Pelican User Guide

Version no : 1.0.0.4

5 April 2021

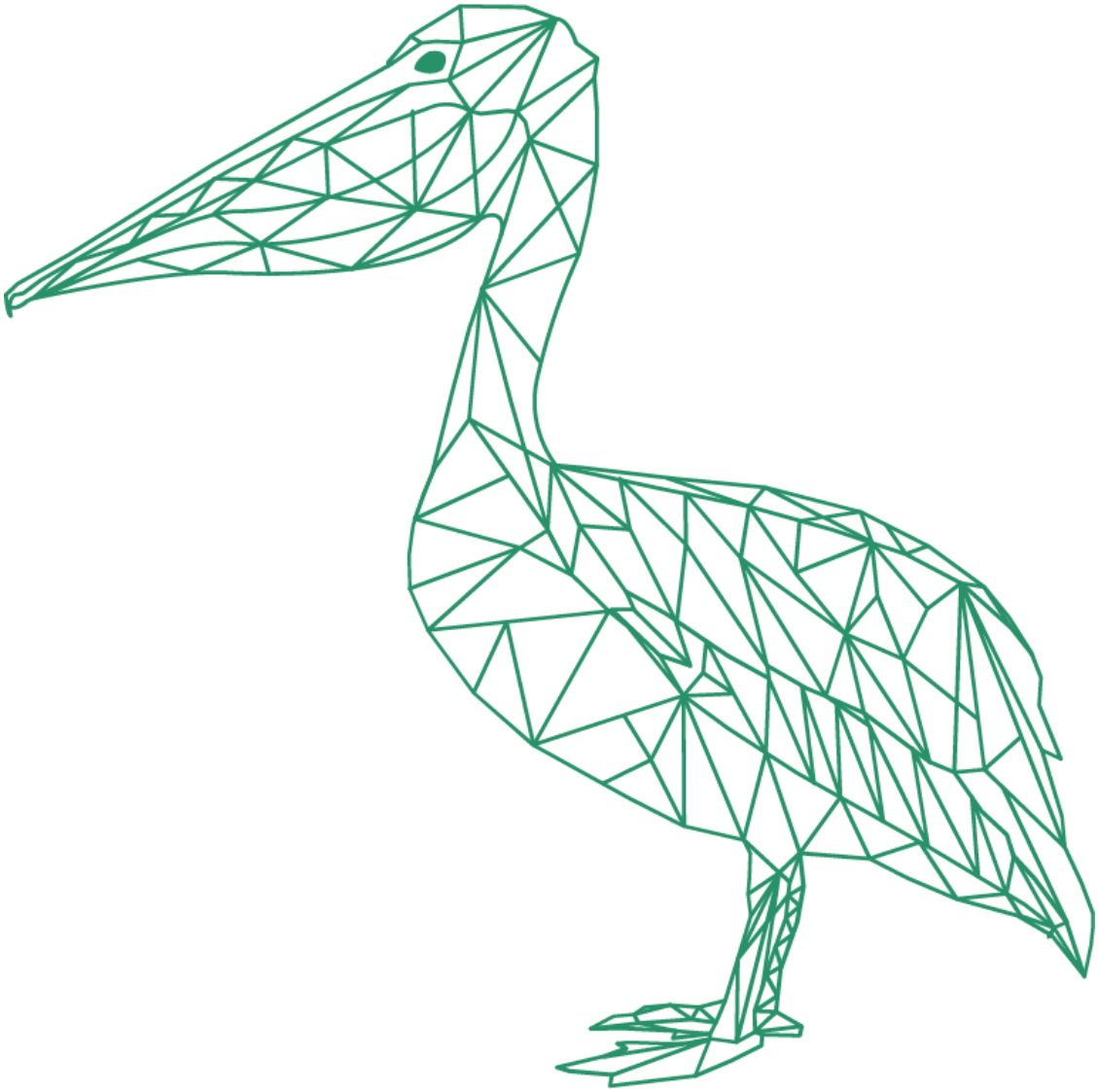


Table of Content

1. Preface

The *Pelican user guide* is written for developers and software engineers who are responsible for comparing and validating data across various data stores. The *Pelican user guide* assumes that you have an understanding of your operating systems, relational database concepts, database engines, flat files, and how the mainframe system works in your environment. This guide also assumes that you are familiar with the interface requirements for your supporting applications.

2. Pelican Overview

Pelican is an innovative enterprise application that ensures business that the data migrated from reference to destination data store accurately and reliably by performing validation over migrated data. Pelican helps the user to compare and validate data across different data stores quickly.

Pelican helps users to validate large data sets without any data transfer from source to target. Pelican allows the user to copy a large dataset (historical and real time data) to a destination data store for validation. It uses an innovative approach to compare the data without moving it across data stores. It also shows the records in data which are having differences.

The user can automate the comparison using schedulers and API. Pelican uses the Phonetic Matching and Approximate Matching algorithms to search the best possible combination of tables from the target datastore. It also facilitates configuring and execute various schedulers as per the business requirement. These Schedulers are created to compare and validate table dataset periodically.

You can configure an email notification for schedulers that sends an email automatically to various users when the scheduler executes. Pelican provides lineage support, which enables the user to track the movement of data across various nodes. This information is retrieved from the job history server of various data stores configured in the Pelican.

Once the validation is done, the application generates statistics with the following information:

- Count of total rows at source
- Count of total rows at destination
- Count of mismatch rows at destination
- Count of extra rows at destination

- Count of missing rows at destination
- Total mismatch row count
- Validation Status
- Sample of mismatch data

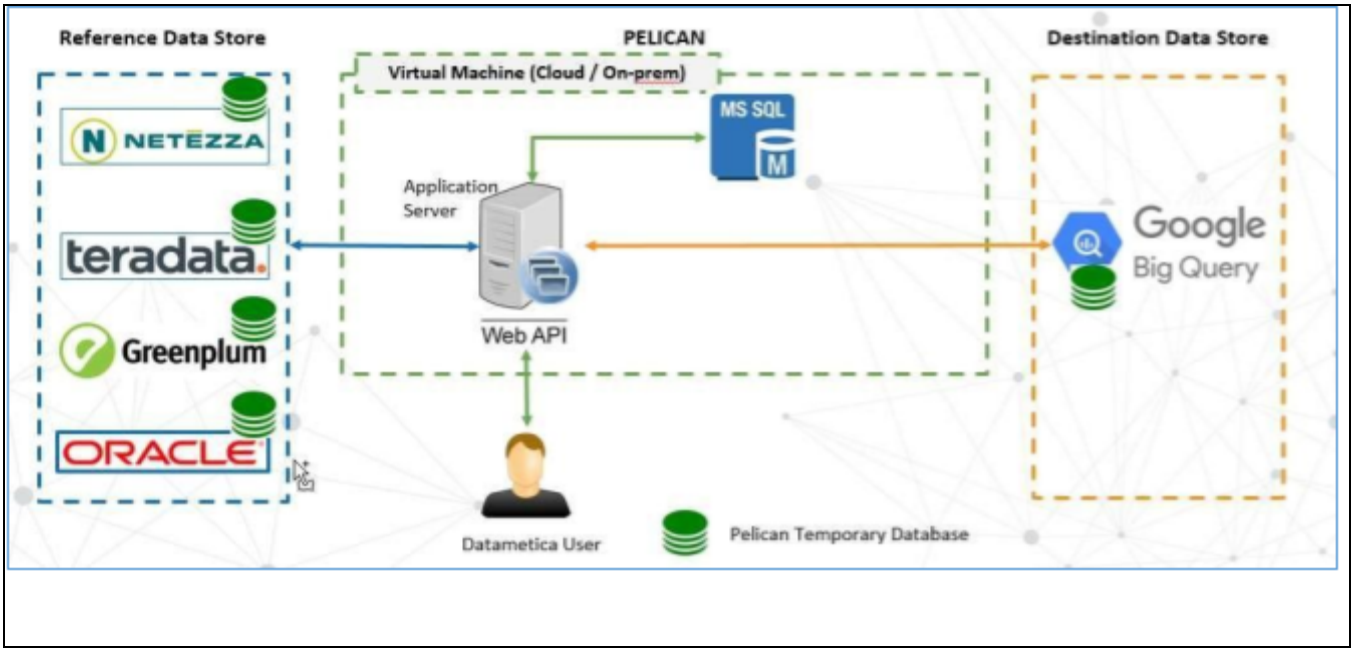
3. Data Store Support

Pelican supports comparison between following data stores using an innovative approach:

Sr. No	Source Data store	Destination Data store
1	Teradata	Big Query
2	Netezza	Big Query
3	Hive	Big Query
4	Oracle	Big Query
5	BigQuery	Big Query
6	DB2	Big Query
7	Oracle	Hive
8	Teradata	Hive
9	Netezza	Hive
9	Hive	Hive
10	Big Query	Hive
11	Oracle	Oracle

4. Architecture

Here is the complete architecture on how Pelican works.



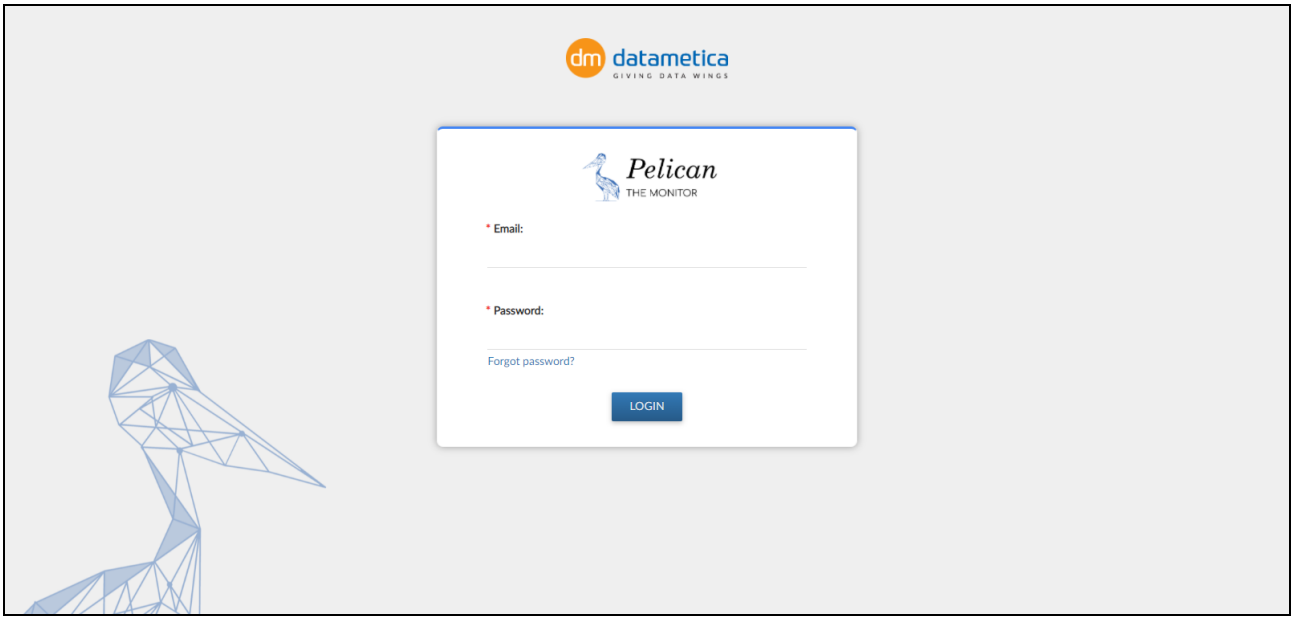
5. User Management

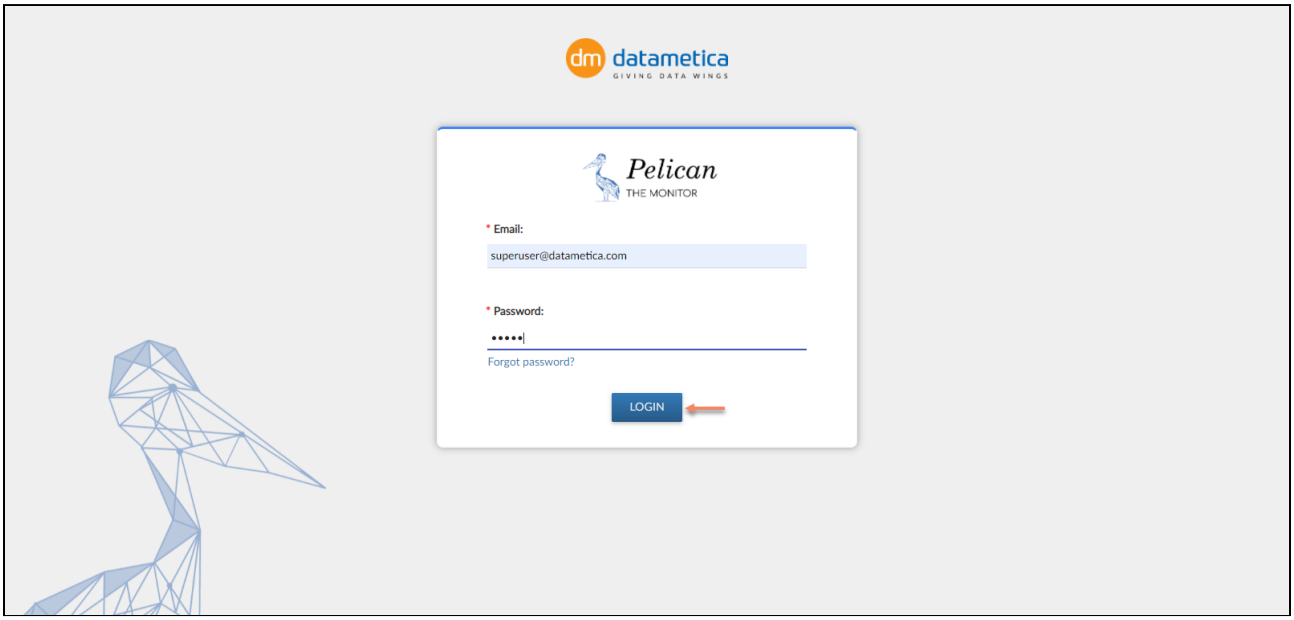
In User Management the whole administration of the application is accessed, providing secured authorization to the users. Each user is assigned a unique identity to authenticate the application.

5.1. Login

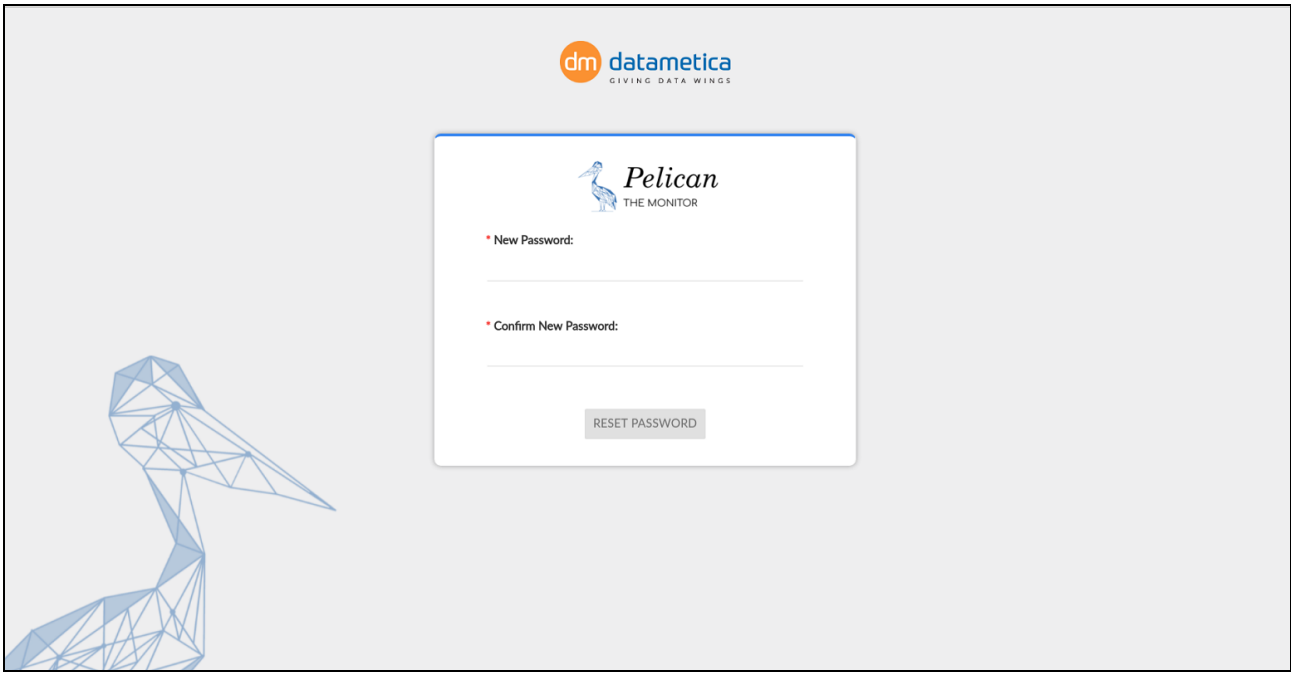
Pelican needs two details on the login page

- User’s email id
- Password





Once the user logged in for the first time, Pelican will ask them to reset the default password.



After completing the steps the user will navigate to the Validation Result page.

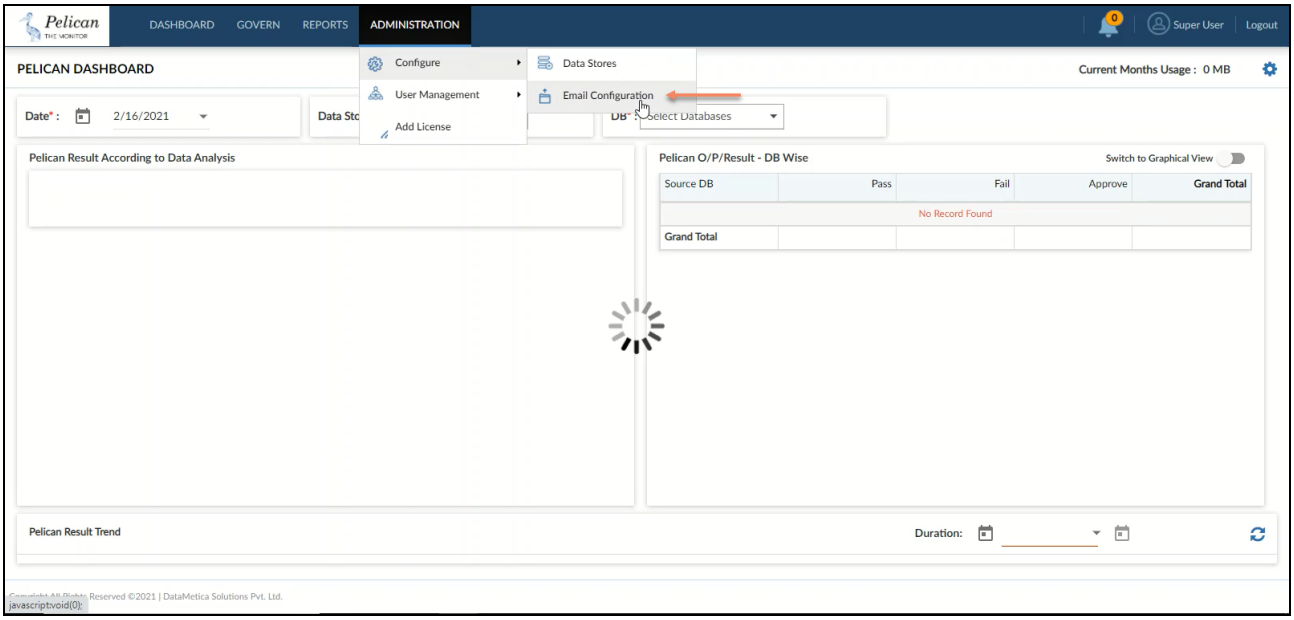
5.2. Configure SMTP server

SMTP server configuration functionality allows the administrator to configure SMTP server and it will be used for sending passwords to newly added users.

