

Configuring ELT Maestro for AWS Reshift

Requirements

- Redshift Connection String & Credentials
- Redshift Database Master User Access
- AWS S3 Secret and Access Keys
- SSH Access to ELT Maestro Server (system.cfg config)

Database Objects & ETL Service Account

ELT Maestro requires service account and database objects to perform data loads and ETL operations. Consider below table as a template and use that as reference to change object names. Additionally, an ETL user group can be created on redshift instance in which the service account (etl_user) is a member of for better management.

Name	Type	Description
etl_user	USER	Database user service account used by ELT Maestro
dev	DATABASE	Default redshift database. This is the configuration value for \$SYSTEM_DEFAULT_DATABASE on system.cfg file
integrator	SCHEMA	Default schema for in-database transformations and loads on database dev . This is the configuration value for \$SYSTEM_DEFAULT_SCHEMA on system.cfg file

Based on above configuration assumptions, the JDBC connection string template would look like below for user etl_user & database dev. This is helpful when configuring ELT Maestro windows client application.

```
jdbc:redshift://redshift-cluster-1.cdj0l8zmk59h.us-east-1.redshift.amazonaws.com:5439/dev
```

Creating redshift database objects for ELT Maestro

Run below query as **master user**

```
create user etl_user with password 'abcD1234!';  
  
create schema integrator authorization etl_user;
```

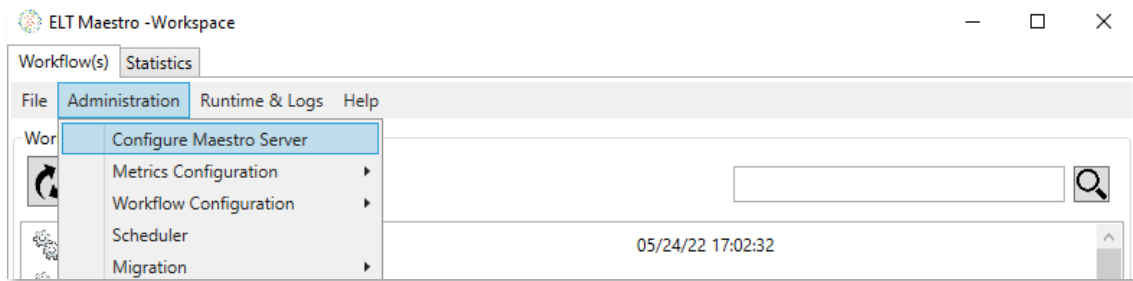
Run below query as **etl_user** to create two functions

```
create or replace function public.b64_encode (argument varchar) returns varchar  
stable as  
$$  
    import base64  
    if argument is None:  
        return ''  
    return base64.b64encode(argument.encode('utf-8'))  
$$ language plpythonu;  
  
create or replace function public.b64_decode (argument varchar) returns varchar  
stable as  
$$  
    import base64  
    if argument is None:  
        return ''  
    return (base64.b64decode(argument).decode('utf-8')).encode('utf-8')  
$$ language plpythonu;
```

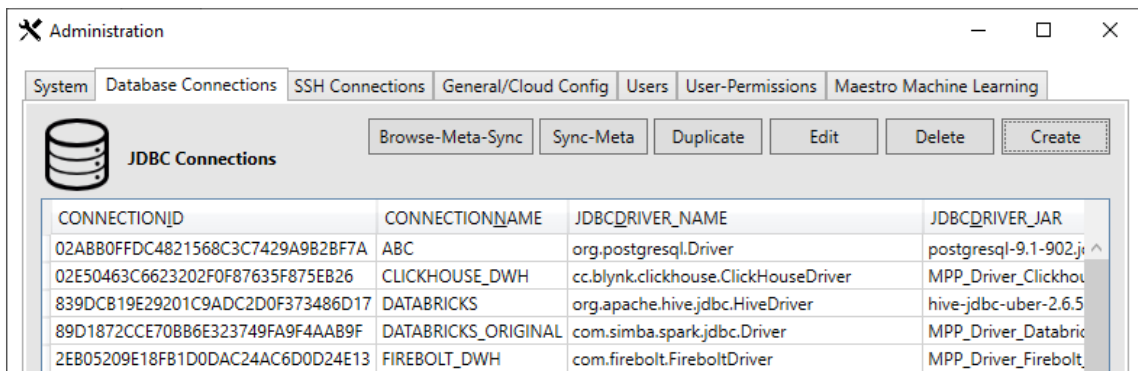
Note: The functions are used to decode ELT Maestro generated load files during lossless load process.

