

# **УСТАНОВКА И ЗАПУСК BI.QUBE METASTAGING**

Москва, 2023

## ОГЛАВЛЕНИЕ

|                                        |                                       |                |           |
|----------------------------------------|---------------------------------------|----------------|-----------|
| ВВЕДЕНИЕ.....                          | 3                                     |                |           |
| ГЛОССАРИЙ.....                         | 3                                     |                |           |
| 1. ЦЕЛИ И НАЗНАЧЕНИЕ METASTAGING ..... | 4                                     |                |           |
| 2. УСТАНОВКА И ЗАПУСКО <b>ОШИБКА!</b>  | <b>ЗАКЛАДКА</b>                       | <b>НЕ</b>      |           |
| <b>ОПРЕДЕЛЕНА.</b>                     |                                       |                |           |
| 2.1. Доступ к площадке .....           | <b>Ошибка! Залкада не определена.</b> |                |           |
| 2.2. Запуск компонента на linux        | <b>Ошибка! Залкада не определена.</b> |                |           |
| 2.3. Запуск компонента на windows      | <b>Ошибка!</b>                        | <b>Залкада</b> | <b>не</b> |
|                                        | <b>определена.</b>                    |                |           |

## ВВЕДЕНИЕ

**Компонент MetaStaging** позволяет консолидировать в стейджинговом слое хранилища данные из гетерогенных источников с поддержанием целостности и унифицированности метаданных, также уменьшает нагрузку на операционные базы при выполнении запросов, а кроме того, обеспечивает надежное подключение различных БД из разнородных источников для помещения данных в единый слой стейджинга (staging area) с поддержанием целостности метаданных в системе-назначения.

В документе приведено описание компонента и принципы работы с ним. Рассмотрены примеры загрузки данных с помощью компонента из разных источников.

Изучение данного документа позволит понять принцип работы компонента.

## ГЛОССАРИЙ

|    |                                                   |                                                                                                                                                                                                           |
|----|---------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. | MetaStaging - BI.Qube                             | Инструмент, предназначенный для транспортировки данных.                                                                                                                                                   |
| 2. | Хранимая процедура                                | Объект базы данных, представляющий собой набор SQL-инструкций, который компилируется один раз и хранится на сервере                                                                                       |
| 3. | Представление                                     | Виртуальная таблица, содержимое которой определяется запросом                                                                                                                                             |
| 4. | Бизнес-представление                              | Представление, в котором собраны Hub, Satellite и Link для сущности                                                                                                                                       |
| 5. | Материализация                                    | Процесс сохранения результата запроса бизнес-представления в таблицу для ускорения выборки.                                                                                                               |
| 6. | Инкрементальная загрузка (загрузка с параметрами) | Регулярная загрузка данных в Greenplum. Извлекаются актуальные данные с даты последней загрузки. В таблице stg.session базы данных settings.db можно отследить историю всех загрузок.                     |
| 7. | Полная загрузка (снэпшоты)                        | Загрузка данных в Greenplum без параметризации. Применяется, когда необходима полная перезагрузка всех данных в таблице на источнике (например, при отсутствии столбца, подходящего для секционирования). |
| 8. | Полная загрузка с сохранением истории             | Загрузка данных в Greenplum без параметризации. Но представления на Greenplum перенацеливаются на новые Parquet-файлы, а старые не удаляются из S3.                                                       |
| 9. | Профиль                                           | Добавляются в таблице stg.profile                                                                                                                                                                         |

|     |                     |                                                                                                                                                                                                       |
|-----|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 10. | Экстрактор          | Компонент системы для извлечения данных из источников в S3. Исполняемый файл находится в директории LoadingToS3. Вызывается в Airflow в соответствии с командами в настройочной БД (settings.db).     |
| 11. | External Table (ET) | Вид таблицы в Greenplum, обеспечивающий доступ к внешним источникам данных, как к объекту самой БД Greenplum. В системе используется для получения доступа к файлам в S3. Используется фреймворк PXF. |
| 12. | Сервисные процедуры | Процедуры, вызываемые автоматически в процессе работы компонента.                                                                                                                                     |

## 1. ЦЕЛИ И НАЗНАЧЕНИЕ METASTAGING

Цель MetaStaging – обеспечить транспортировку данных из систем источников в файловое S3-совместное хранилище данных (HDFS, ObjectStorage) с автоматической генерацией в СУБД Greenplum объектов типа «представление» на каждый полученный файл хранилищем.

Компонент MetaStaging, предназначен для передачи данных из различных источников, как правило, из учетных систем в целевое корпоративное хранилище данных (КХД) с поддержкой целостности метаданных систем-источников, при формировании промежуточного физического слоя хранения учитываются особенности целевой платформы.

Компонент MetaStaging входит в состав системы BI.Qube и может эксплуатироваться как отдельный компонент, так и в составе системы, так и под управлением компонента MetaOrchestrator, в такой конфигурации использование компонента является наиболее эффективной.