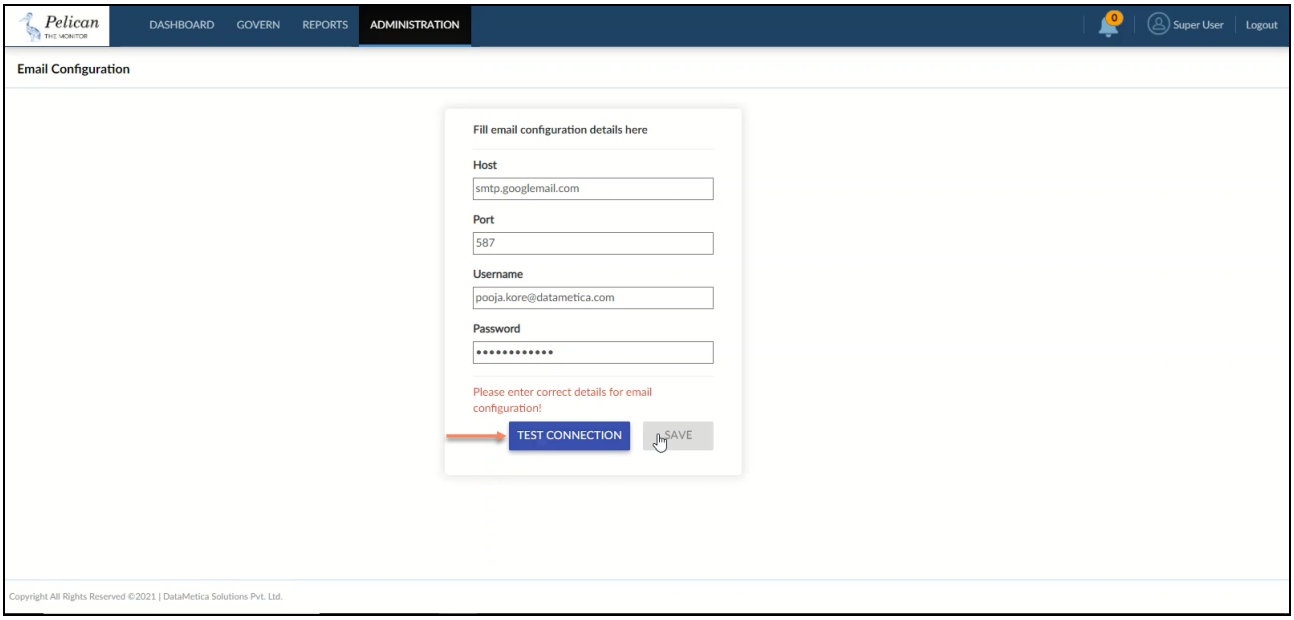
Configure SMTP server and Sender Email Address ;

To configure the SMTP Server and senders email address:

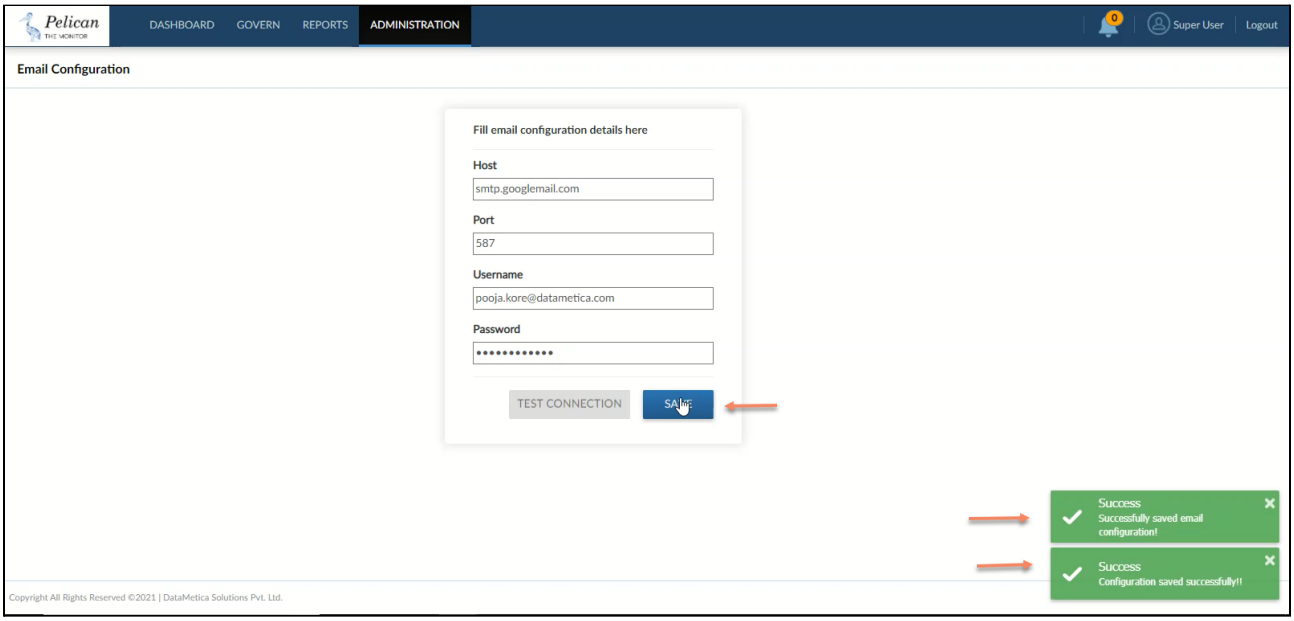
1. Go to the **Administration → Configure → Email Configuration**.

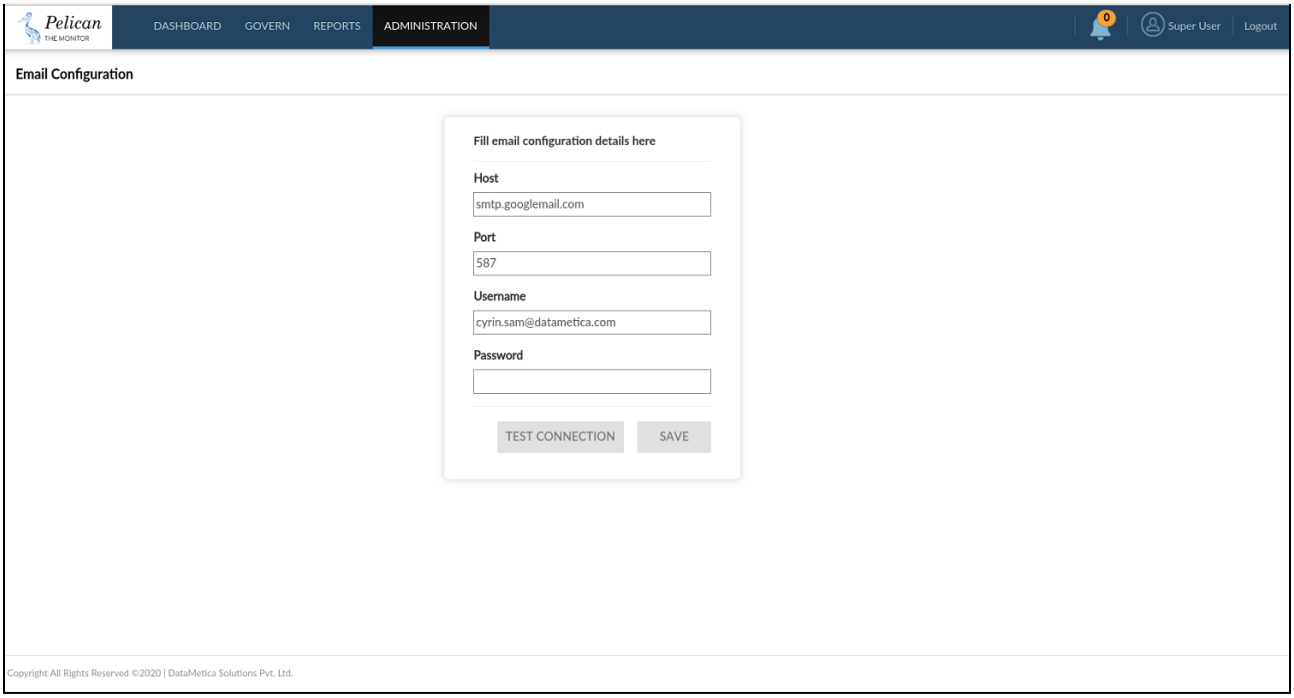


2. Fill the below details in the Email Configuration window.
    - Enter **Host Name** (http://smtp.googlemail.com/)
    - Enter **Port Name** (547)
    - Enter **User Name** (firstname@datametica.com)
    - Enter **Password** in the provided field. ••••••••••



3. Click on Test Connection
4. Click the **Save** button to save the configuration.



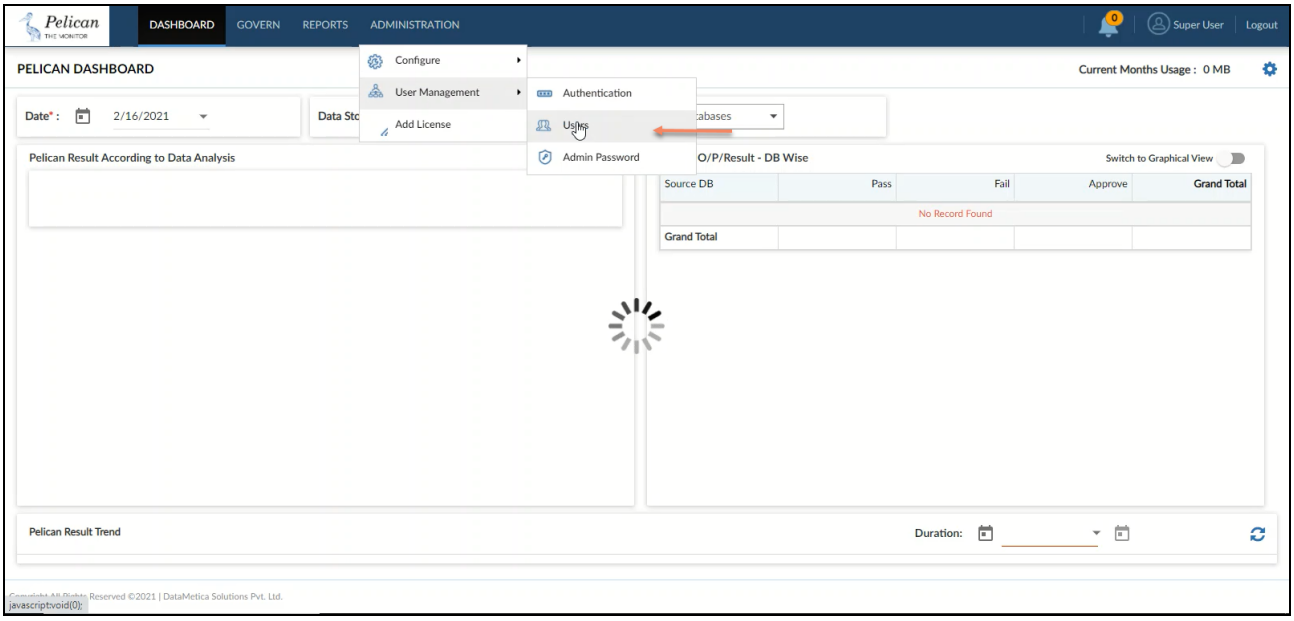


5.3. User creation

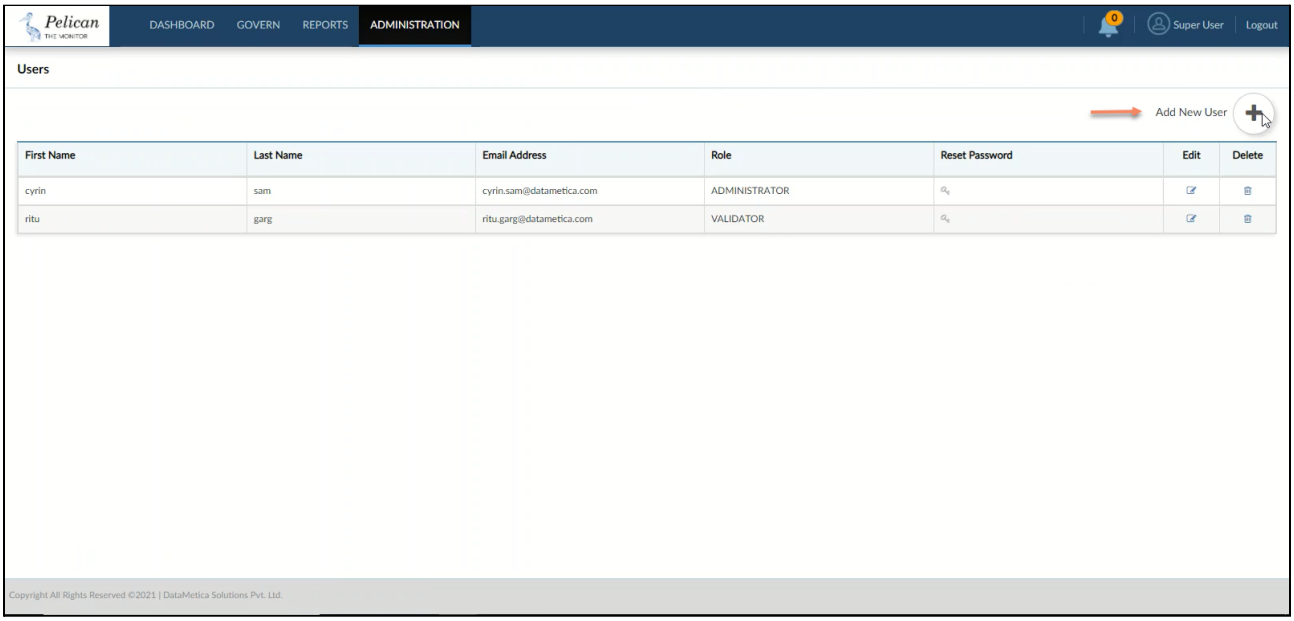
The Users tab allows you to create a new user in the role of Validator or Admin. You can edit or delete an existing user as well. This option will be available if LDAP configuration is not set up.

Steps to add new user:

1. Go to **Administration --> User Management → Users**



2. Click **Add New User**

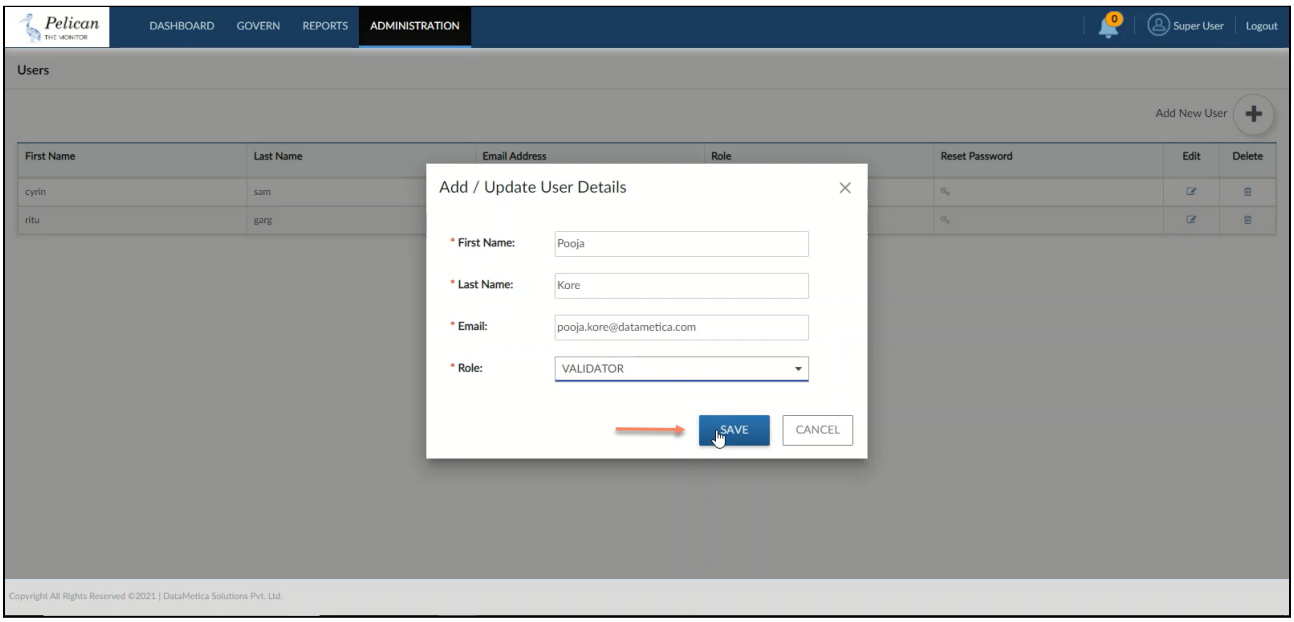


3. In Add / Update User Details Form, enter First Name, Last Name, Email, and select Role from the drop-down list.

Roles -

1. Super user

2. Validator

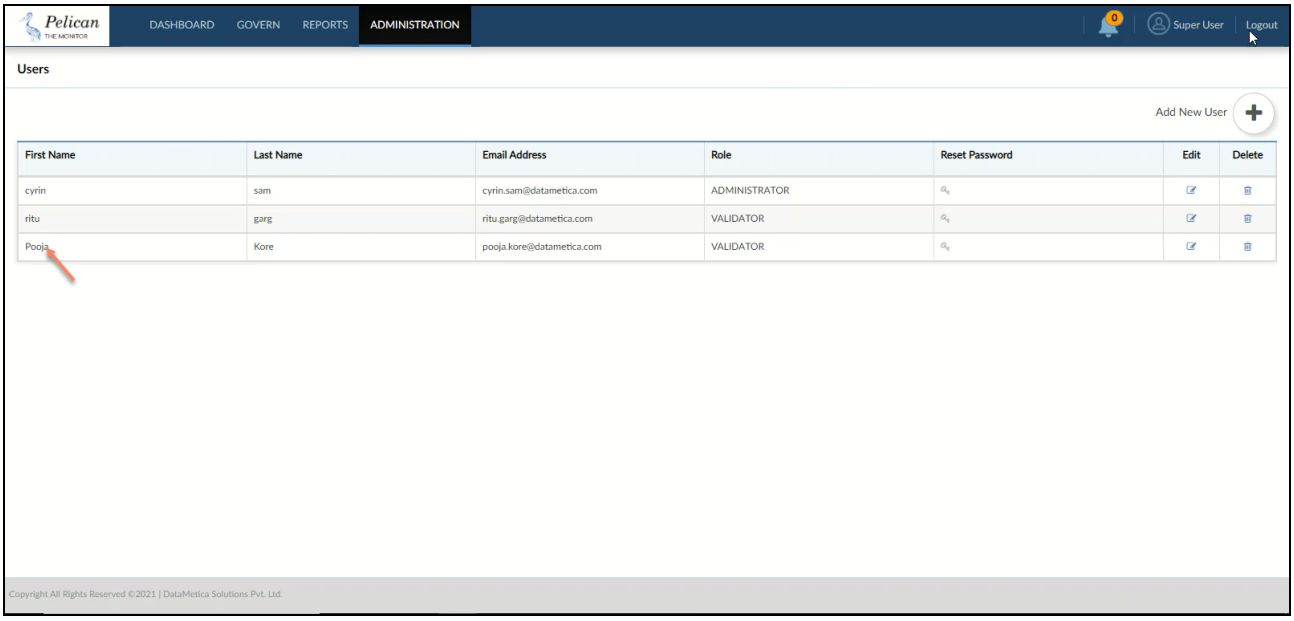


Note:

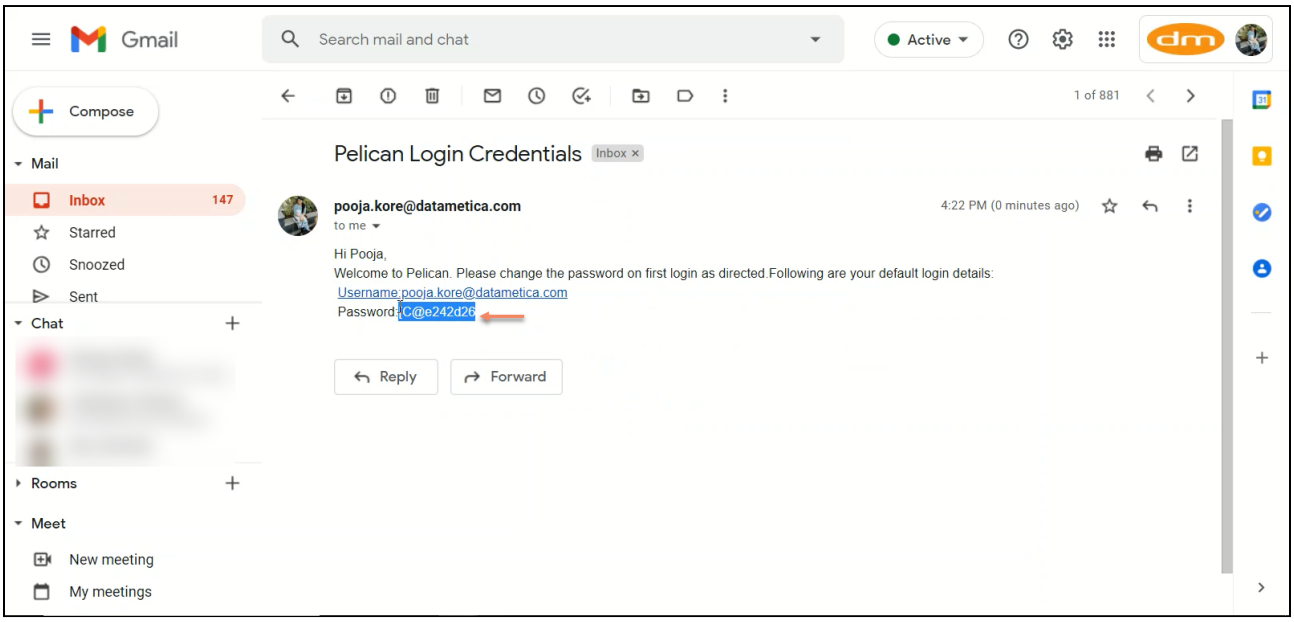
Admin: Users having the role as **admin** have the same rights as Super user. They can create new users having roles as Admin or Validator. They can create data stores for validation and configure email for sending mails.