ELT Maestro Client Redshift Connection

- Open ELT Maestro Windows Client
- Click [Administration] -> [Configure Maestro Server]



- Click [Database Connections Tab]



- Click [Create], then enter [jdbc driver class & jar] as shown below, enter [connection string and credentials]
- Then Click [Save]

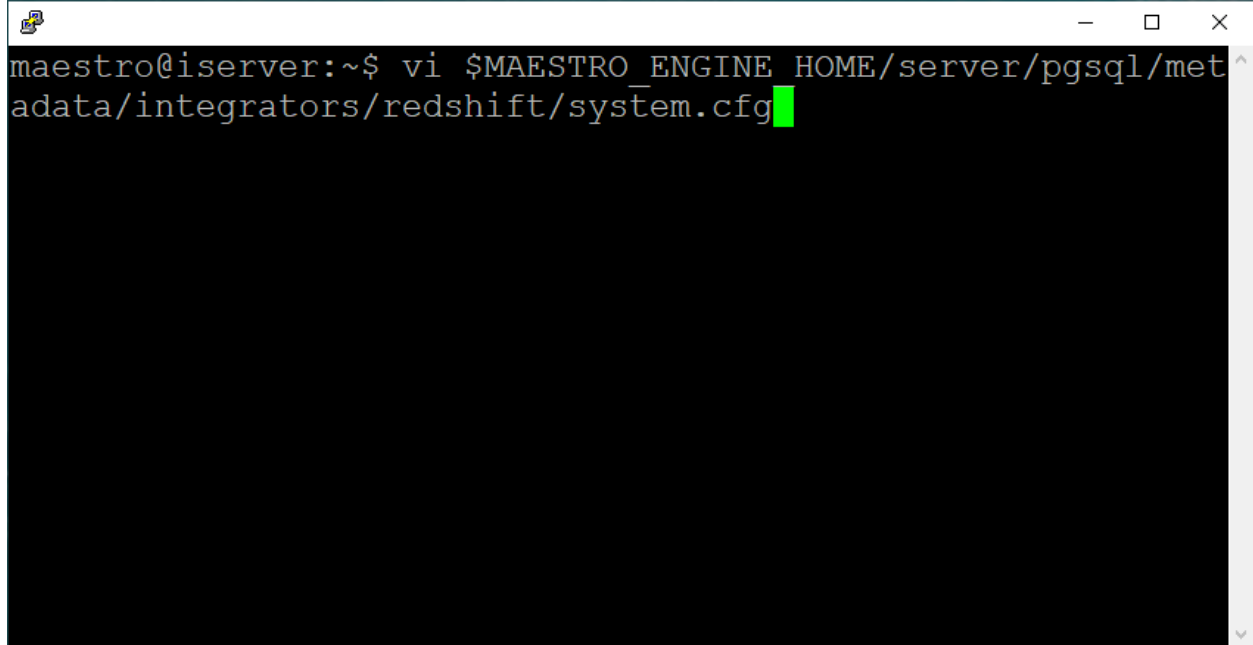


Note: To test connection, select the connection, Click [Edit] then Click [Test]

