External Data Accounts Setup and API Guide—Internal

Table of Contents

External Data Accounts Setup and API Guide

Document Data:

Jira ID: TWTSD-28

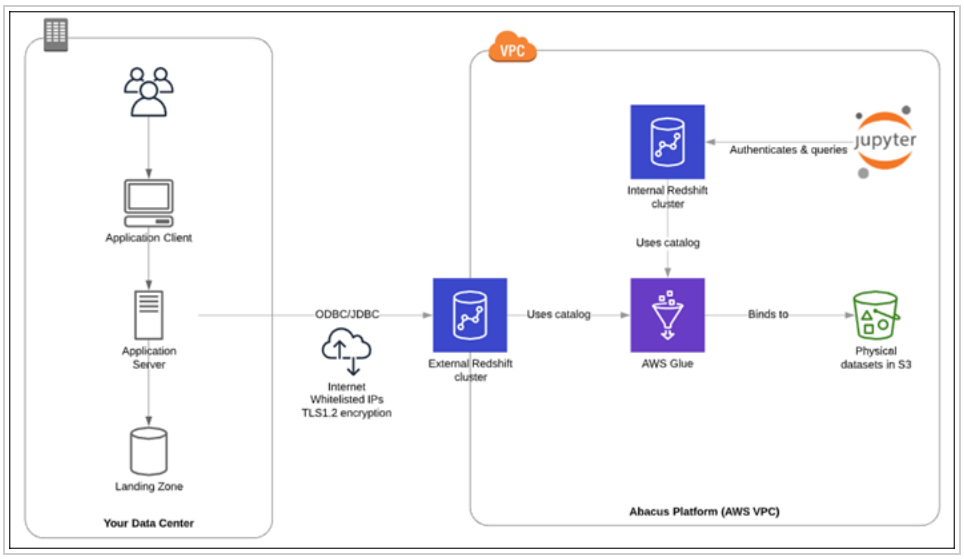
SMEs: Anusha Tomar, Zubair Magdum

Tech Writer: Tony Vinciguerra

Status: Marked complete by SME 11/17/2020.

Customers can connect to data marts on the Abacus Insights Platform using external data accounts (EDAs). Using the API calls described in this document, you can give an EDA can access to an existing data mart. The EDA will expose the data mart through an external Redshift Spectrum cluster that is distinct from the cluster used by Jupyter. Customers can access this external Redshift cluster over the internet with a JDBC or ODBC protocol.

We have tested this integration pattern specifically with SAS, Tableau, and PowerBI.



General EDA Instructions for DataOps and Support

The following steps always require a support email directly from the customer:

- EDA account creation
- Granting data mart entitlements
- Requesting that SRE whitelist an IP address

Connecting to the Cluster

Customers can connect using either ODBC or JDBC.

- ODBC - https://docs.aws.amazon.com/redshift/latest/mgmt/configure-odbc-connection.html
    - Install the AWS ODBC driver.

      The hostname is the URL for the public Redshift cluster, for example: `mg1-public.crsqde5qx9wj.us-east-1.redshift.amazonaws.com`
- JDBC - https://docs.aws.amazon.com/redshift/latest/mgmt/configure-jdbc-connection.html
    - Install the AWS JDBC driver.
    - Configure the JDBC URL using the URL for the public Redshift cluster. For example: `jdbc:redshift://mg1-public.crsqde5qx9wj.us-east-1.redshift.amazonaws.com:5439/dcm`

These are instructions that we provide to the customer:

1. Submit a support email to request the URL for your environment.
2. Connect using Port 5439.
3. Set the database to "dcm".

   Note: Do *not* use "abacus" as the database.

EDA Setup Workflow

1. Log in to Postman, and authenticate using IDM, following the instructions in "Access Control FAQ."

   Note: All of the following calls are in the IDM collection.

2. Create EDA account.

   ```
   POST {{IDM_ENDPOINT}}/eda-accounts
   ```

   Password is returned.

3. Set entitlement (access to a data mart as created in DCM).

   ```
   POST {{IDM_ENDPOINT}}/eda-accounts/{{EDA_ACCOUNT_
   NAME}}/entitlements
   ```

   target_urn example: `urn:ai:idm::data-view/eda/test27/member_csv`

4. Disable user.

   ```
   PUT {{IDM_ENDPOINT}}/eda-accounts/{{EDA_ACCOUNT_
   NAME}}/disable
   ```

   This sends an event to DCM, which in turn sets the Redshift account to DISABLE.

5. Enable user.

   ```
   PUT {{IDM_ENDPOINT}}/eda-accounts/{{EDA_ACCOUNT_NAME}}/enable
   ```

   As we don't store the password, this will return a new password.