Validator: Users having the role as **validator** do not have access to create new datastore and can't create new users, but they can create mappings, run schedulers, and view reports.

4. Click **SAVE**.

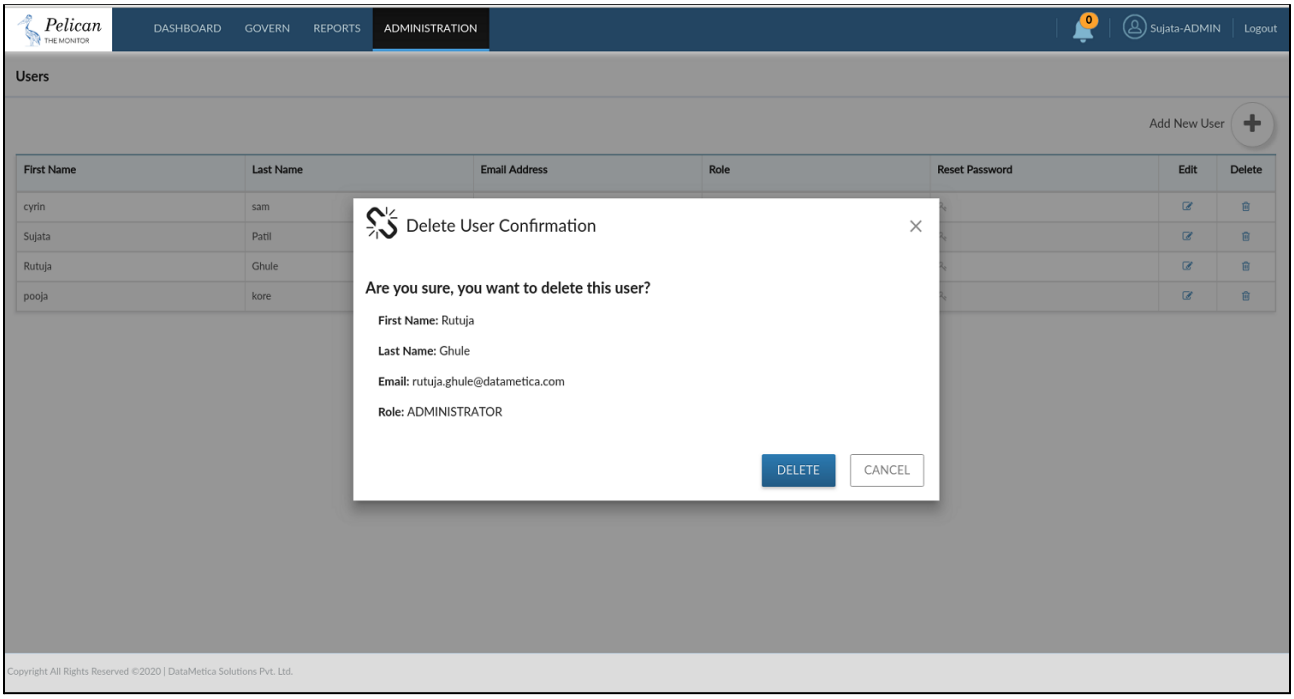


A system generated Pelical Login Credentials email is generated and sent to the first-time login user which includes username and password. The user should login with those credentials and create a new password.



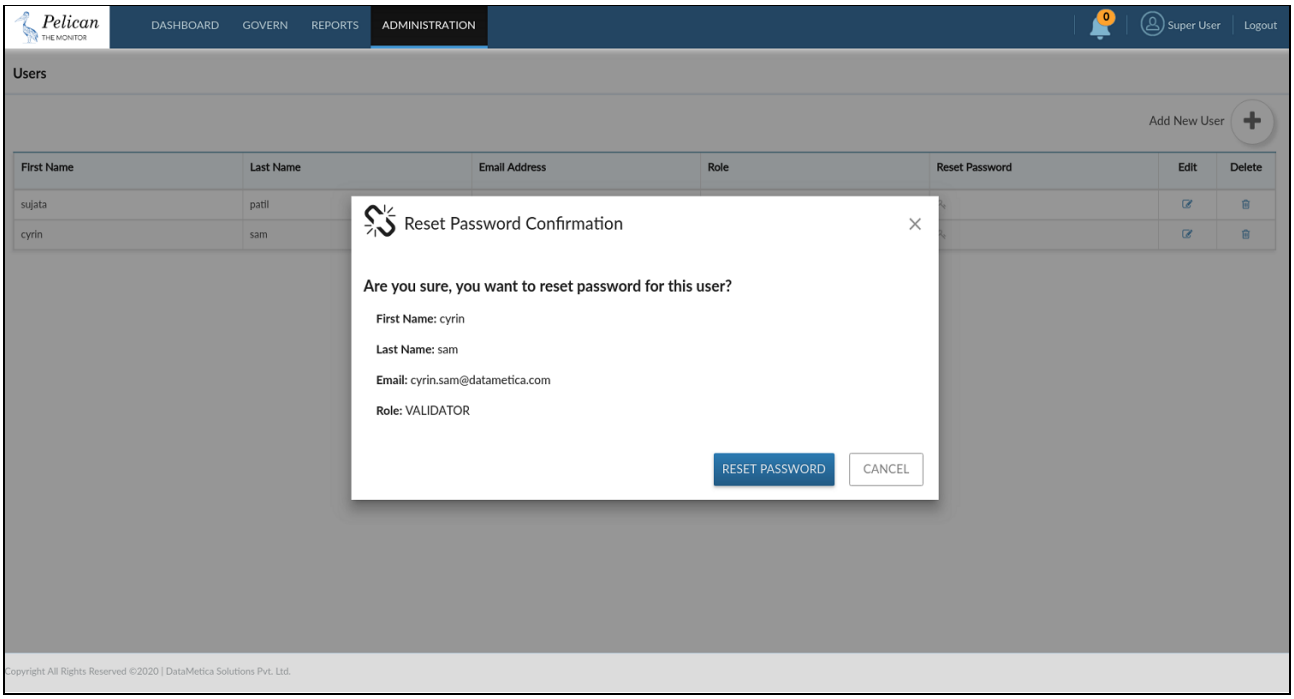
5.4. Modify existing user

You can edit the existing user details by clicking the edit icon on the *Manage Users* screen. So make the necessary changes and click SAVE in the **Add / Update User Details** form. You can delete the user details by clicking the delete icon on the same page of *Manage Users.* The system displays a deletion confirmation dialog box, hence, click the **DELETE** button as shown in the image.



Confirmation message will appear when the admin selects an action.

Reset Password-

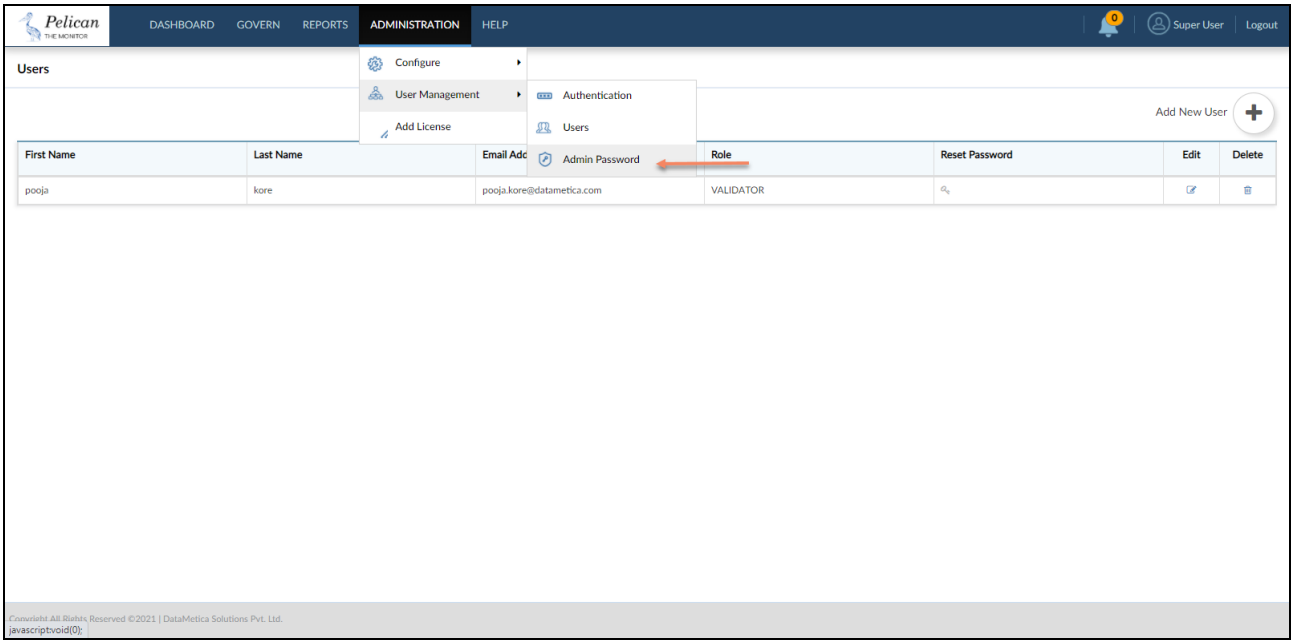


5.5. Admin Password

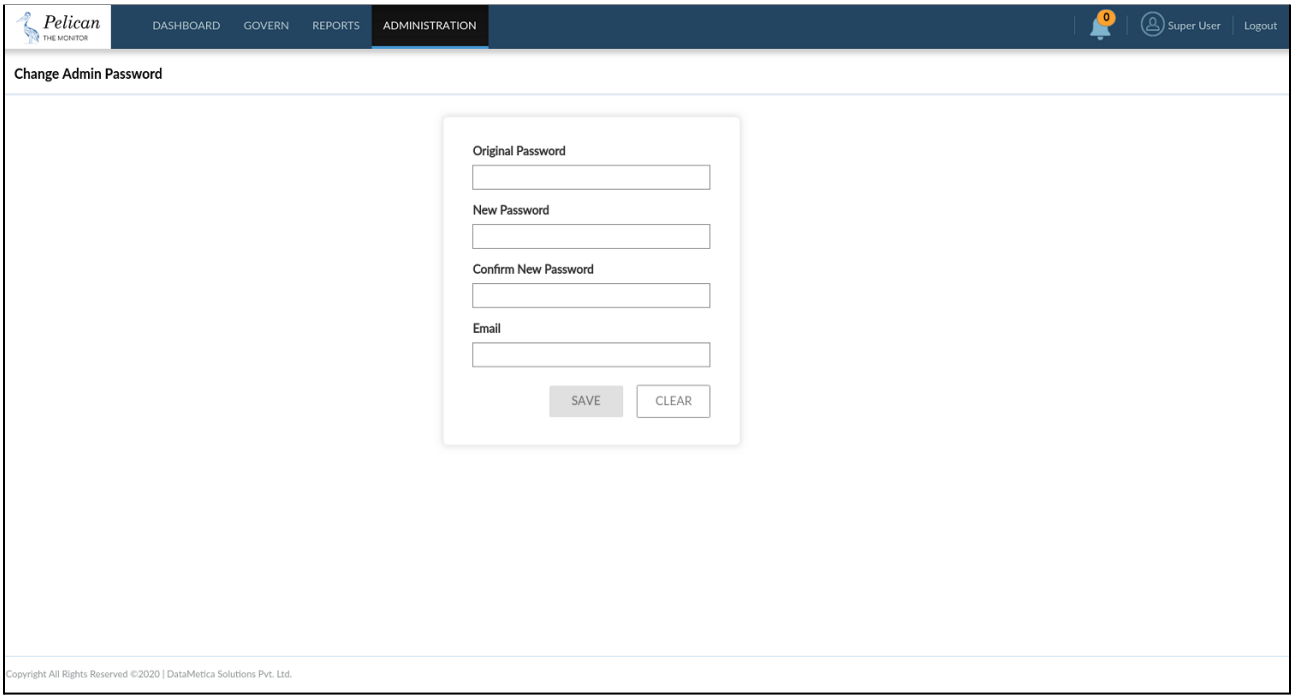
In the Admin Password of User Management, the administrator can change the passwords from original to new password.

Steps

1. Go to **Administration > User Management > Admin Password**.



2. Enter Original Password, New Password, Confirm New Password, Email.
3. Click **SAVE**.



5.6. LDAP User Management

LDAP user authentication is the process of validating a username and password combination with a directory server. Only the superuser has the right to create LDAP users. LDAP users can't create any new users. Only users with Admin access can create new LDAP users.

In User Management you are:

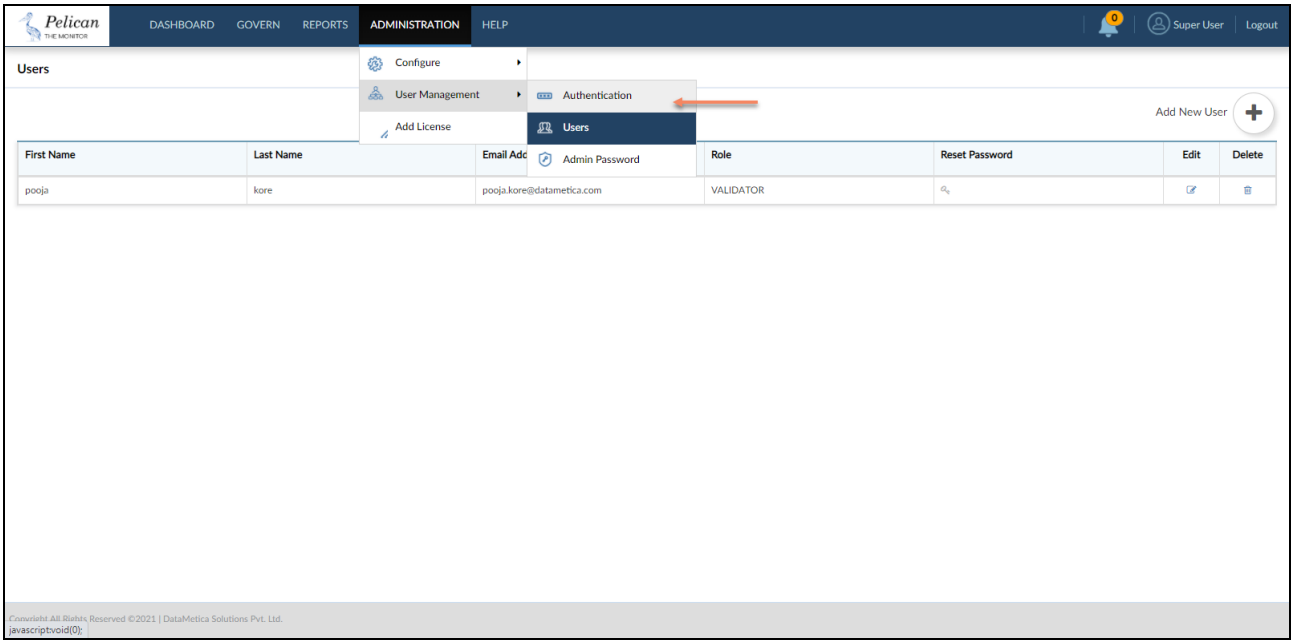
- Setting LDAP credential
- Adding/editing/deleting users
- Accessing control
- Assigning roles to the users
- Administrating Pelican
- Adding/editing roles
- Granting permissions
- Assigning/editing/deleting user groups

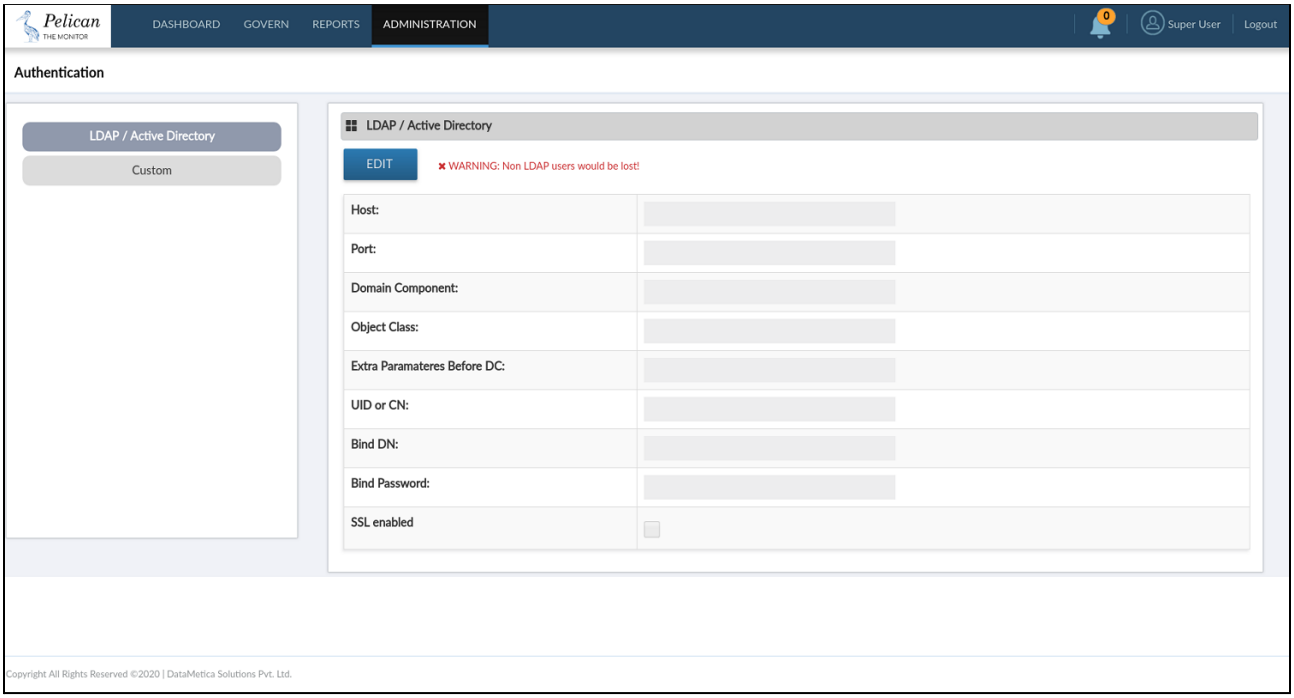
5.6.1. Authentication

You can add the Authentication details for the LDAP/ Active Directory

Steps

1. Go to **Administration > User Management > Authentication**.
2. Click **Edit**.





3. Enter Host, Port, Domain Component, Object Class, Extra Parameters Before DC, and other fields.
4. Select SSL enabled and click **DONE**.

Field Name	Description
Host	Ip or machine name on which Ldap server is running
Port	Port at which connection to the ldap server is to be created
Following fields are used to create the BaseDN, in which user should be searched	
Domain Component	The comma separated values that are put with the attribute “dc”
Object Class	Object class decides the attributes of the entry and is part of the search criteria
Extra Parameters Before DC	Any extra parameters that are present between common name & domain component
UID or CN	Value of this field will be either uid or cn. This attribute denotes the common name of the user.
Bind DN	This field will hold the qualified value of the user which may be required to connect to ldap
Bind Password	This field will hold the password corresponding to the user mentioned in above field
SSL Enabled	If ldap is secured, check the checkbox. Certificate should also be configured in the pelican jvm

Note: Post Ldap/AD successful connection, we will need to restart ldap so that added users can log in to the application.

After Configuring Authentication Page the user has to navigate to the bin folder where Pelican is installed. Stop the Pelican application using ./shutdown.sh

Then navigate to /webapps/ROOT/WEB-INF/classes/config.properties file.

