WEBSITE_URL=http://localhost:4200
```

```
CONFIG_CASDOOR_FRONTEND_URL=http://localhost:8000
```

```
CONFIG_MONITORING_URL=http://localhost:3000
```

6.2 Сменить логин и пароль для входа в Менеджер управления кластером, изменив параметры:

```
DEFAULT_USER` - имя администратора
```

```
DEFAULT_USER_PASSWORD` - пароль администратора
```

7. Выполнить команду установки и запуска контейнеров, выполнив:

```
docker-compose -f docker-compose-standalone.yaml up -d
```

8. По окончании запуска контейнеров их работу можно проверить, выполнив:

```
docker ps -a
```

9. Для запуска Интерфейса Менеджера управления кластером, необходимо запустить браузер и открыть URL: *http://<IP-адрес Сервера>:4200*

Имя УЗ по умолчанию: *web_console_superadmin*

Пароль по умолчанию: *1234*

Обновление файла лицензии

Для добавления или обновления лицензии в Менеджере управления кластером, необходимо:

1. Открыть Менеджер управления кластером.
2. В навигационной панели нажать на «Лицензия».
3. В открывшемся окне нажать «Обновить лицензию».
4. В открывшейся оснастке указать путь к файлу и нажать открыть.

Добавление Виртуальных Машин

Для добавления виртуальных машин в кластер платформы хранения данных, необходимо:

1. Открыть Менеджер управления кластером.
2. В навигационной панели нажать на «Виртуальные машины».
3. В открывшемся окне нажать «Новая виртуальная машина».
4. В открывшейся оснастке указать:
 - a. Название VM.
 - b. Публичный IP-адрес.
 - c. Внешний IP-адрес.
 - d. Пользователь.
 - e. SSH Ключ или Пароль.
5. Для проверки корректности введённой информации нажать «проверить доступ по SSH».
6. Нажать «Сохранить»

Добавление Мастер Сервера

Для добавления Мастер Серверов в кластер платформы хранения данных, необходимо:

1. Открыть Менеджер управления кластером.
2. В навигационной панели нажать на «Управление кластером».

3. Напротив раздела «Фронтенды» нажать кнопку




4. В открывшейся оснастке выбрать виртуальную машину, на которую будет устанавливаться ПО и указать какую конкретную роль будет брать на себя добавляемый мастер сервер:

- Лидер
 - Последовать
 - Наблюдатель
5. Нажать сохранить.
 6. При первом добавлении Мастер Сервера с ролью Лидер, будет открыта вкладка, в которой необходимо задать `priority networks` - сеть, в которой будет функционировать кластер. Формат CIDR, для маскировки допускается использовать только последний сегмент. Пример: 10.0.0.0/24.
 7. По окончании установки напротив добавленного компонента будет указан статус установки.


Добавление Вычислительного Сервера

Для добавления Вычислительных Серверов в кластер платформы данных Селена, необходимо:

1. Открыть Менеджер управления кластером.
2. В навигационной панели нажать на «Управление кластером».
3. Напротив раздела «Бэкенды» нажать кнопку  .
4. В открывшейся оснастке выбрать виртуальную машину, на которую будет устанавливаться ПО.
8. Нажать сохранить.
9. По окончании установки напротив добавленного компонента будет указан статус установки.

Добавление узла хранения

При отсутствии у заказчика собственного хранилища S3 Compatible, может быть использован S3 Minio, идущая в комплекте поставки Платформы Данных Селена. Для добавления компонента хранения данных, необходимо:

1. Открыть Менеджер управления кластером.
2. В навигационной панели нажать на «Управление кластером».
3. Напротив раздела «Хранилища» нажать кнопку  .
4. В открывшейся оснастке указать:

- a. тип кластера - SINGLE_NODE или Multi_NODE
 - b. access key
 - c. private key
 - d. Директорию установки
5. Нажать сохранить.
 6. По окончании установки напротив добавленного компонента будет указан статус установки.

Добавление Метакаataloga

Для добавления Метакаataloga в кластер платформы данных Селена, необходимо:

1. Открыть Менеджер управления кластером.
2. В навигационной панели нажать на «Управление кластером».

3. Напротив раздела «Внешние каталоги» нажать кнопку



4. В открывшейся оснастке указать:

- e. Наименование каталога
- f. Ключ шифрования данных
- g. Адрес хранилища данных
- h. Access key
- i. Secret key
- j. Пароль Базы данных

Установка компонента загрузки и оркестрации данных (ETL)

Для установки компонента загрузки и оркестрации данных, необходимо:

1. На сервере, где планируется установка, скопировать дистрибутив.
2. На сервере, где планируется установка компонента загрузки и оркестрации данных убедиться, что установлен docker и docker compose.

Пример его установки:

```
sudo apt install docker.io docker-compose
```

```
sudo usermod -aG docker $USER
```

```
exec su - $USER
```

3. Перейти в директорию, где выложен дистрибутив.
4. Загрузить образы контейнеров, выполнив:

```
docker load -i ./etl-alert.tar
```



```
docker load -i ./etl-api.tar
```



```
docker load -i ./etl-master.tar
```



```
docker load -i ./etl-worker.tar
```



```
docker load -i ./etl-tools.tar
```
5. Перейти в директорию deploy-custom.

```
cd deploy-custom
```
6. Выполнить команду установки и запуска контейнеров, выполнив:

```
docker-compose up -d
```
7. Дождаться успешного поднятия всех контейнеров.
8. Открыть браузер и перейти по адресу `http://<IP-адрес сервера>:12345/etl/ui`
логин по умолчанию: admin
пароль по умолчанию: etladmin

Установка AI-Copilot

Для установки компонента AI-Copilot, необходимо:

1. На сервере, где планируется установка скопировать дистрибутив.
2. На сервере, где планируется установка компонента загрузки и оркестрации данных убедиться, что установлен docker и docker compose.

Пример его установки:

```
sudo apt install docker.io docker-compose
```

```
sudo usermod -aG docker $USER
```

```
exec su - $USER
```

3. Перейти в директорию, где выложен дистрибутив.
4. Загрузить образы контейнеров, выполнив

```
docker load -i ./selena-ai-sql-generator-backend.tar
```

```
docker load -i ./selena-ai-sql-generator-frontend.tar
```

5. Подготовить БД. Для этого необходимо скопировать `ddl.sql` из дистрибутива на клиент, с которого будет выполнено подключение к Платформе Данных Селена и выполнить следующие команды:

```
create database ai_sql_generator;
```

```
use ai_sql_generator;
```

```
source /путь к директории/ddl.sql
```

6. Открыть на редактирование файл `docker-compose.yml`

```
vi docker-compose.yml
```

Изменить параметрах:

```
SELENA_URL=http://<IP-адрес Мастер сервера>:8030
```

```
SELENA_USER=для подключения к Мастер Серверу
```

```
SELENA_PASSWORD=пароль
```

```
SELENA_CATALOG=каталог данных
```

```
AUTH_USERNAME= логин
```

```
AUTH_PASSWORD= пароль
```

7. Выполнить команду установки и запуска контейнеров, выполнив:

```
docker-compose up -d
```

8. Дождаться успешного поднятия всех контейнеров.

9. Открыть браузер и перейти по адресу `http://<IP-адрес сервера>:4200`