6. Remove entitlement.

   ```
   DELETE {{IDM_ENDPOINT}}/eda-accounts/{{EDA_ACCOUNT_
   NAME}}/entitlements/{{EDA_ACCOUNT_ENTITLEMENT_ID}}
   ```

7. List EDA accounts.

```
GET {{IDM_ENDPOINT}}/eda-accounts?paginated_by=20&page=1
```

8. Get EDA account.

```
GET {{IDM_ENDPOINT}}/eda-accounts/{{EDA_ACCOUNT_NAME}}
```

9. Get EDA entitlements.

```
GET {{IDM_ENDPOINT}}/eda-accounts/{{EDA_ACCOUNT_
NAME}}/entitlements?paginated_by=20&page=1
```

10. Delete EDA.

```
DELETE {{IDM_ENDPOINT}}/eda-accounts/{{EDA_ACCOUNT_NAME}}
```

Extracting Server Information for the Public Cluster

1. In the AWS Console, visit the Redshift page.
2. Go to Clusters > Clusters.
3. Click the cluster named `{environment-public}`.
4. Go to Properties > Connection details > View all connection details.
5. Use the server information provided in the connection details.

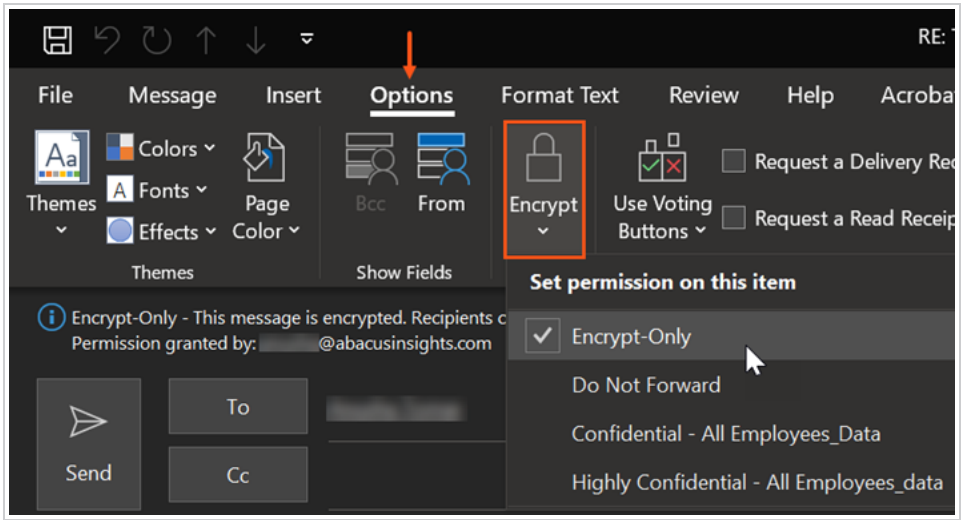
   Note: Tell the customer to set the database to "dcm". Do *not* use "abacus" as the database.

Distributing Credentials

Important: EDA credentials can give the recipient access to PHI. Take care when distributing these credentials.

1. Ensure that the customer has sent an email requesting EDA credentials for a specific recipient.
2. Place the credentials in a text file.
3. Encrypt the file with AES-256 encryption and password-protect it. (We suggest using 7Zip.)
4. Using *encrypted* email, send the following items to the customer admin that is configuring their tool to connect to the EDA. Send each in a *separate* email.
    1. Encrypted email 1: Encrypted text file
    2. Encrypted email 2: Decryption password

       Note: In Outlook, select Options > Encrypt.



5. Respond to the request email confirming that you have sent the credentials to the specified recipient.
6. Once the customer confirms receipt, delete the local copy of the credentials file from your computer.
7. Delete the randomized password that you used to encrypt the file.

Environment Instructions for SRE

Whitelisting IP Addresses

1. Ensure that you have an email from the customer requesting the whitelisting.
2. Whitelist IP addresses by updating the Terraform layer for the environment. Set the `redshift_public_whitelist_ips` variable in the storage/Redshift layer.

What does Abacus log? What does it not log?

The Abacus Insights Platform logs activity that happens within the platform boundary, including:

- Creation of a new external data account (EDA)
- Granting of data mart entitlements to an EDA
- Any edits or updates to an EDA
- Every query made through an EDA against a data mart

The Abacus Insights Platform *cannot* log user activity that occurs outside its boundary, including:

- End users being granted access to a data mart within a third-party analytics application
- Which of those end users executed a specific query through the EDA
- Which analytics applications are connecting to the platform using a given set of EDA credentials