And add the following properties and corresponding values in it:

1) authenticationByDistinguishedName = false
2) enterpriseIdAttribute = sAMAccountName

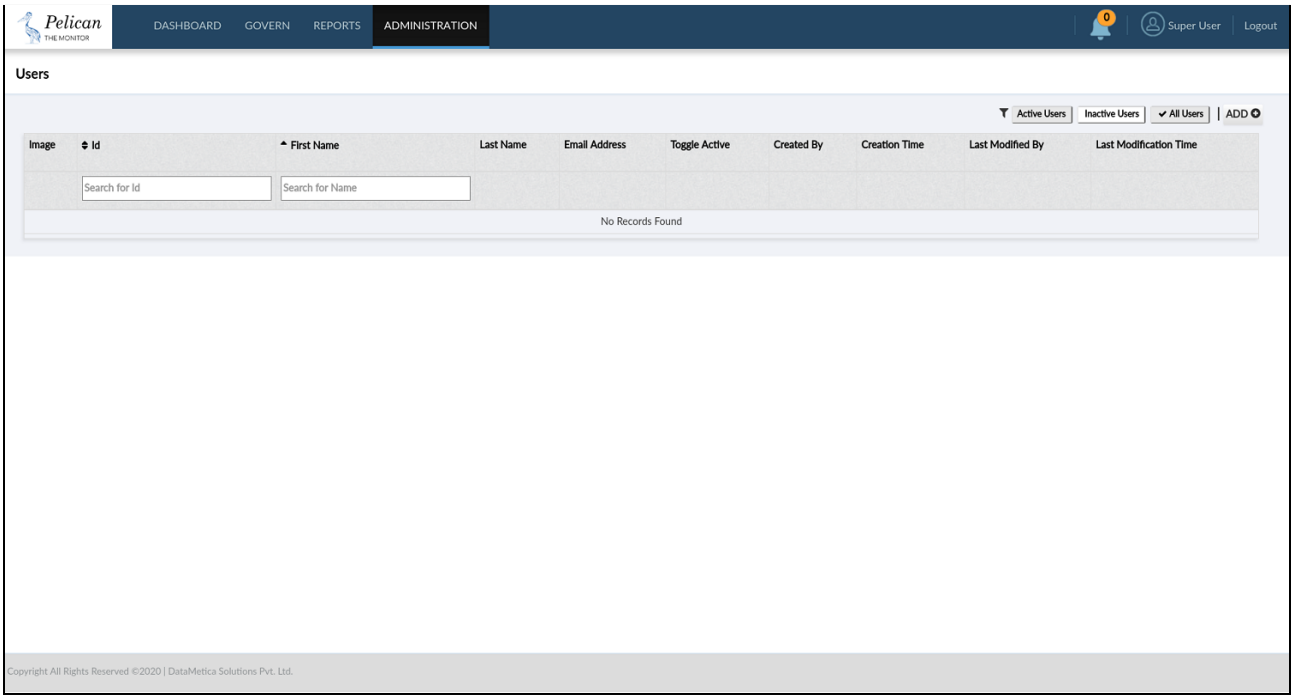
Note: Once you have added the credential super user can only create ldap user and can’t create non-ldap users having role as admin or validator.

5.6.2. Adding Ldap Users

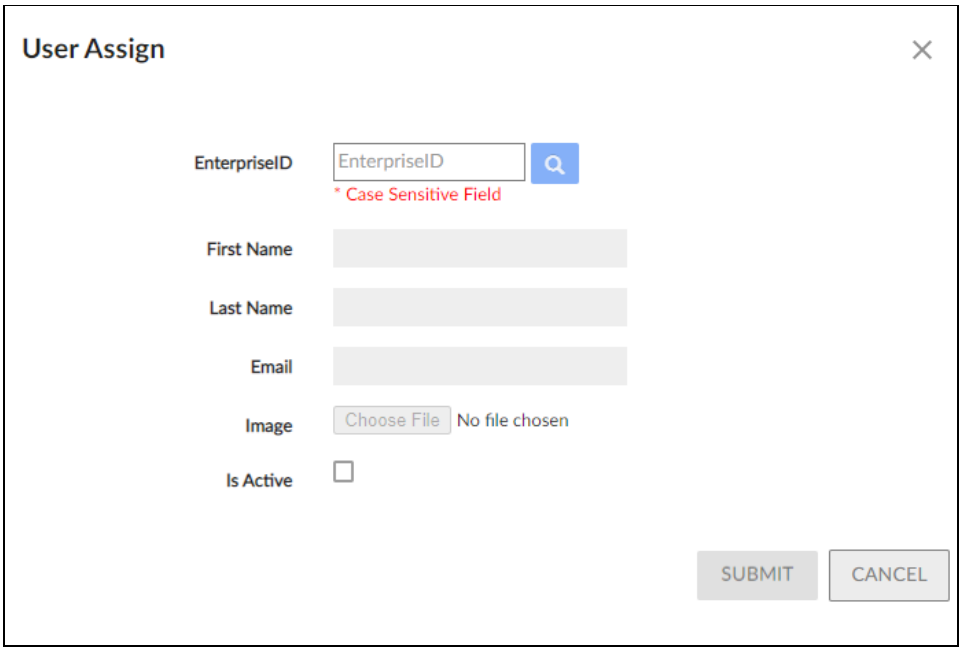
In **User Management** you can add new LDAP users and view the details on the *User* page.

Steps

1. Go to **Administration > User Management > Users**.
2. Click the **ADD + button**.



3. On the ***User Assign*** form enter EnterpriseID and all other details are visible automatically in the fields.
4. Select the Is Active checkbox and click **SUBMIT**.



Tips: You can apply filters to view all the active and inactive users by clicking Active or Inactive users.

5.6.3. Access Control

The Access Control features to:

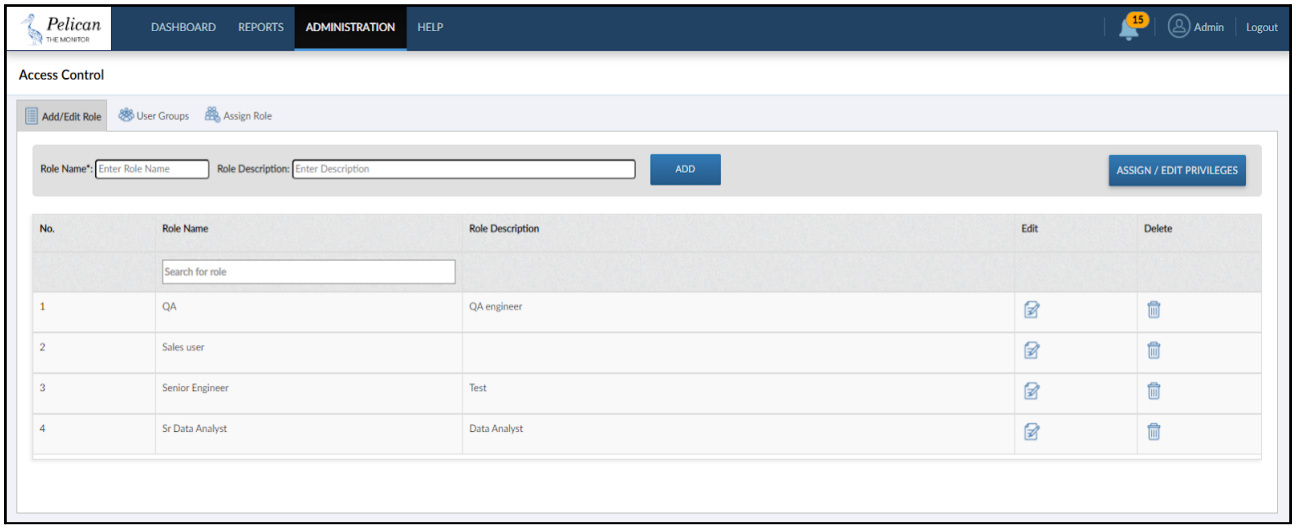
- Add a new role and editing an existing role
- Add user groups and assign users to the groups
- Editing or deleting existing user group
- Assigning or editing privileges
- Assigning role

5.6.4. Modify User

Admin can add new users and hence follow the steps as explained below.

Steps

1. Go to **Administration > User Management > Access Control**.
2. Click **ADD** in the *Add/Edit Role* tab after you Role Name and Role Description.



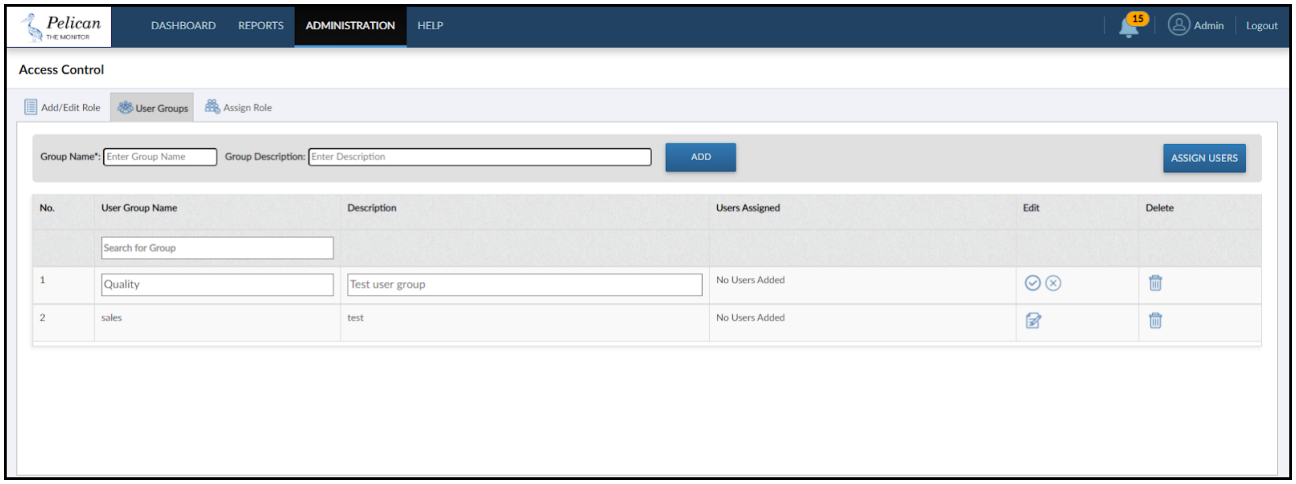
3. Click the edit icon for editing the Role Name and Role Description. Click  to save the changes.
4. Click the delete icon for deleting the Role Name and Role Description. The system displays a confirmation dialog box and then clicks **OK**.

5.6.5. Modify Groups

You can add a user group and assign users to the users to any specific group.

Steps for adding a new group and editing or deleting existing group

1. Go to **Administration > User Management > Access Control**.
2. In the *User Groups* tab, enter the **Group Name** and **Group Description**.
3. Click **ADD**.

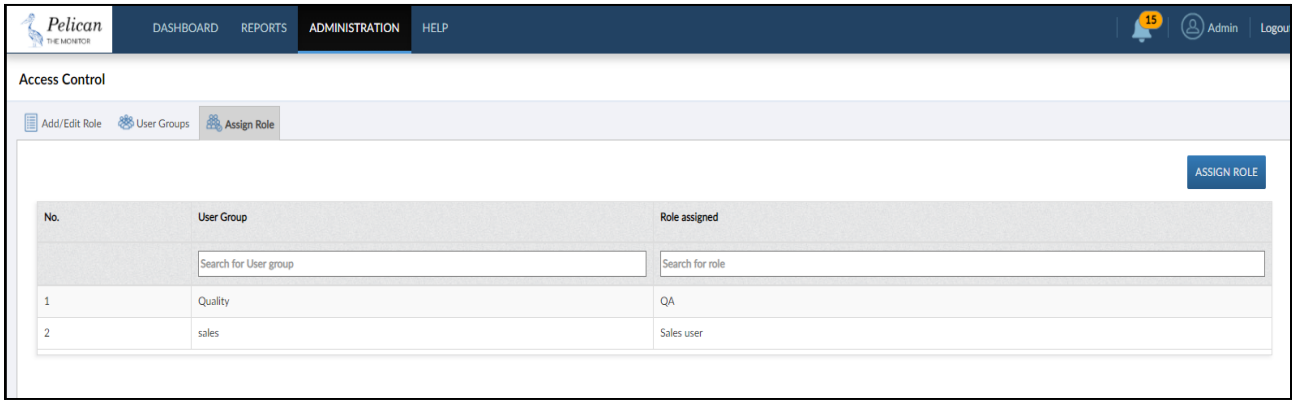


4. Click the edit icon for editing the User Group Name and Group Description. Click  to save the changes.
5. Click the delete icon for deleting the user group name and description. The system displays a confirmation dialog box and then clicks on **OK**.

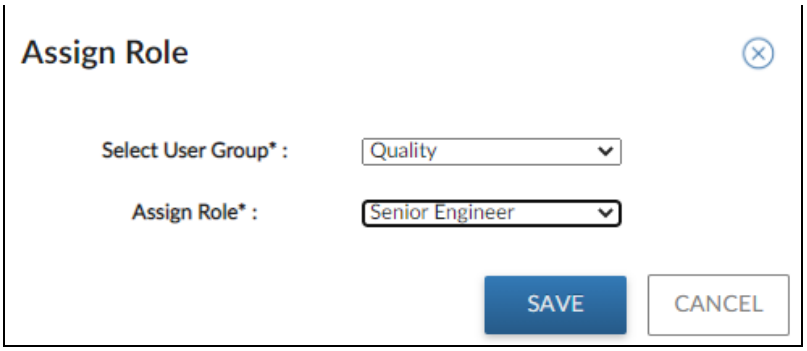
5.6.6. Assign Roles

You can assign specific roles to any group by following the below steps.

1. Go to **Administration > User Management > Access Control**.
2. In the *Assign Role* tab, enter the **User Group** and click **ASSIGN ROLE**.



3. Select User Group and Assign Role from the drop-down list.

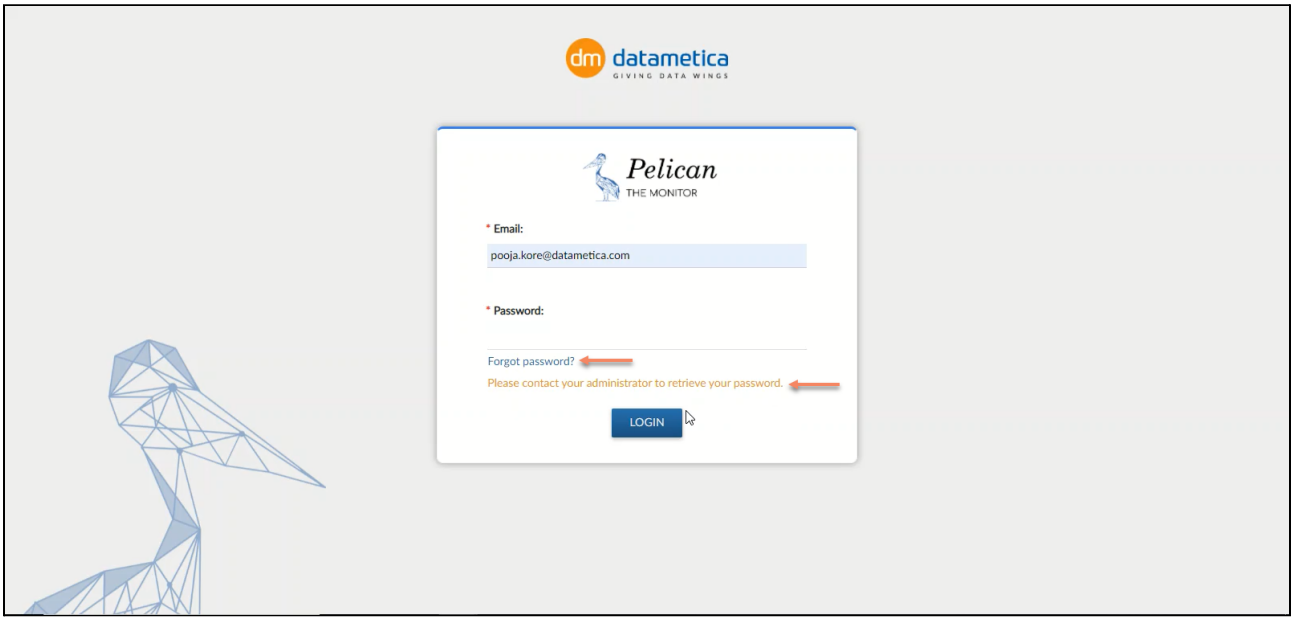


4. Click **SAVE**.

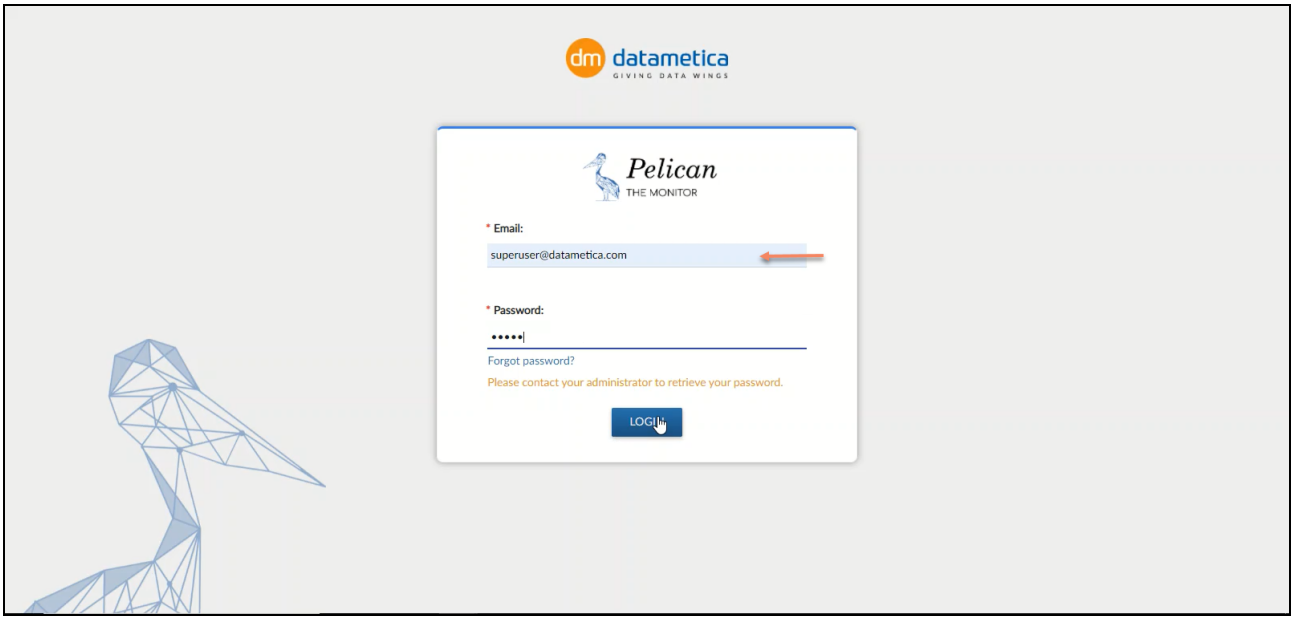
5.7. Forget Password

If you forget the password,for resetting the password, perform following steps.

Step 1: Click on the forget password.

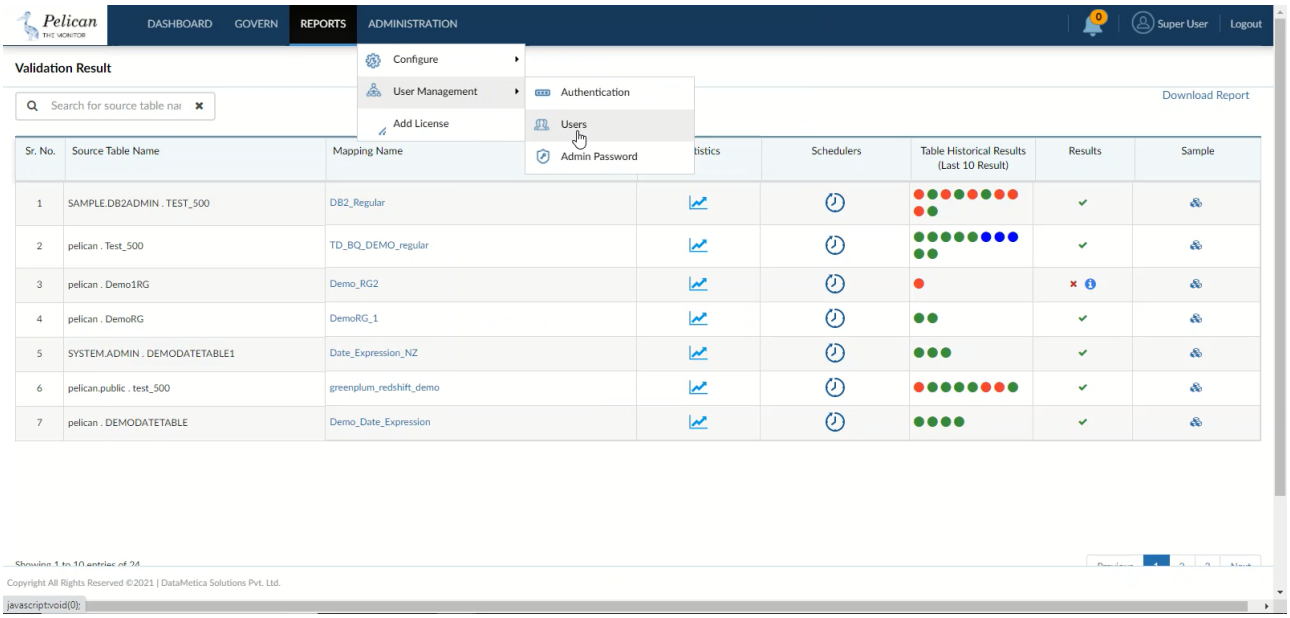


Step2: Contact one of the superusers or administrators and ask them to reset the password.