ELT Maestro Server [system.cfg] configuration file

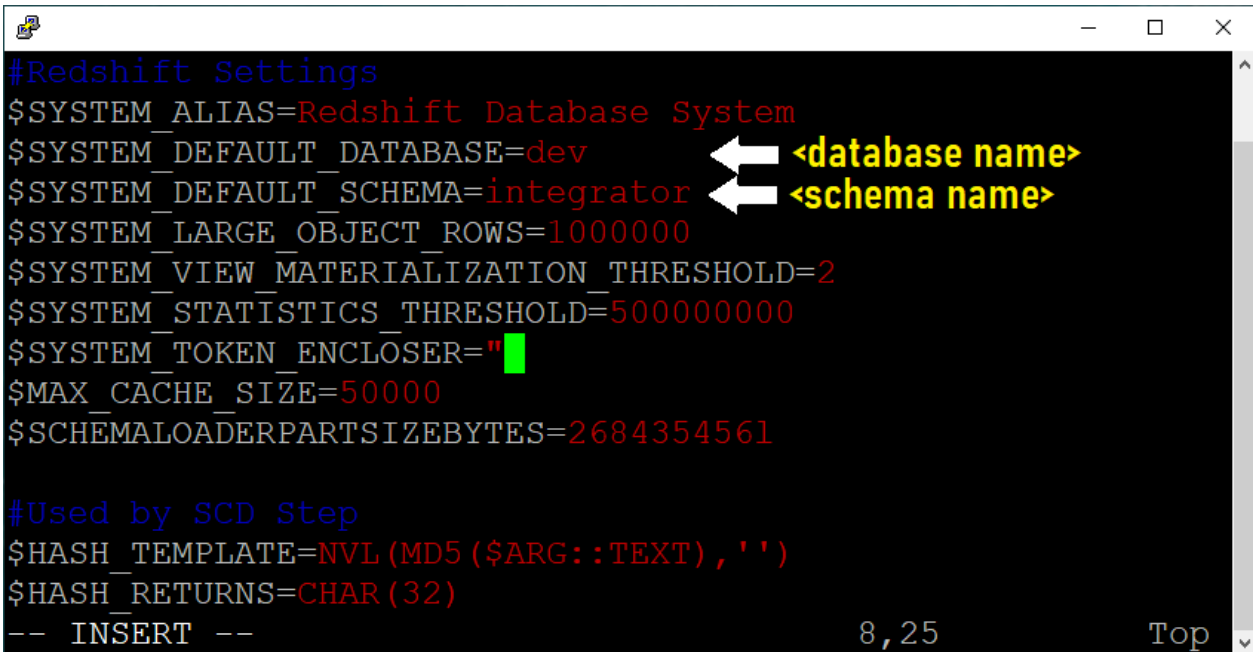
- SSH Login into ELT Maestro linux server as user maestro & load environment file `~/ .env_integrator`
- Edit file

```
[$MAESTRO_ENGINE_HOME/server/pgsql/metadata/integrators/redshift/system.cfg]
```



```
maestro@iserver:~$ vi $MAESTRO_ENGINE_HOME/server/pgsql/metadata/integrators/redshift/system.cfg
```

- Modify database name and schema name



```
#Redshift Settings
$SYSTEM_ALIAS=Redshift Database System
$SYSTEM_DEFAULT_DATABASE=dev ← <database name>
$SYSTEM_DEFAULT_SCHEMA=integrator ← <schema name>
$SYSTEM_LARGE_OBJECT_ROWS=1000000
$SYSTEM_VIEW_MATERIALIZATION_THRESHOLD=2
$SYSTEM_STATISTICS_THRESHOLD=50000000
$SYSTEM_TOKEN_ENCLOSER=""
$MAX_CACHE_SIZE=50000
$SCHEMALoaderPARTSIZEBYTES=2684354561

#Used by SCD Step
$HASH_TEMPLATE=NVL(MD5($ARG::TEXT), '')
$HASH_RETURNS=CHAR(32)
-- INSERT --
```