### 1.1. Запуск системы

Система разворачивается силами вендора. Развернутый экземпляр (в режиме веб-интерфейса) доступен по адресу:

<https://metacomponents-rr.itprocomp.ru/>

логин: admin

пароль: admin

### 1.2. Развертывание системы

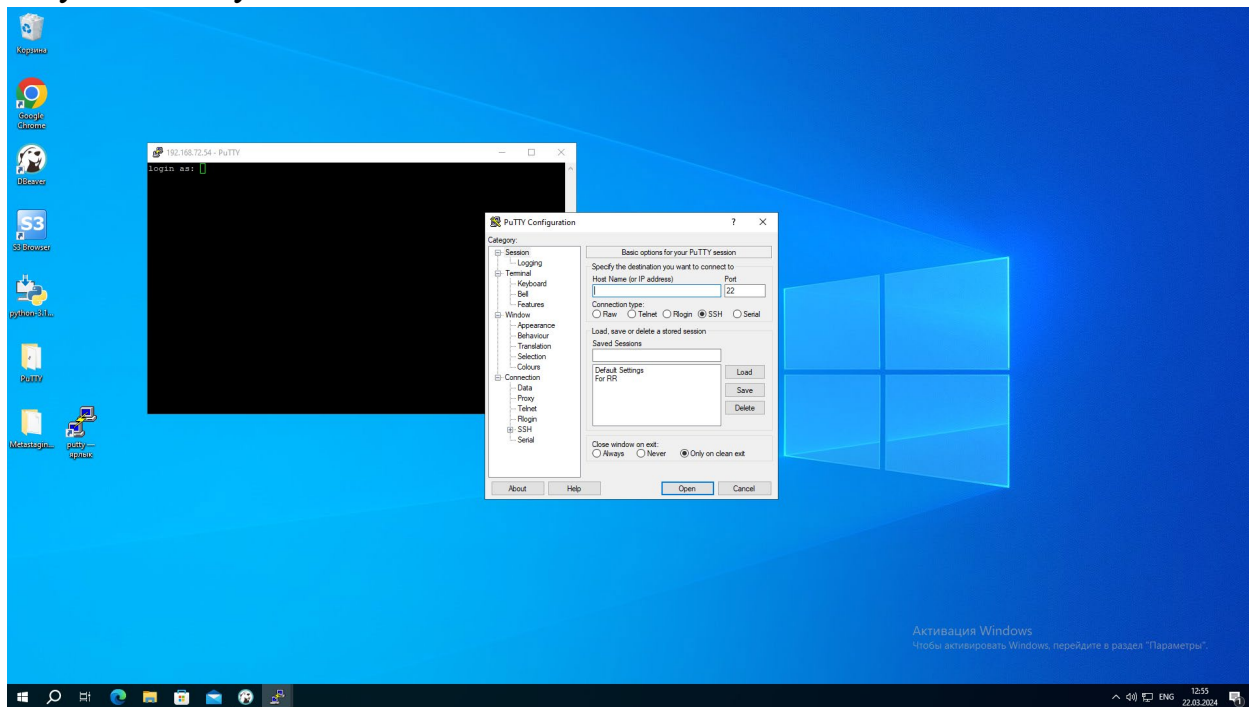
Для получения доступа к исполняемым объектам необходимо зайти по протоколу RDP

212.103.126.82:38154

логин: TestRR01

пароль:

запустить Putty



Ввести (выбрать saved session “For RR”)

Логин itpro\_admin

Пароль MNBrewq123

Система выполняется на сервере, расположенном по адресу:  
192.168.72.54

Доступ к конфигурации NGINX: `sudo nano /etc/nginx/sites-enabled/metacomponents`

Расположение компоуз файла конфигурации докер: `cd /opt/metacomponents/`

Выполнить команду `ll`

```
itpro_admin@biqube-etl-01: /opt/metacomponents$ ll
total 48
drwxr-xr-x  8 itpro_admin itpro_admin 4096 Mar 21 14:49 ./
drwxr-xr-x  6 root        root        4096 Mar 21 11:22 ../
drwxr-xr-x  2 root        root        4096 Mar 21 12:12 certs/
drwxr-xr-x  4 root        root        4096 Mar 21 12:52 data/
-rwxrwxrwx  1 itpro_admin itpro_admin 6926 Mar 21 12:31 docker-compose.yaml*
-rw-r--r--  1 itpro_admin itpro_admin 1698 Mar 21 12:43 .env
-rw-rw-r--  1 itpro_admin itpro_admin 2492 Mar 21 12:02 key.json
drwxr-xr-x  2 root        root        4096 Mar 21 12:16 postgres_backup/
drwx----- 19 lxd         root        4096 Mar 21 12:41 postgres_data/
drwxr-xr-x  2 root        root        4096 Mar 21 12:33 realms/
drwxr-xr-x  7 root        root        4096 Mar 21 12:11 vault/
```

Просмотр содержимого докер-контейнера `nano docker-compose.yaml`

Просмотр всех строк подключения доступен в файле nano .env

Для просмотра содержимого контейнеров:

- Фронтенд: `docker exec -it metacomponents-biqube-frontend-1 /bin/bash` Команда для просмотра содержимого ls

```
itpro_admin@biqube-etl-01:/opt/metacomponents$ docker exec -it metacomponents-biqube-frontend-1 /bin/bash
root@d8d9fe7ce098:/usr/share/nginx/html# ls
50x.html  assets  env.sh  index.html  locales
```

- Бекенд Metastaging: `docker exec -it metacomponents-staging-backend-1 /bin/bash` Команда для просмотра содержимого ls

```
itpro_admin@biqube-etl-01:/opt/metacomponents$ docker exec -it metacomponents-staging-backend-1 /bin/bash
root@019d28b3a4c6:/app# ls
AWS SDK.Core.dll                      Microsoft.Extensions.Configuration.EnvironmentVariables.dll
AWS SDK.S3.dll                       Microsoft.Extensions.Configuration.FileExtensions.dll
Antlr4.Runtime.Standard.dll           Microsoft.Extensions.Configuration.Json.dll
Azure.Core.dll                       Microsoft.Extensions.Configuration.UserSecrets.dll
Azure.Identity.dll                   Microsoft.Extensions.Configuration.dll
BouncyCastle.Crypto.dll              Microsoft.Extensions.DependencyInjection.Abstractions.dll
ClassLibrary1C.deps.json              Microsoft.Extensions.DependencyInjection.dll
ClassLibrary1C.dll                   Microsoft.Extensions.DependencyModel.dll
ClassLibrary1C.pdb                   Microsoft.Extensions.FileProviders.Abstractions.dll
ClassLibrary1C.runtimeconfig.json     Microsoft.Extensions.FileProviders.Composite.dll
ClickHouse.Client.dll                 Microsoft.Extensions.FileProviders.Physical.dll
CommandLine.dll                      Microsoft.Extensions.FileSystemGlobbing.dll
Common.Core.dll                      Microsoft.Extensions.Hosting.Abstractions.dll
Confluent.Kafka.dll                  Microsoft.Extensions.Hosting.Systemd.dll
DajetMetadata.deps.json               Microsoft.Extensions.Hosting.dll
DajetMetadata.dll                    Microsoft.Extensions.Http.dll
DajetMetadata.pdb                    Microsoft.Extensions.Localization.Abstractions.dll
Dapper.dll                           Microsoft.Extensions.Localization.dll
EFCore.NamingConventions.dll          Microsoft.Extensions.Logging.Abstractions.dll
EntityFrameworkCore.ClickHouse.dll    Microsoft.Extensions.Logging.Configuration.dll
ExcelDataReader.DataSet.dll           Microsoft.Extensions.Logging.Console.dll
ExcelDataReader.dll                  Microsoft.Extensions.Logging.Debug.dll
ExcelFiles                           Microsoft.Extensions.Logging.EventLog.dll
Google.Api.Gax.Rest.dll               Microsoft.Extensions.Logging.EventSource.dll
Google.Api.Gax.dll                   Microsoft.Extensions.Logging.dll
Google.Apis.Auth.PlatformServices.dll Microsoft.Extensions.ObjectPool.dll
Google.Apis.Auth.dll                 Microsoft.Extensions.Options.ConfigurationExtensions.dll
Google.Apis.Bigquery.v2.dll           Microsoft.Extensions.Options.dll
Google.Apis.Core.dll                  Microsoft.Extensions.Primitives.dll
Google.Apis.dll                      Microsoft.Extensions.WebEncoders.dll
Google.Cloud.BigQuery.V2.dll          Microsoft.Graph.Core.dll
Google.Protobuf.dll                  Microsoft.Graph.dll
Humanizer.dll                        Microsoft.IO.RecyclableMemoryStream.dll
ItPro.MetaStaging.Lib.deps.json        Microsoft.Identity.Client.Extensions.Msal.dll
ItPro.MetaStaging.Lib.dll              Microsoft.Identity.Client.dll
ItPro.MetaStaging.Lib.pdb              Microsoft.IdentityModel.Abstractions.dll
ItPro.MetaStaging.Backend              Microsoft.IdentityModel.Clients.ActiveDirectory.dll
ItPro.MetaStaging.Backend.deps.json    Microsoft.IdentityModel.JsonWebTokens.dll
ItPro.MetaStaging.Backend.dll           Microsoft.IdentityModel.Logging.dll
ItPro.MetaStaging.Backend.pdb           Microsoft.IdentityModel.Protocols.OpenIdConnect.dll
ItPro.MetaStaging.Backend.runtimeconfig.json Microsoft.IdentityModel.Protocols.dll
ItPro.MetaStaging.Cli                  Microsoft.IdentityModel.Tokens.dll
ItPro.MetaStaging.Cli.deps.json         Microsoft.Net.Http.Headers.dll
ItPro.MetaStaging.Cli.dll               Microsoft.OpenApi.dll
```