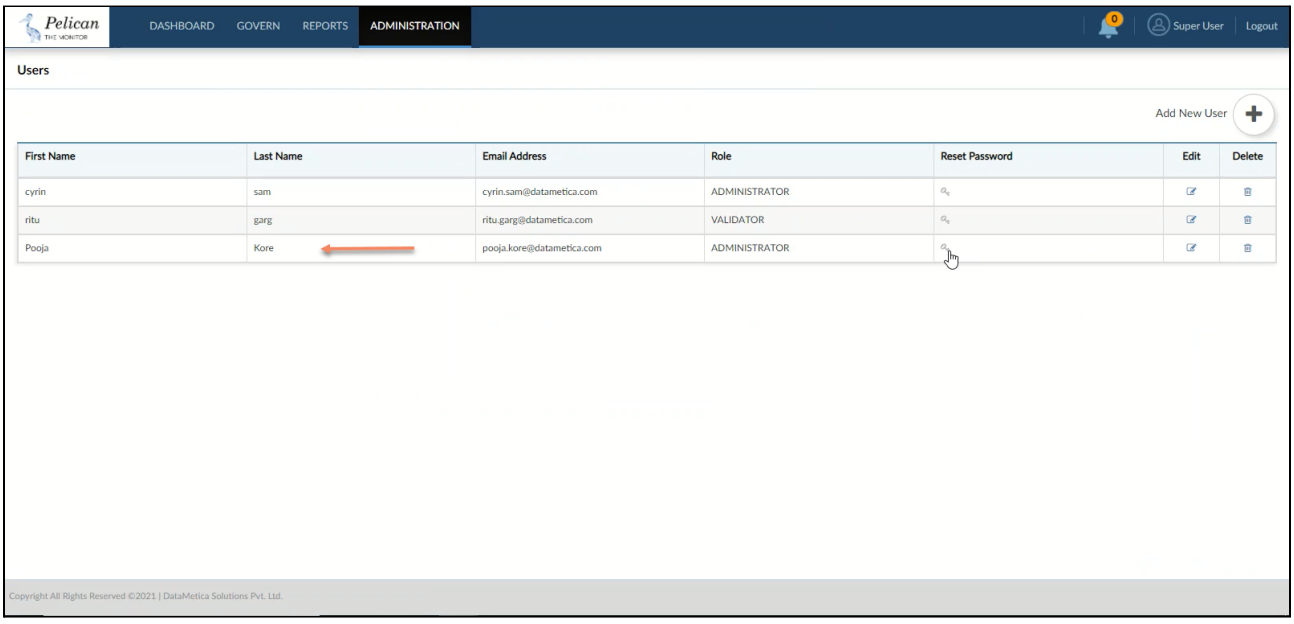
Step3: The administrator with the authority to reset the password will login in the system. Followed following steps,

Login → Administrator → User Management → Users.

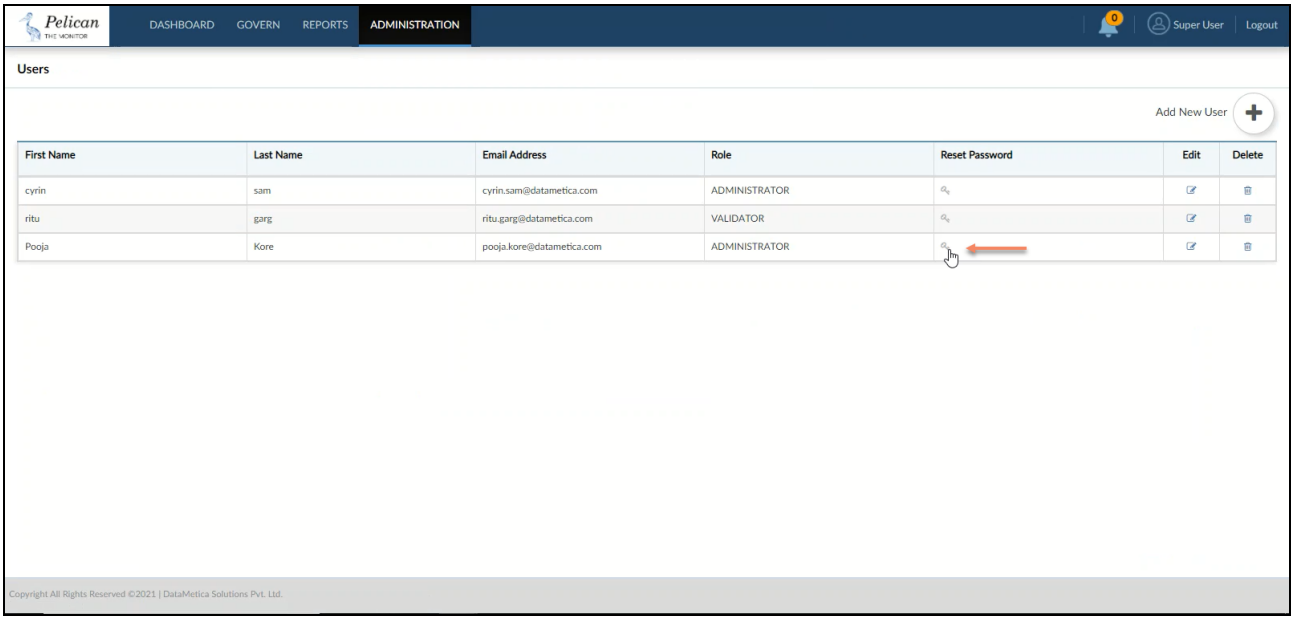


After clicking on the Users, the new screen will appear with the list of the users with the following parameter.

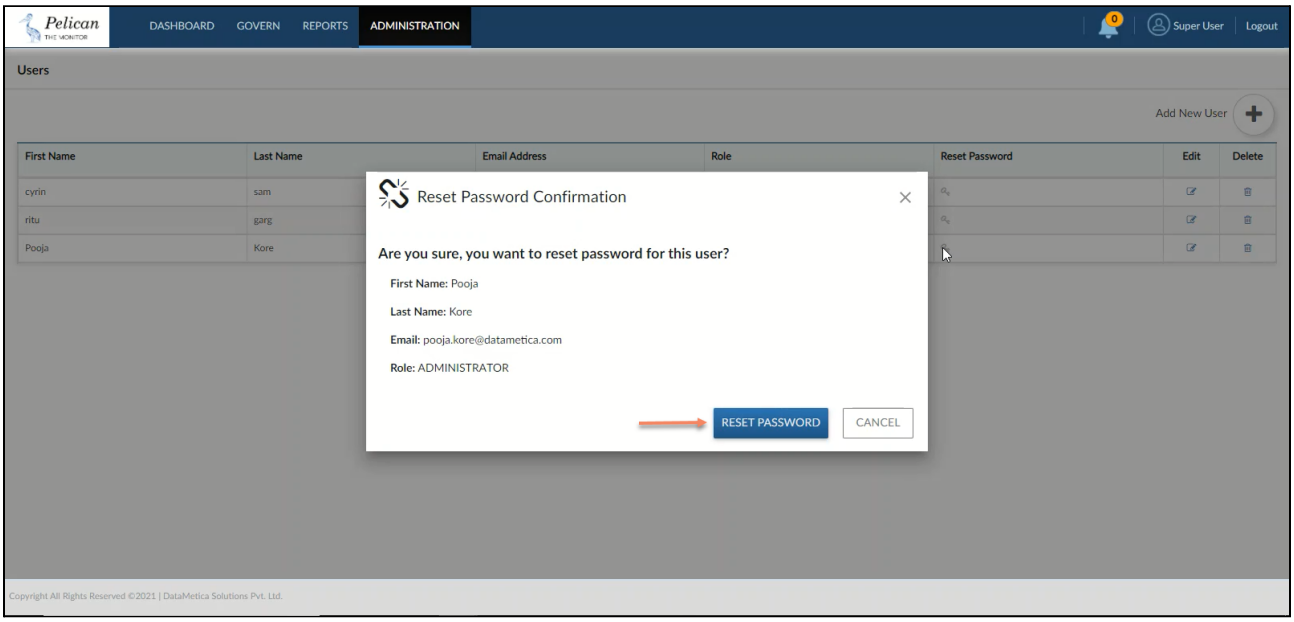
1. First Name
2. Last Name
3. Email Address
4. Role
5. Reset Password
6. Edit
7. Delete



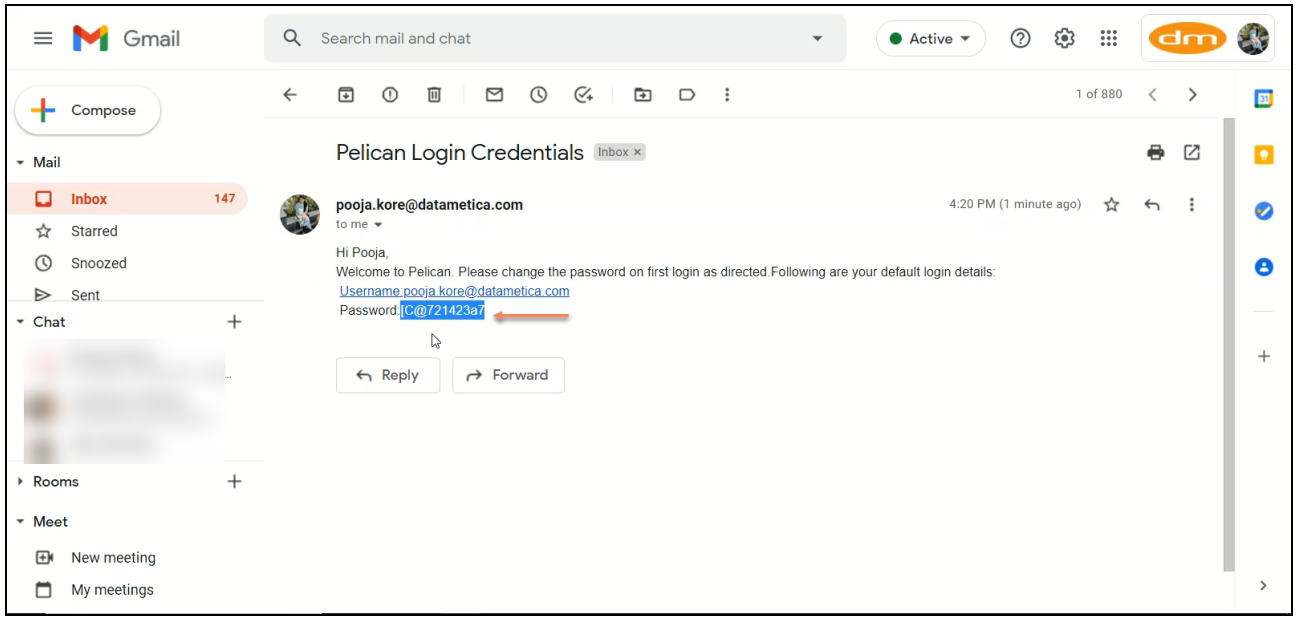
Step 4: The superuser has to look for the validator who has requested to reset the password. Then click on the reset icon.



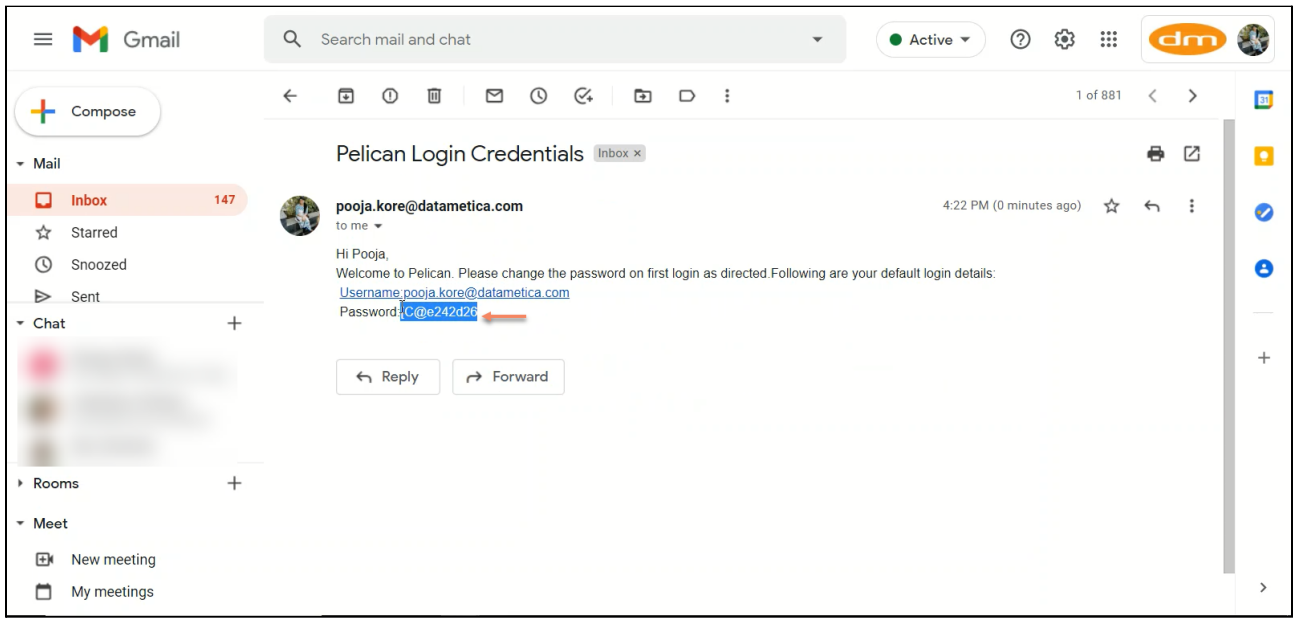
Step 5: The Reset Password Confirmation pop window will appear, it will ask for the confirmation to reset the password. Click on Reset Password.

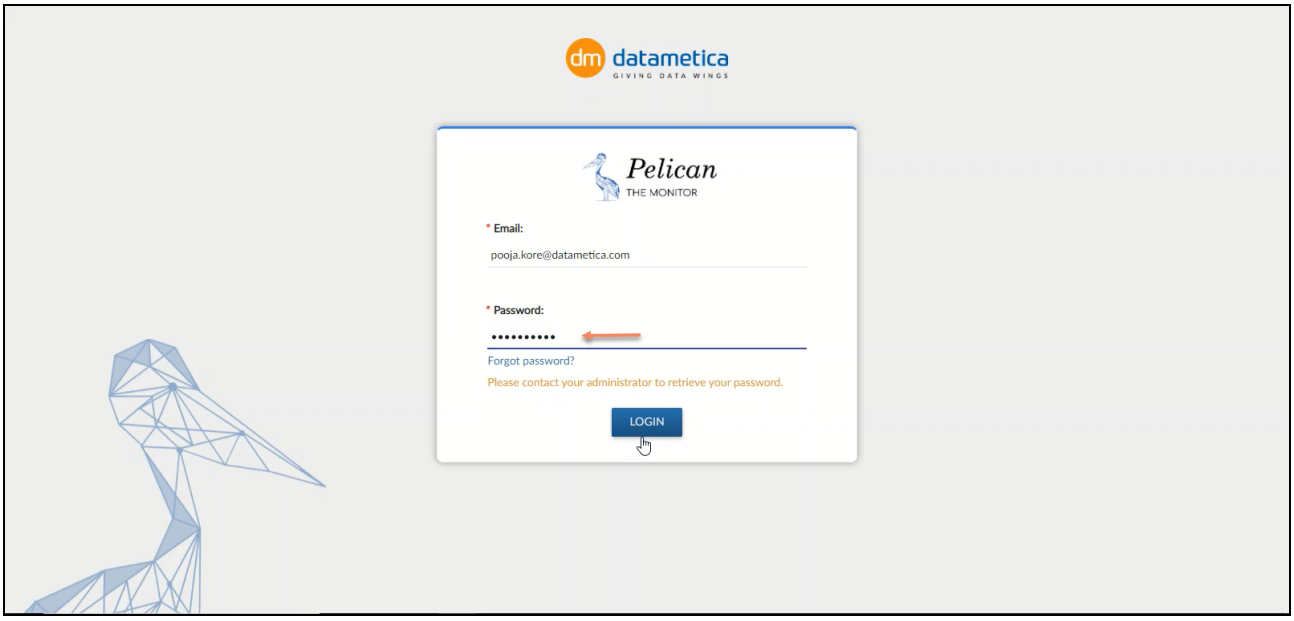


Step 6: The validator or superuser who has forgotten the password will receive an email with a new credential/ password.

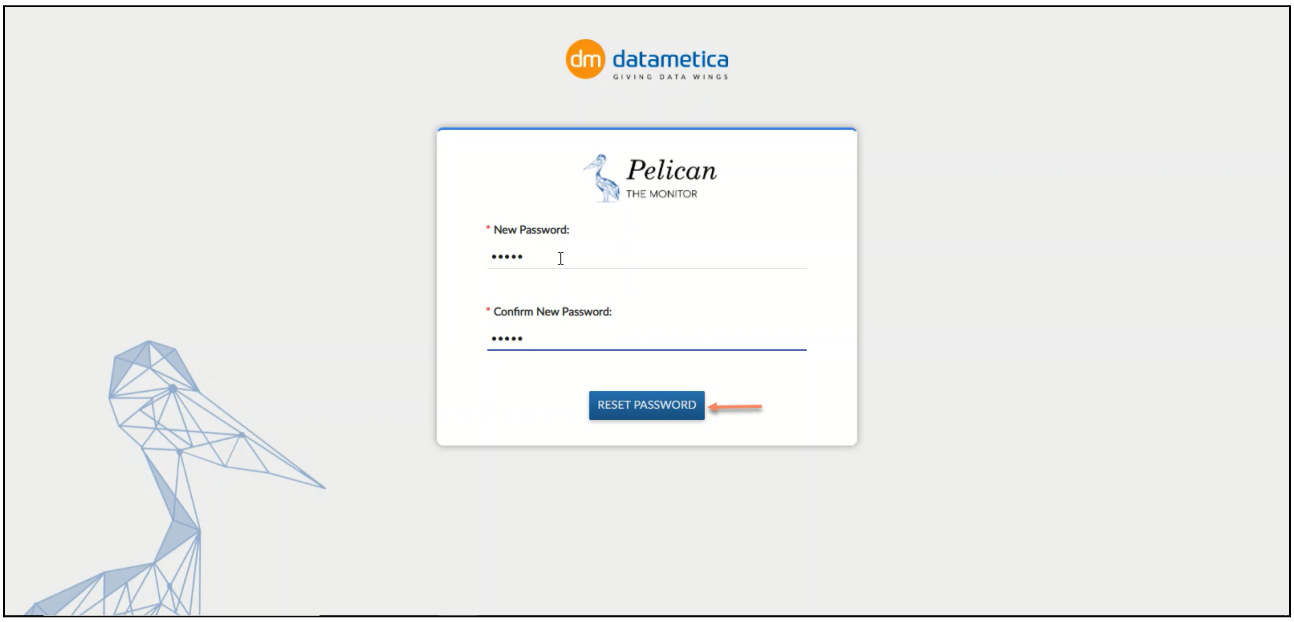


Step 7: Copy and paste the credentials form email to the login window and click on Login.