- Save the changes

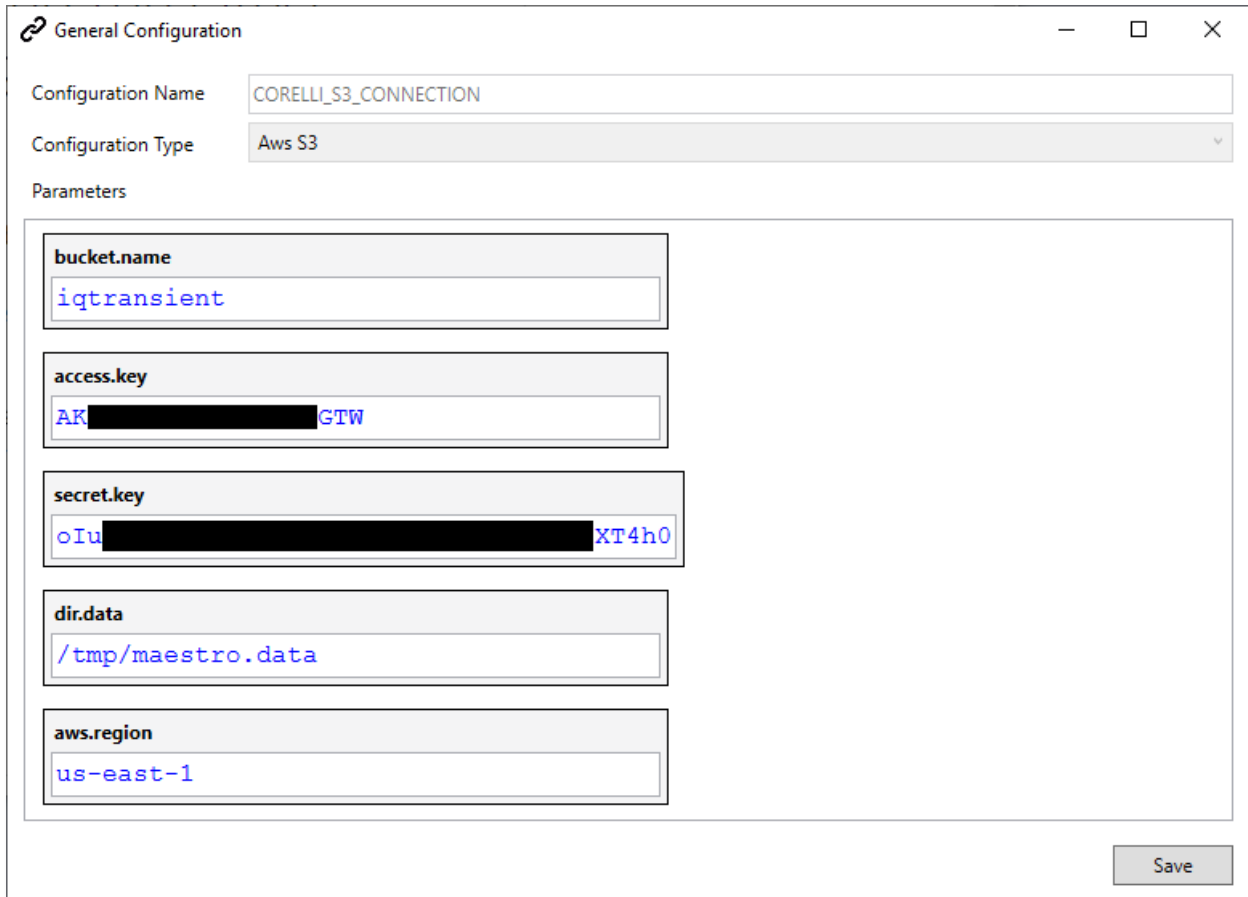
Note:

- Value for `$SYSTEM_DEFAULT_DATABASE` is redshift database name.
- Value for `$SYSTEM_DEFAULT_SCHEMA` is redshift schema name for ELT Maestro to store transient data during loads and in-situ transformation.

AWS Cli & S3 Connection

SSH Login into ELT Maestro Linux server as user maestro and install/configure aws cli package as described on <https://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html>

Using ELT Maestro Client create a new cloud connection with name **CORELLI_S3_CONNECTION** with S3 access and secret keys. Click [Administration] -> [Configure Maestro Server] -> [General / Cloud Config] -> [Create] -> [Configuration Type: Aws S3] -> Input bucket name & keys -> [Save]



The screenshot shows a 'General Configuration' window with the following fields:

- Configuration Name: CORELLI_S3_CONNECTION
- Configuration Type: Aws S3
- Parameters:
 - bucket.name: iqtransient
 - access.key: AK [REDACTED] GTW
 - secret.key: oIu [REDACTED] XT4h0
 - dir.data: /tmp/maestro.data
 - aws.region: us-east-1

A 'Save' button is located at the bottom right of the window.

Note: AWS S3 connection with name "CORELLI_S3_CONNECTION" must exist. This connection name is used by ELT Maestro engine to stage data in S3 bucket during load and unload process.

Support

Email: zdave@maestro-analytics.com

Website: www.maestro-analytics.com