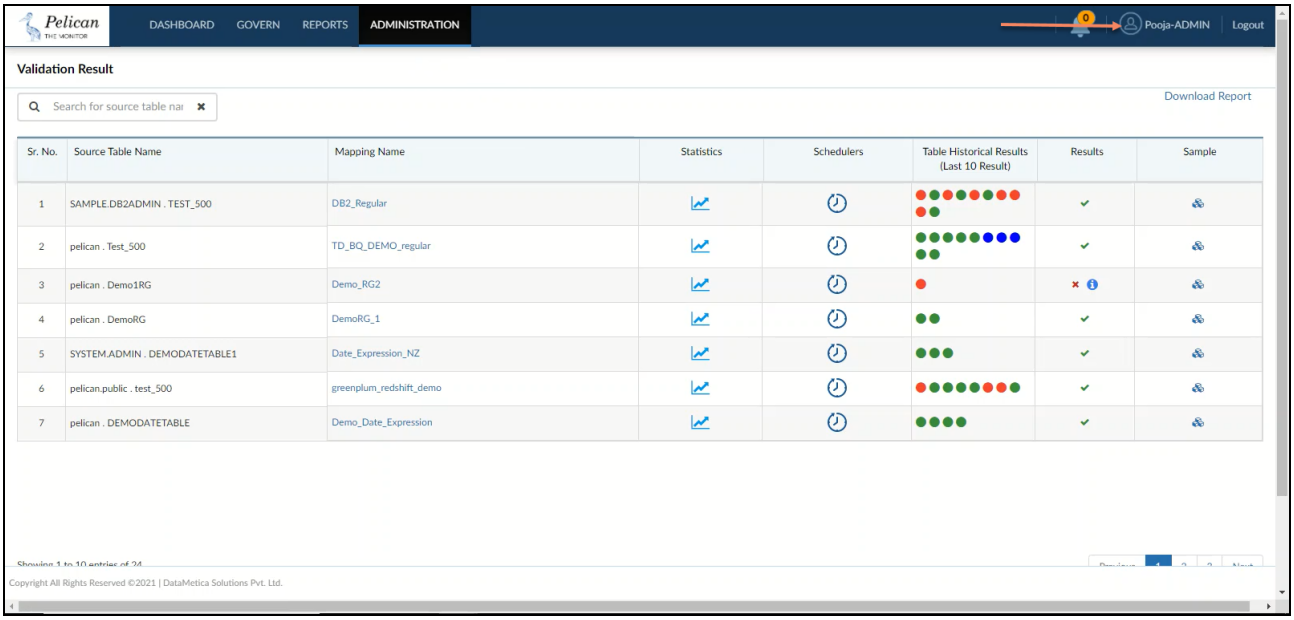




Step 8: The Reset Password window will appear, enter the New Password and Confirm New Password.



Step 9: Now login with a new Password. Pelican will be ready to use.



6. Datastore Configurtion

To compare and validate data across various data stores, first, you need to configure source and destination datastores in the Pelican. Pelican provides predefined data store types under which you can configure data stores as per the business requirement.

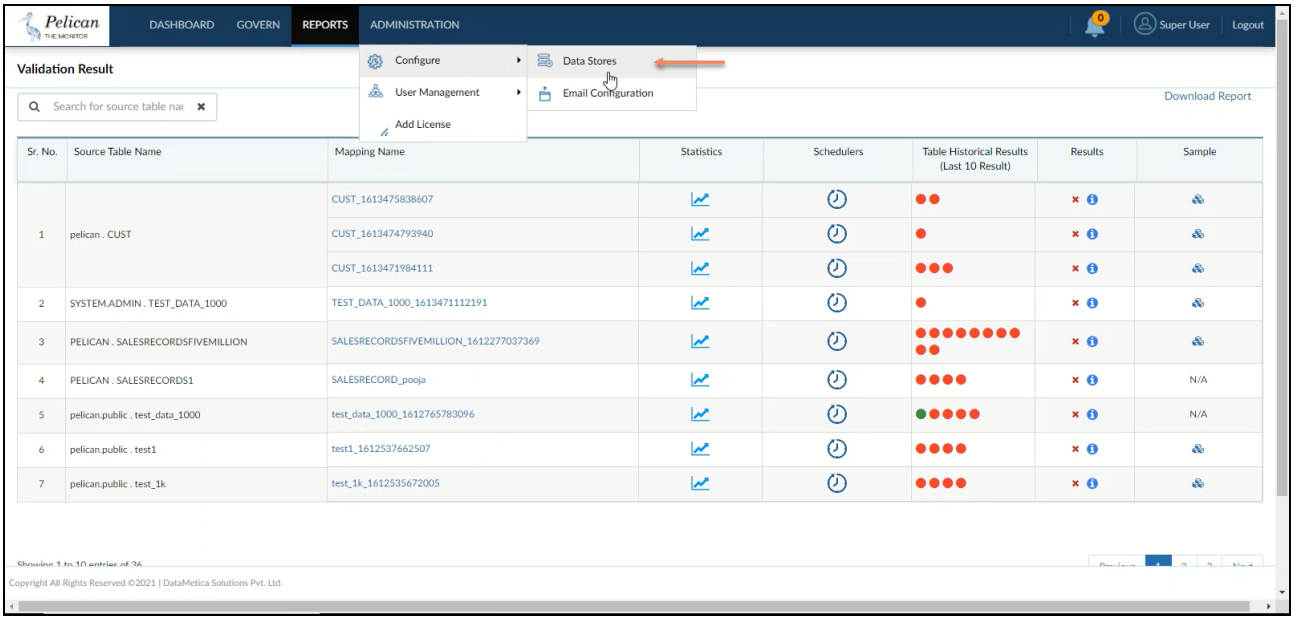
There are different data stores as listed in here.

- HIVE
- Netezza
- Teradata
- Oracle
- BigQuery
- DB2
- Snowflake
- Azure Synapse
- MS_SQL_Server
- Greenplum
- Redshift

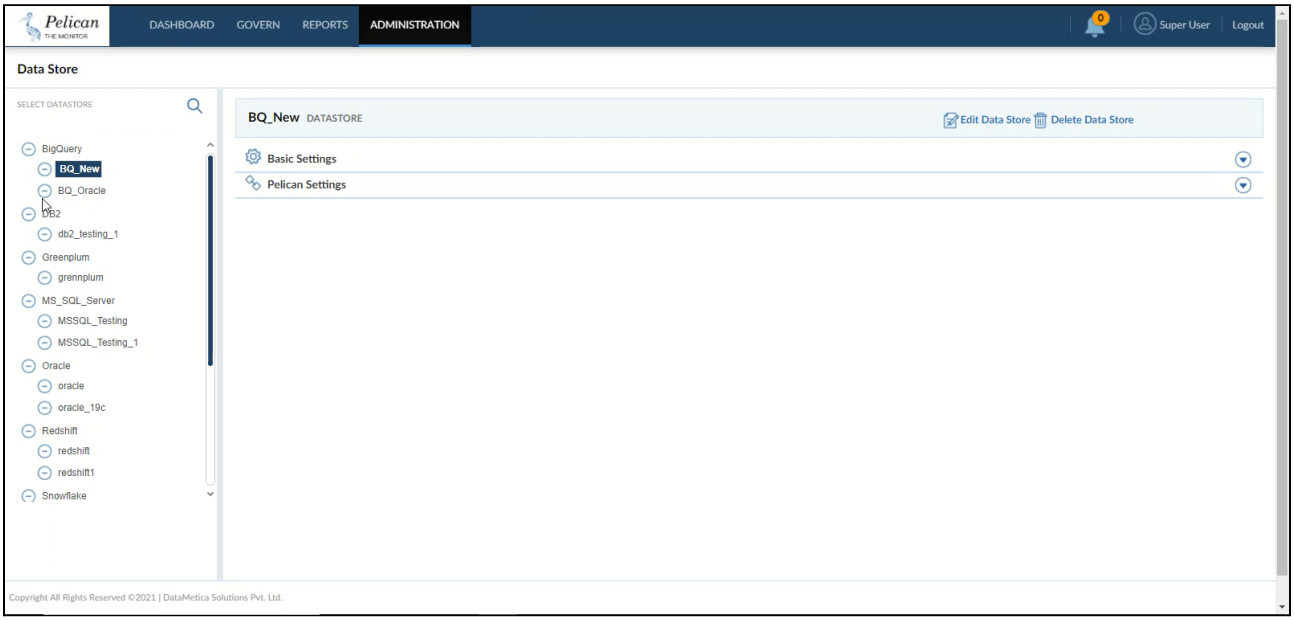
The data configuration is similar for all the data stores except HIVE which has few more steps.

Here, lets see how to do data configuration. To narrate the steps I have,

Go to **Administration** □ **Configure** □ **Data Store**.



This displays the **Data Store screen** as shown below.

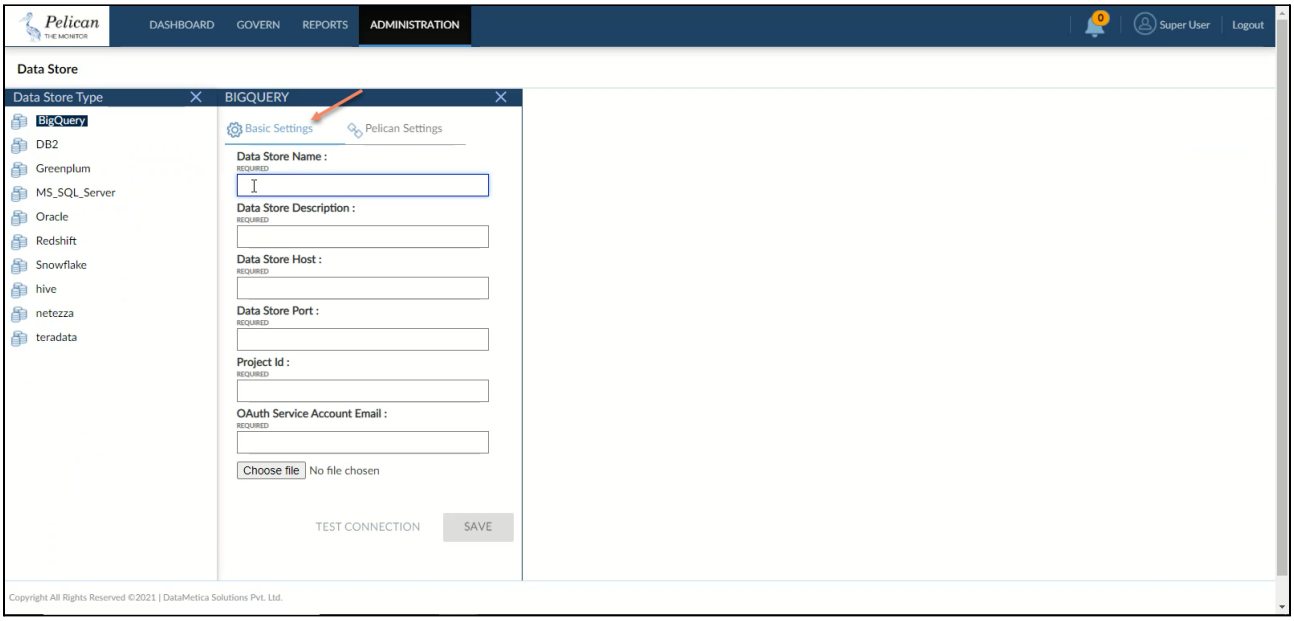


6.1. Steps for Database configuration

To perform the data store activities, you need to update the basic settings and pelican settings.

Step 1: To start the process, click on the datastore you want to create.

Step 2: Go to the Basic Setting tab.



Step 3: Enter the following information,

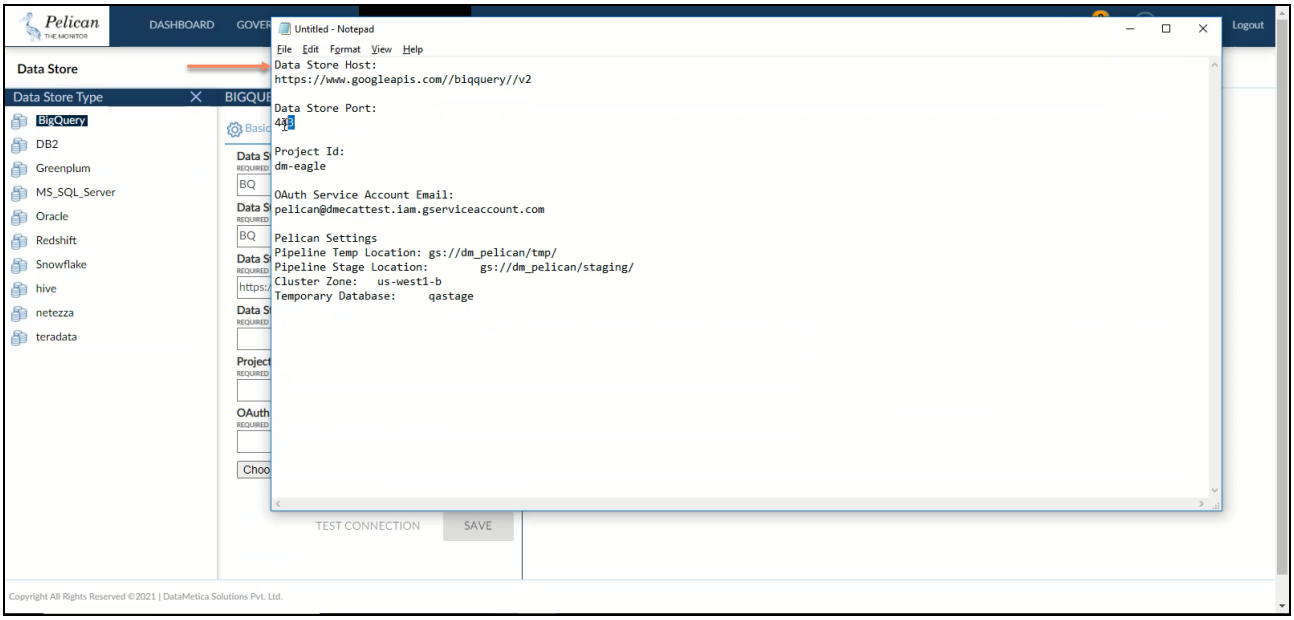
Basic setting:

1. Data Store Host: https://www.googleapis.com//biqquery//v2
2. Data Store Port: 443
3. Project Id: dm-eagle
4. O Auth Service Account Email: pelican@dmecattest.iam.gserviceaccount.com

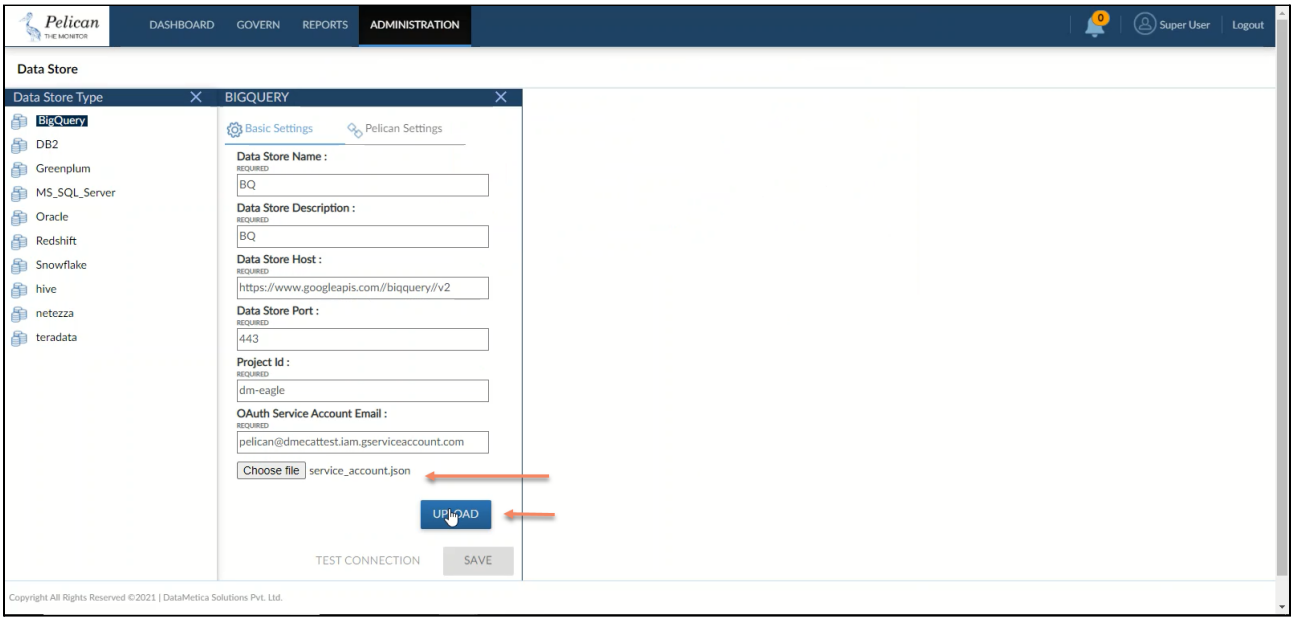
Pelican setting:

1. Pipeline Temp Location: gs://dm_pelican/tmp/
2. Pipeline Stage Location: gs://dm_pelican/staging/
3. Cluster Zone: us-west1-b
4. Temporary Database: qastage

Step 4: Add the Data Store Name and Data Store Description.

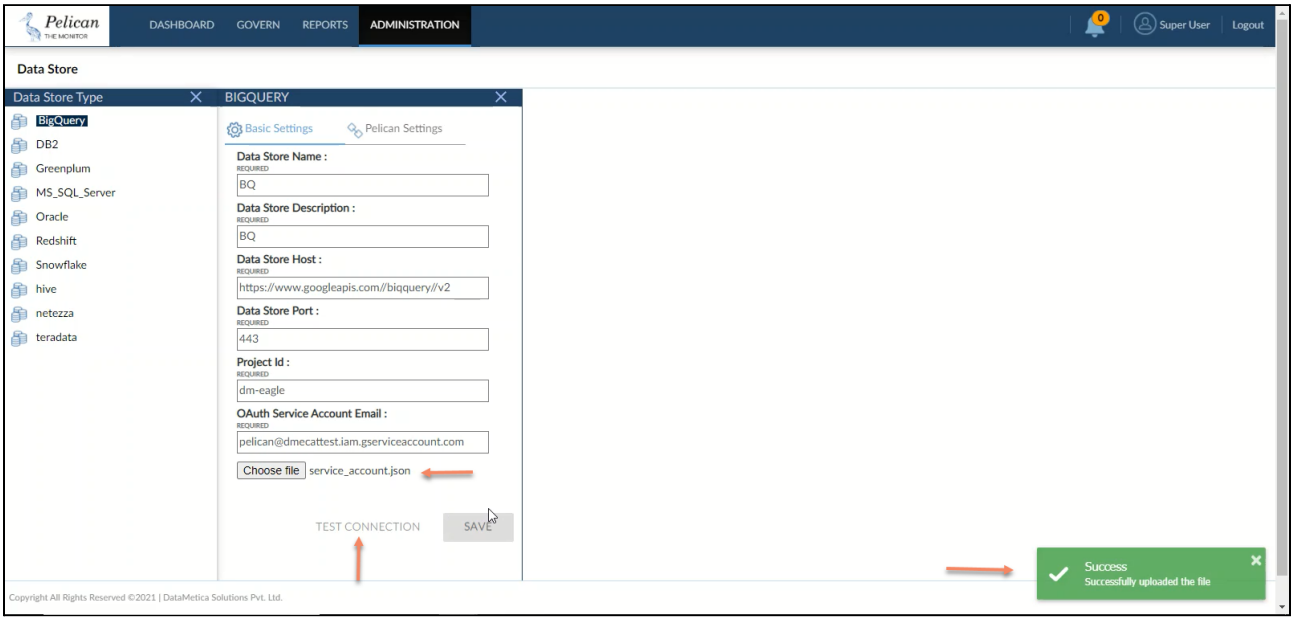


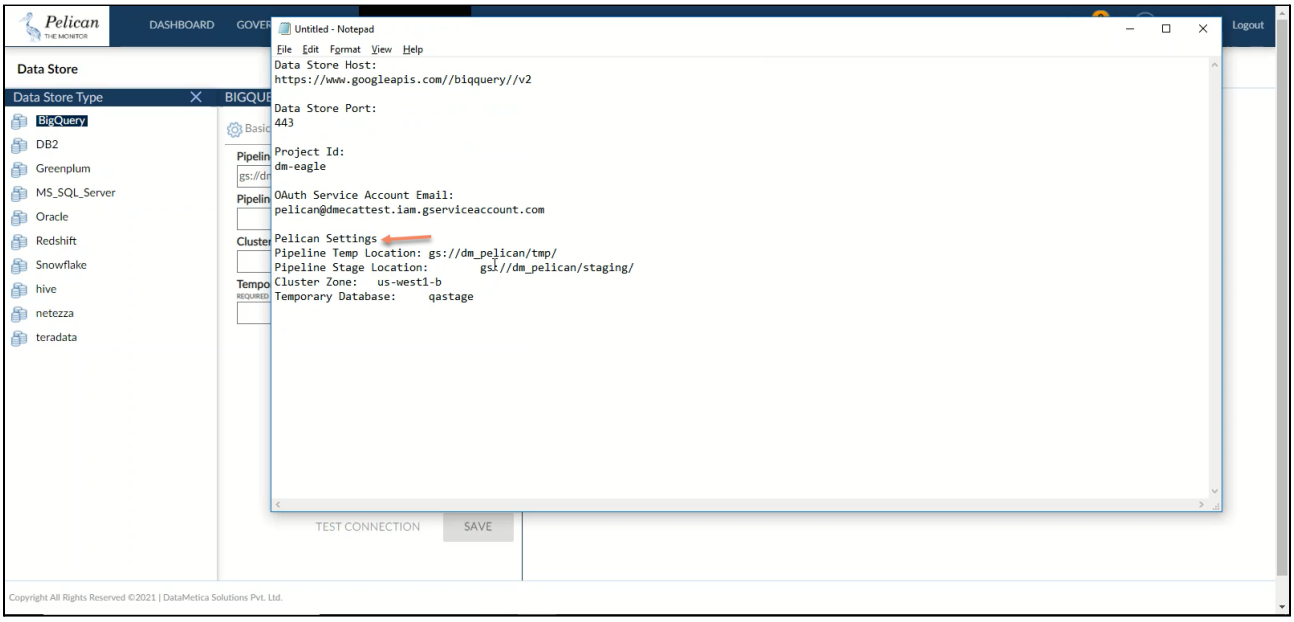
Step 5: Upload the service_account.json file.



Once you upload the file **service.account.json** and add all the details in the field, then click on the **TEST CONNECTION**. If everything is correct then it displays the Success message and enables the Save button.

Step 6: Click on the **Save** button, to save the basic settings.

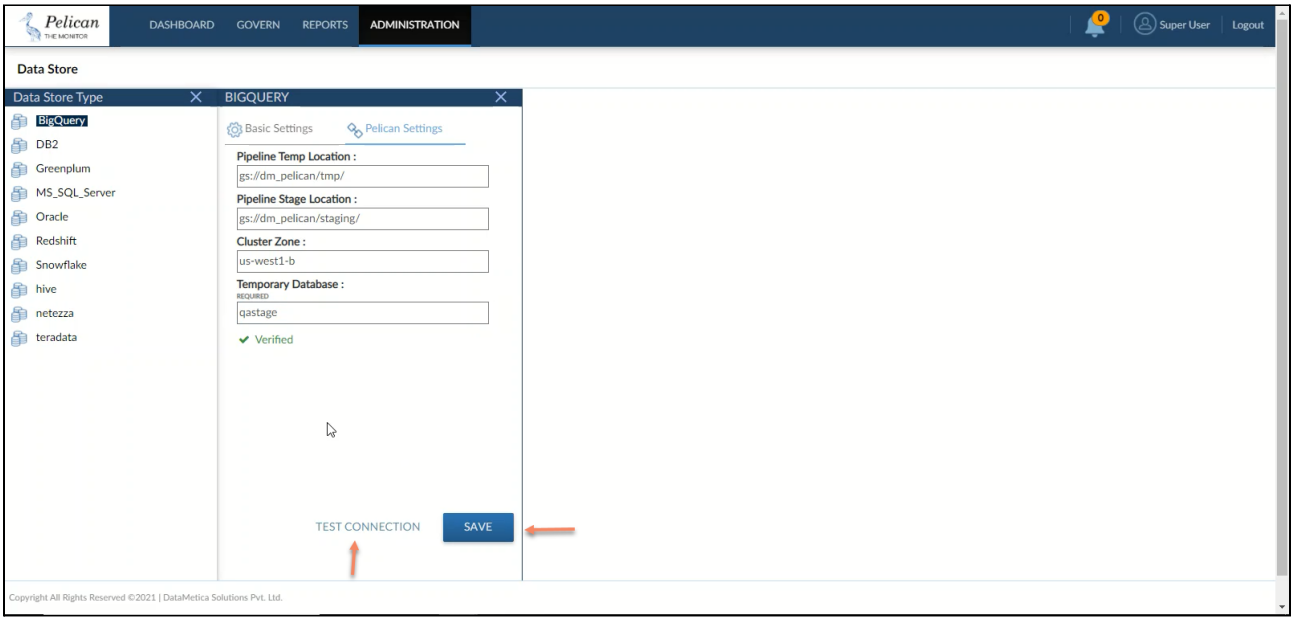




Step 6: Click on the **Pelican settings** button and enter the field details.

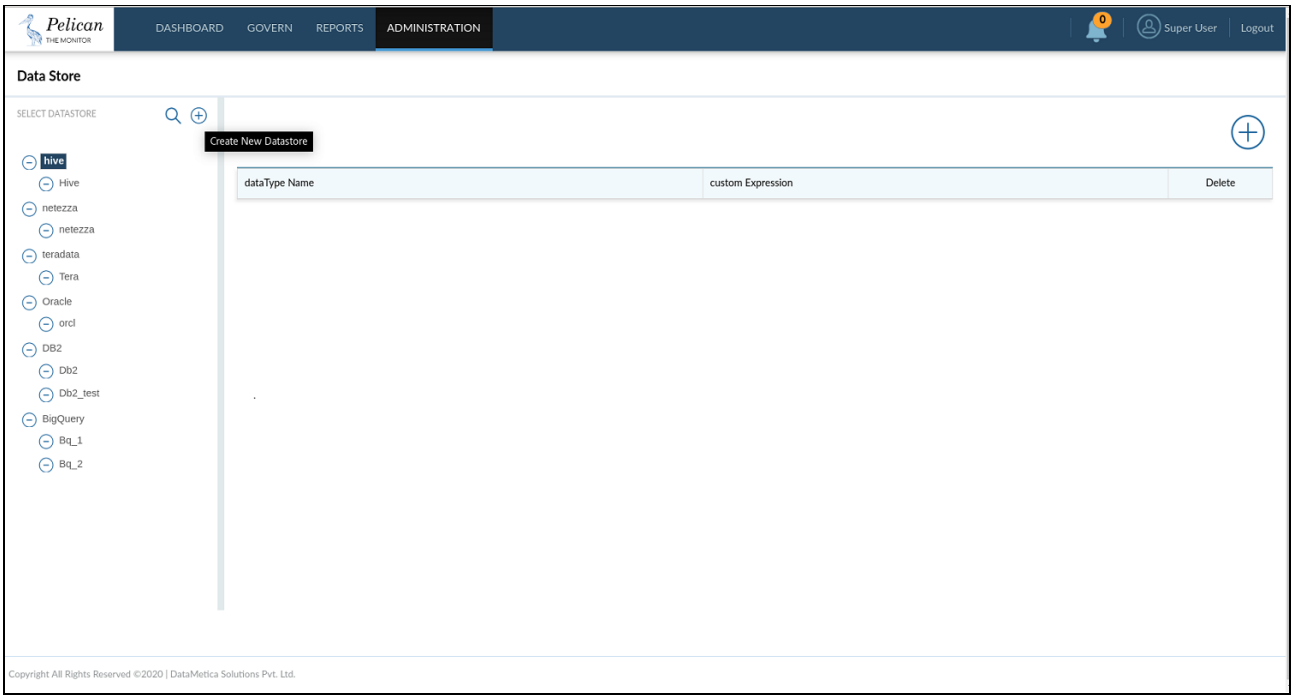
Step 7: Click on the test connection. If the connection is correct then the message will pop up as **Verified** in green color. Once the connection is verified, the **Save** button will be enabled.

S



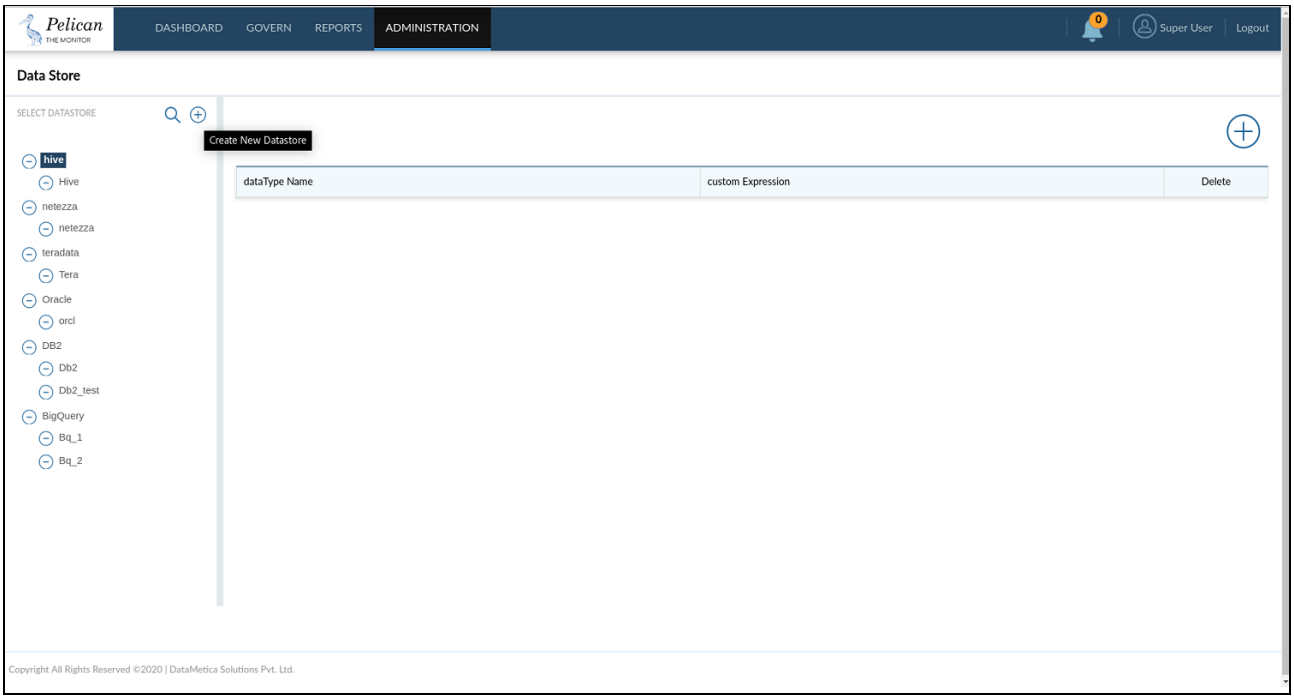
Steps to configure data store :

1. Go to **Administration** □ **Configure** □ **Datastore**. This displays the Data Store screen as shown below.

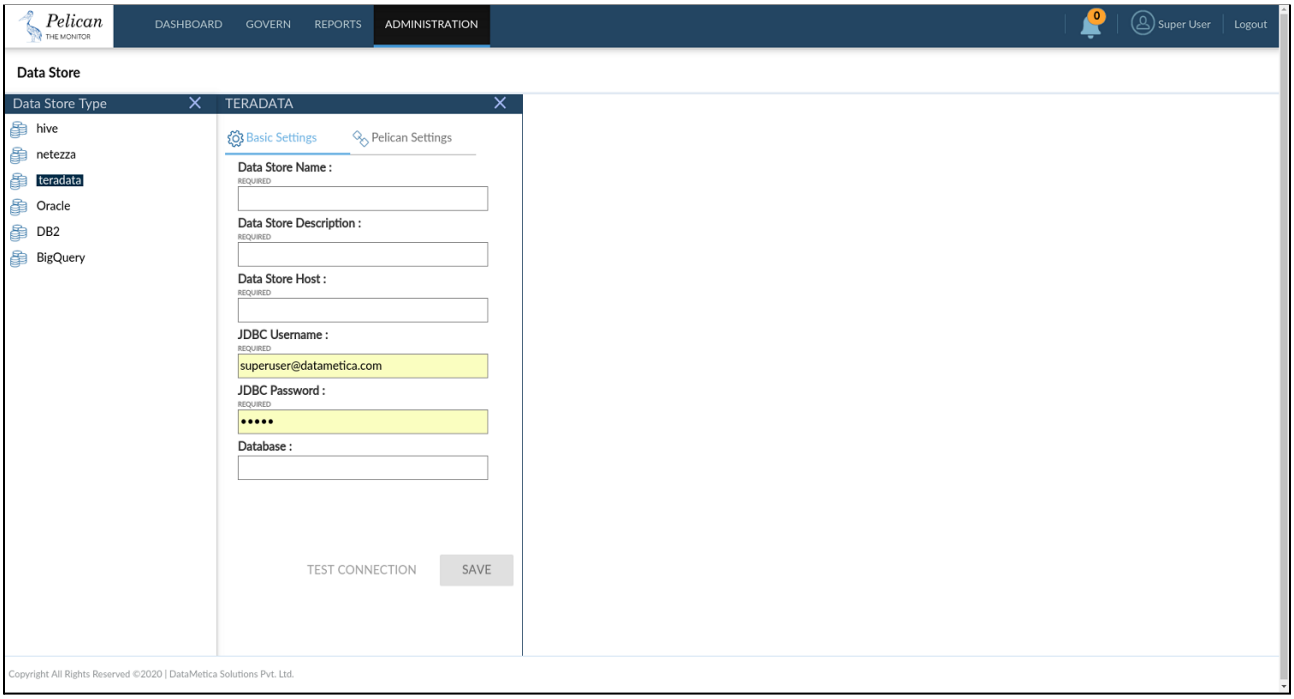


The Data Store screen displays predefined data stores at the left side of the screen under which you can create one or more databases.

2. Select **Datastore Type** and click the **Create New Datastore**  icon



3. Enter information in the various fields to configure the data store and click **Save**.



4. The newly created data store instance will be listed under the respective data store type as shown below.

Similarly, you can configure any number of data store instances under the predefined data store type. The information to be entered while configuring the data store will vary from data store to data store.

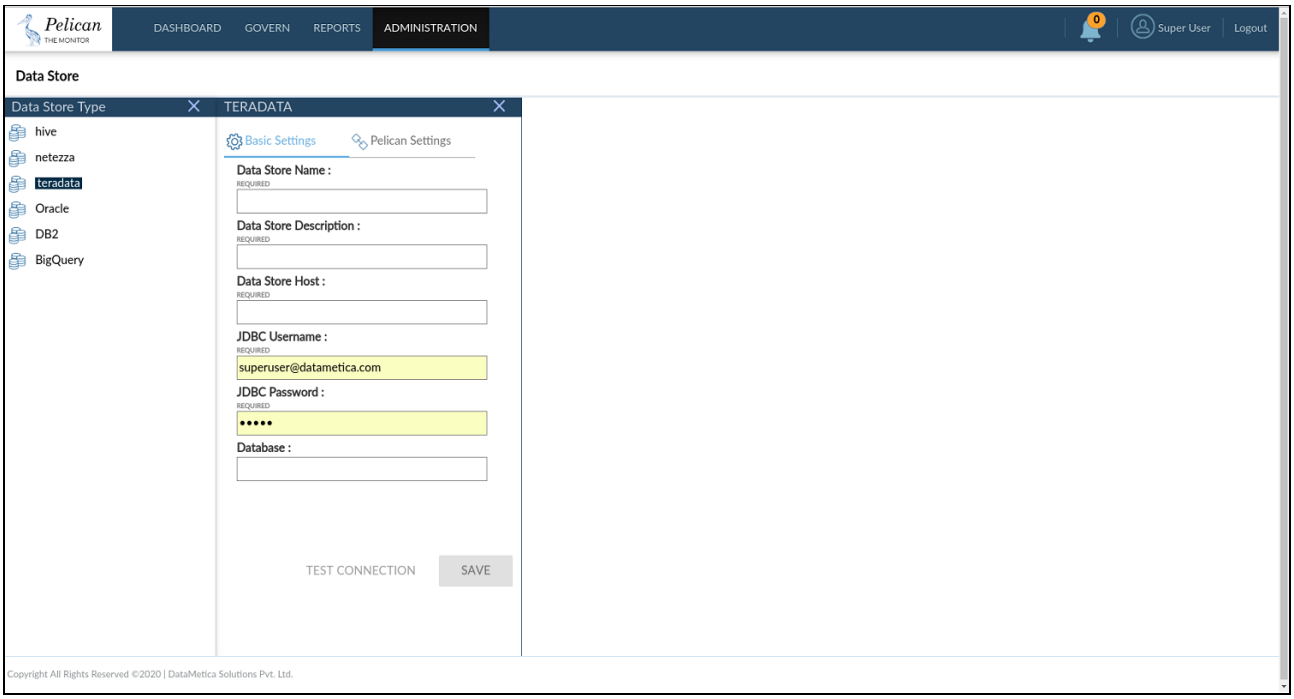
For example, while configuring HIVE Datastore, you need to enter information in Basic Settings, Security Settings, JDBC Settings, Metadata Settings, Report Settings, Lineage Settings, and Pelican Settings sections; whereas while configuring most of the data stores except HIVE you need to enter information in Basic Settings and Pelican Settings sections.

Additionally, the application provides Kerberos and SASL (Simple Authentication and Security Layer) authentication security support in the HIVE data store.

Use the above-mentioned steps to configure source and destination data store by entering the connectivity details and other necessary details.

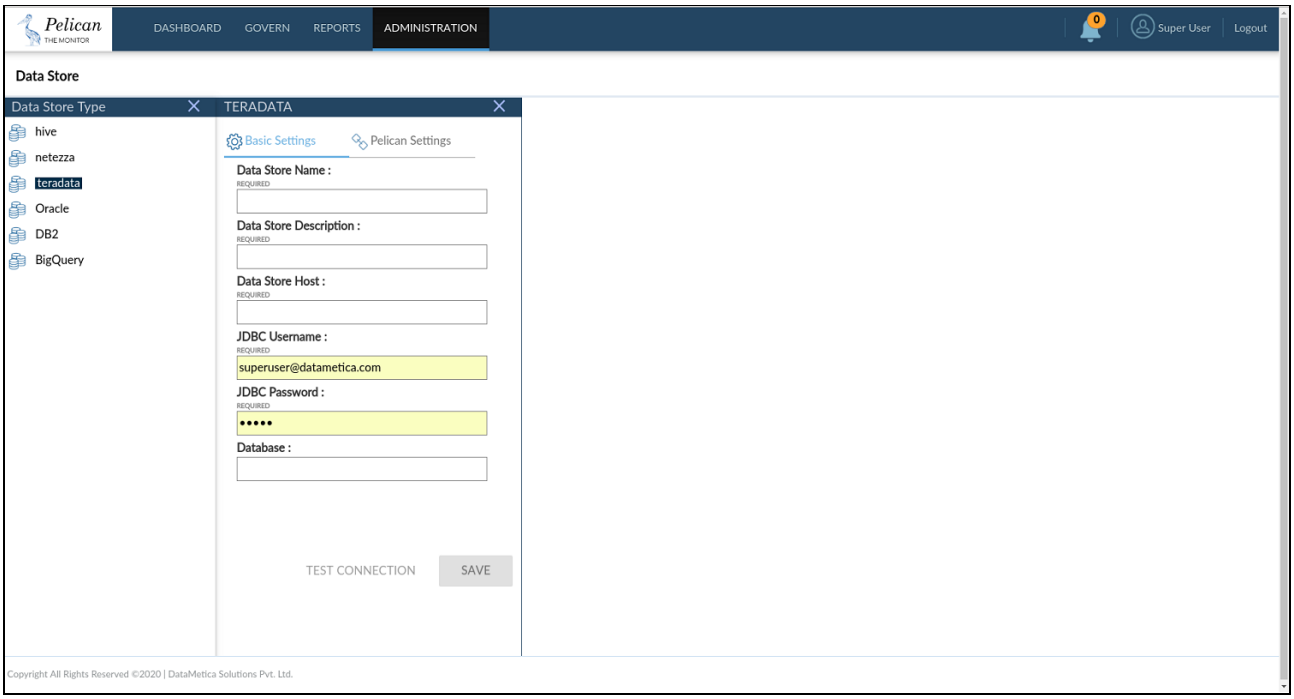
6.2. Steps for Teradata Configuration

1. Go to **Administration > Configure > Datastore**.
2. At the left-hand side pane, a list of predefined data store lists is visible so select the **teradata** in the data store field or enter the data store. Click  icon.



3. Enter all the details in the Basic Settings tab fields and click **SAVE**.
   You may view the created data store in the list under the respective data store type. Similarly, you can configure any number of data store instances under the predefined data store type. The information to be entered while configuring the data store will vary for each data store.

   Information below does not fit the Section as in this section we should only concentrate on Teradata Settings.



Basic Settings Field Descriptions

I have added descriptions for each field as clients are more interested in what to fill in the field. Please add descriptions for other data stores as well most of them are similar.

Field	Description
Data Store Name	The name the user wants given to the data store.
Data Store Description	Description for the data store.
Data Store Host	IP address of the data store host machine.
Data Store Port	JDBC port to connect to the data store.
JDBC Username	JDBC username using which Pelican will connect to the data store.
JDBC Password	JDBC password for the provided JDBC user.

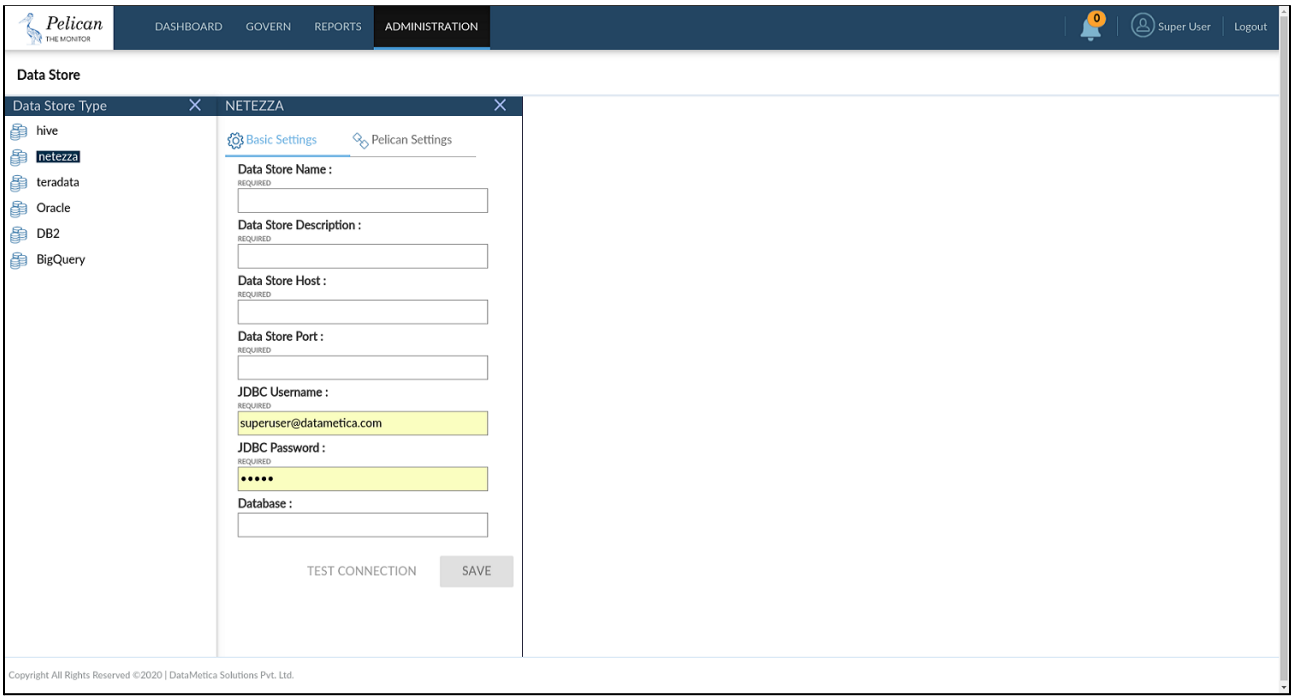
Database	Default database to connect.

Pelican Settings Field Description

Field	Description
Temporary Database	Database used by Pelican for its functioning.
Supplementary DataStore Location	Writable Path on the machine where Pelican is installed.

6.3. Steps for Netezza Configuration

1. Go to **Administration > Configure > Datastore**.
2. At the left-hand side pane, a list of predefined data store lists is visible so select the **netezza** the data store field or enter the data store. Click  icon.



The Data Store screen displays predefined data stores at the left side of the screen under which you can create one or more databases.

3. Enter information in the various fields to configure the data store and click **SAVE**. You may view the created data store in the list under the respective data store type.

Similarly, you can configure any number of data store instances under the predefined data store type. The information to be entered while configuring the data store will vary from with different data stores.

Basic Settings Field Descriptions

Field	Description
Data Store Name	The name the user wants given to the data store.

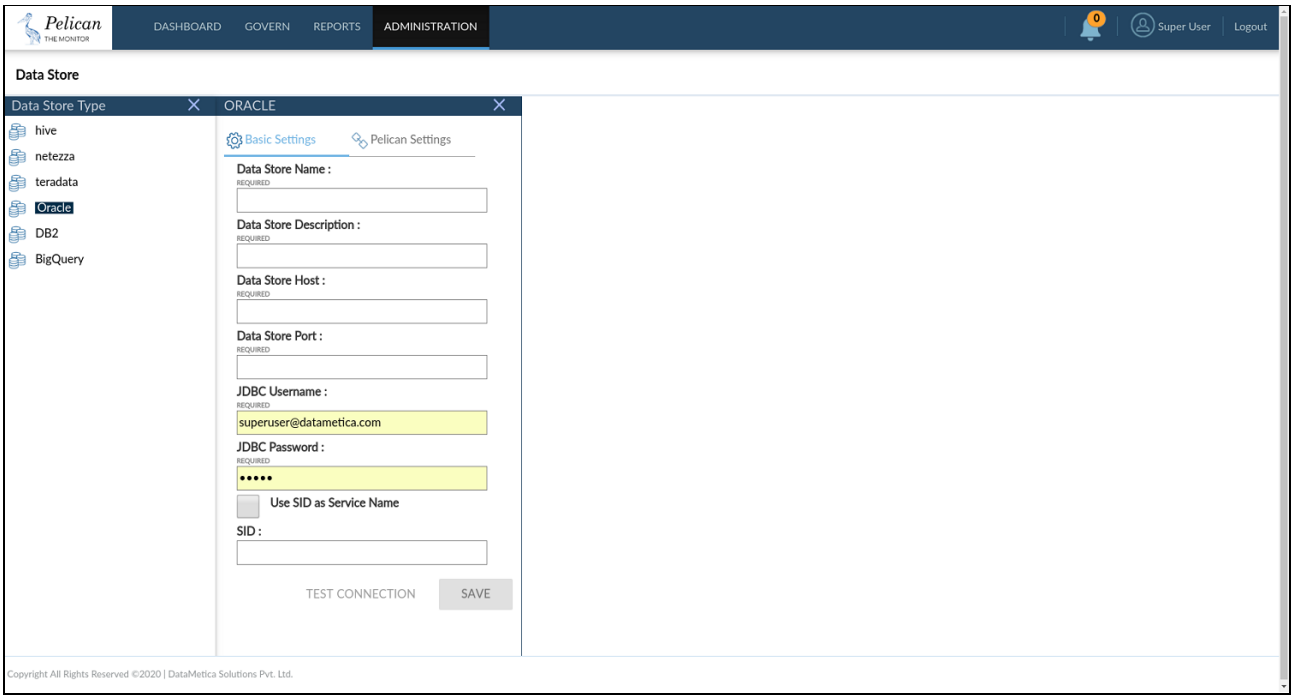
Data Store Description	Description for the data store.
Data Store Host	IP address of the data store host machine.
Data Store Port	JDBC port to connect to the data store.
JDBC Username	JDBC username using which Pelican will connect to the data store.
JDBC Password	JDBC password for the provided JDBC user.
Database	Default database to connect.

Pelican Settings Field Description

Field	Description
Temporary Database	Database used by Pelican for its functioning.
Supplementary DataStore Location	Writable Path on the machine where Pelican is installed.

6.4. Steps for Oracle Configuration

1. Go to **Administration > Configure > Datastore**.
2. At the left-hand side pane, a list of predefined data store lists is visible so select the **Oracle** the data store field or enter the data store. Click  icon.



3. Enter all the details in the Basic Settings tab fields and click **SAVE**.

   You may view the created data store in the list under the respective data store type. Similarly, you can configure any number of data store instances under the predefined data store type. The information to be entered while configuring the data store will vary from with different data stores.

Basic Settings Field Descriptions

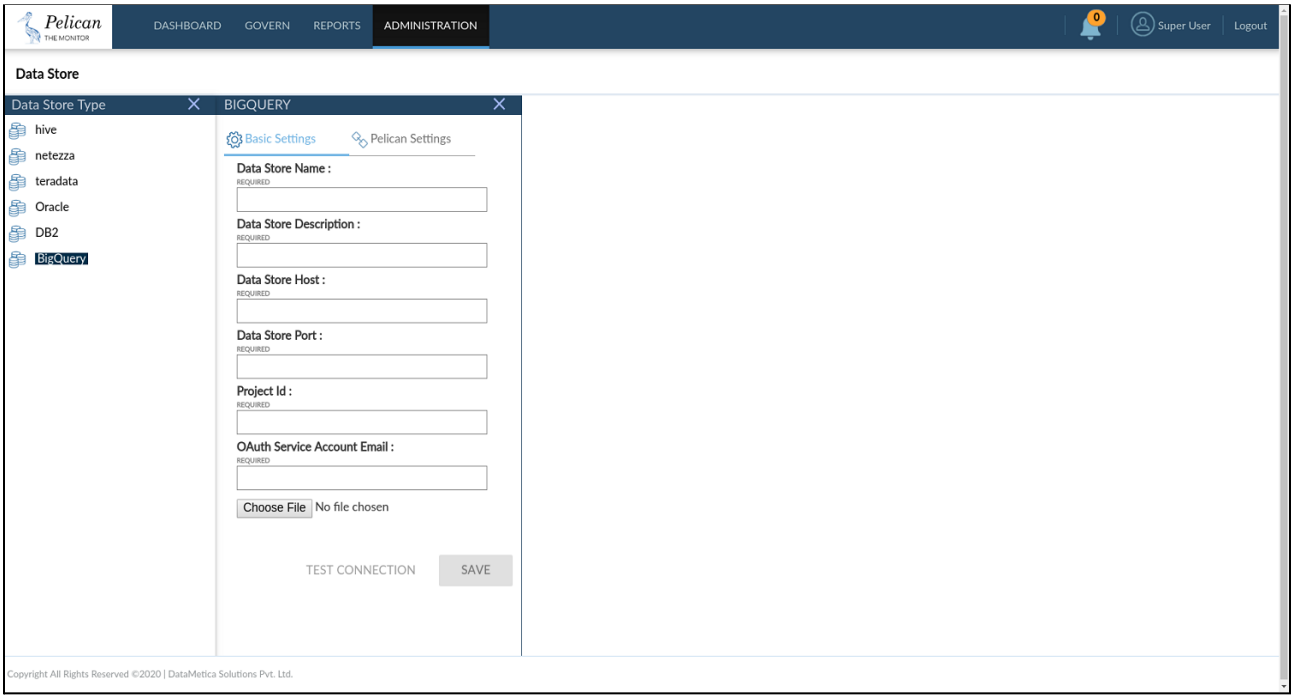
Field	Description
Data Store Name	The name the user wants given to the data store.
Data Store Description	Description for the data store.
Data Store Host	IP address of the data store host machine.
Data Store Port	JDBC port to connect to the data store.
JDBC Username	JDBC username using which Pelican will connect to the data store.
JDBC Password	JDBC password for the provided JDBC user.
Use SID as Service Name	If service name is provided for an Oracle instance, then select this checkbox and enter service name value in the SID textbox.
SID	The **SID** is a site identifier.

Pelican Settings Field Description

Field	Description
Temporary Database	Database used by Pelican for its functioning.
Supplementary DataStore Location	Writable Path on the machine where Pelican is installed.

6.5. Steps for BigQuery Configuration

1. Go to **Administration > Configure > Datastore**.
2. At the left-hand side pane, a list of predefined data store lists is visible so select the **BigQuery** the data store field or enter the data store. Click  icon.



3. Click **Choose File**. And provide the path for google cloud service account key json file.
4. Enter all the details in the Basic Settings tab fields and click **SAVE**.
   You may view the created data store in the list under the respective data store type. Similarly, you can configure any number of data store instances under the predefined data store type. The information to be entered while configuring the data store will vary for each data store.

Basic Settings Field Descriptions

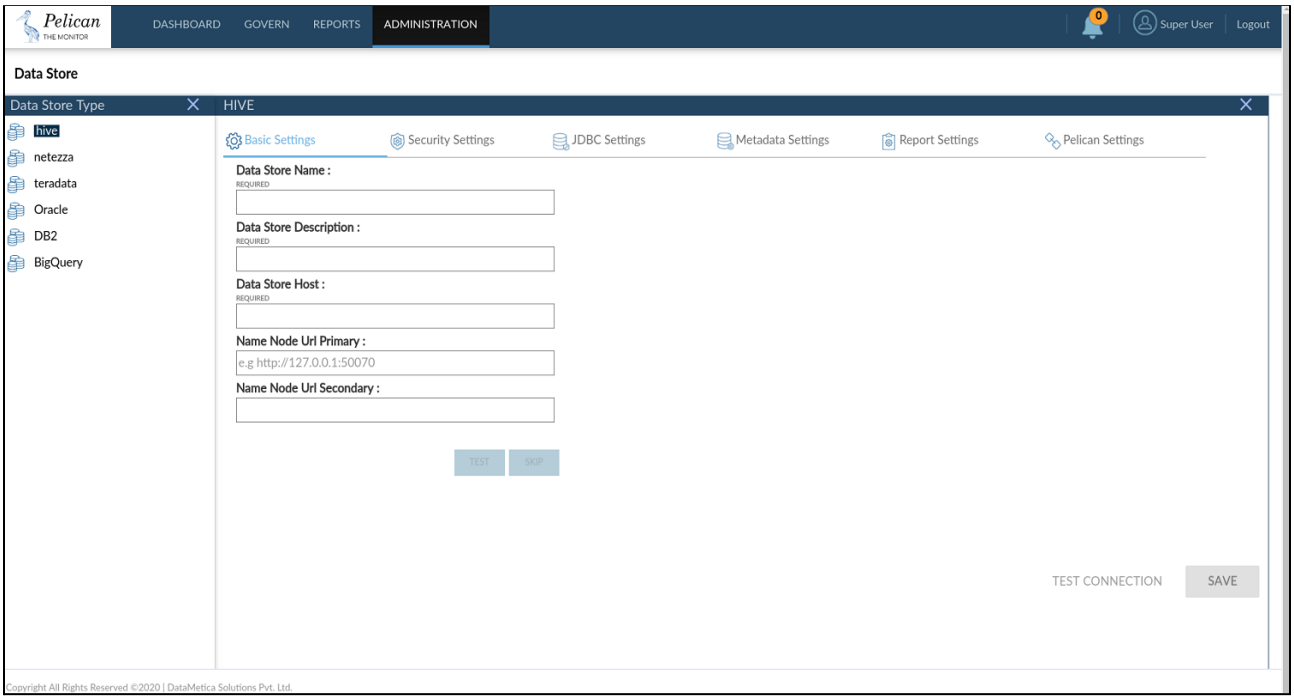
Field	Description
Data Store Name	The name the user wants given to the data store.
Data Store Description	Description for the data store.
Data Store Host	IP address of the data store host machine.
Data Store Port	JDBC port to connect to the data store.
Project Id	BigQuery Project to which Pelican will connect,
OAuth Service Account Email	Service account email for authentication.

Pelican Settings Field Description

Field	Description
Pipeline Temp Location	
Pipeline Stage Location	
Cluster Zone	
Temporary Database	

6.6. Steps for HIVE Configuration

1. Go to **Administration > Configure > Datastore**.
2. At the left-hand side pane, a list of predefined data store lists is visible so select the **HIVE** the data store field or enter the data store. Click  icon.



3. Enter all the details in the: Basic Settings, Security Settings, JDBC Settings, Metadata Settings, Report Settings, and Pelican Settings fields.
4. Click **SAVE**.

   You may view the created data store in the list under the respective data store type. Similarly, you can configure any number of data store instances under the predefined data store type. The information to be entered while configuring the data store will vary from with different data stores.

Basic Settings Field Descriptions

Field	Description
Data Store Name	The name the user wants given to the data store.
Data Store Description	Description for the data store.
Data Store Host	IP address of the data store host machine.
Name Node Url Primary	URL of namenode required to get cluster configurations.
Name Node Url Secondary	URL of secondary namenode

Security Settings Field Description

Field	Description
Kerberos Enabled	Check box mark it true for kerberized Hive.
Use Sasl	Check box based if sasl is required
User Kerberos Principal	Pattern will be like user@organization.com.
Kerberos Service Principal for JDBC	A Kerberos principal is a unique identity to which Kerberos can assign tickets.
Kerberos Service Principal for Meta Store	HDFS meta store path

User Keytab Location	User keytab file location in case of Kerberos Enable.
SASL QOP Enable	Checkbox to enable SASL QOP
SASL QOP	**SASL** Mechanisms
Kerberos Enabled for JDBC	Can be checked true if Kerberos Enable is true.
Kerberos Enabled for Metastore	Can be checked true if Kerberos Enable is true.
Kerberos Enabled for Name node	Can be checked true if Kerberos Enable is true.
Kerberos Enabled for Resource Manager	Can be checked true if Kerberos Enable is true.
Kerberos Enabled for Job History Server	Can be checked true if Kerberos Enable is true.
Kerberos Enabled for Application Timeline Server	Can be checked true if Kerberos Enable is true.
Kerberos Enabled for Spark History Server	Can be checked true if Kerberos Enable is true.
Kerberos Enabled for Ranger	Can be checked true if Kerberos Enable is true.

JDBC Settings Field Descriptions

Field	Description
Enable Spark SQL Support	Checkbox if need to enable spark
JDBC Username	jdbc username.
JDBC Password	jdbc password.
JDBC Port	jdbc url and port.
File System	hdfs url and port.
Warehouse Directory	hdfs path.
Queue Name	Hive queue name
Queue Value	Hive queue value

Metadata Settings Field Descriptions

Field	Description
RPC Value	Authentication information
Metastore Thrift URL	Hive IP with 9083 port.
Ranger URL	Apache ranger url
Ranger User	Apache ranger username
Ranger Password	Apache ranger password

Report Settings Field Descriptions

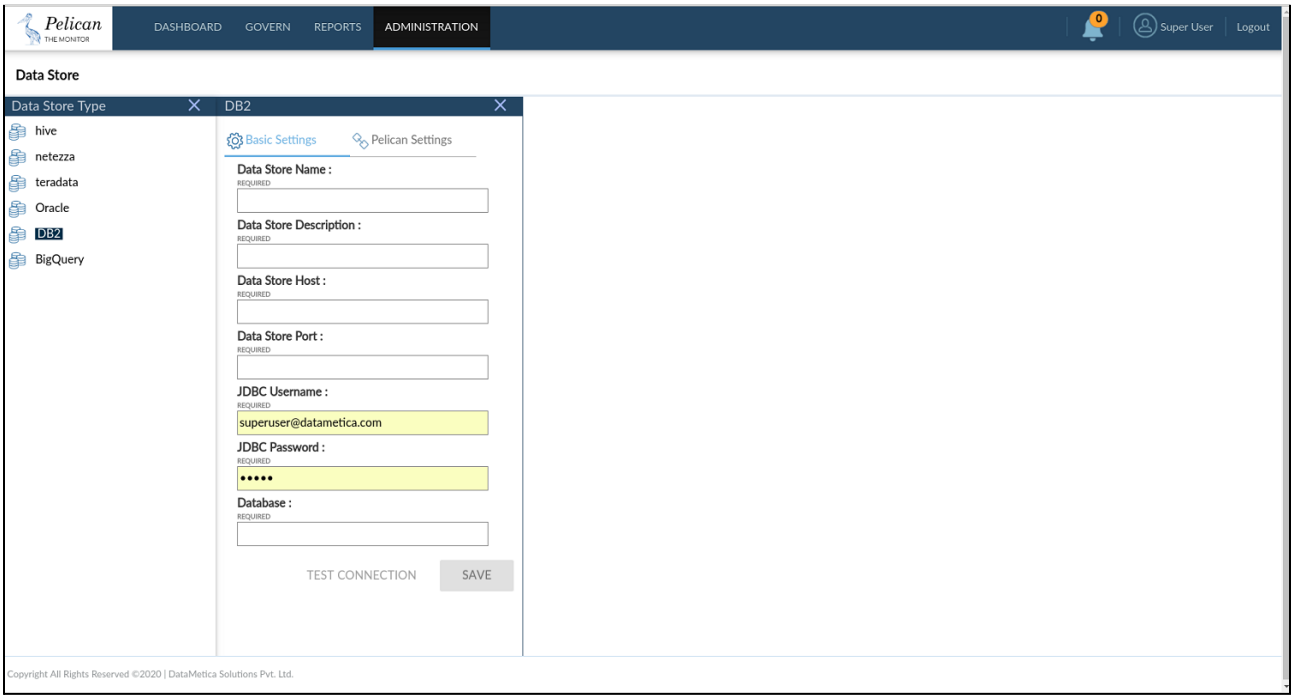
Field	Description
ResourceManager Server URL	Provide resource manager URL if you did not specify Name node URL under basic settings.

Pelican Settings Field Description

Field	Description
Temporary Database	Database used by Pelican for its functioning.
Create Hive UDFs	Checkbox if user wants Pelican to create UDFs (user defined functions in their Temporary Hive database)
Auxiliary Jar File	Jar file path for Hive Pelican UDFs.
Supplementary DataStore Location	Writable Path on the machine where Pelican is installed.

6.7. Steps for DB2 Configuration

1. Go to **Administration > Configure > Datastore**.
2. At the left-hand side pane a list of predefined data store list is visible so select the **DB2** the data store field or enter the data store. Click on  icon.



3. Enter all the details in the: Basic Settings,Pelican Settings fields.Click **SAVE**. You may view the created data store in the list under the respective data store type. Similarly, you can configure any number of data store instances under the predefined data store type. The information to be entered while configuring the data store will vary from with different data stores.

Basic Settings Field Description

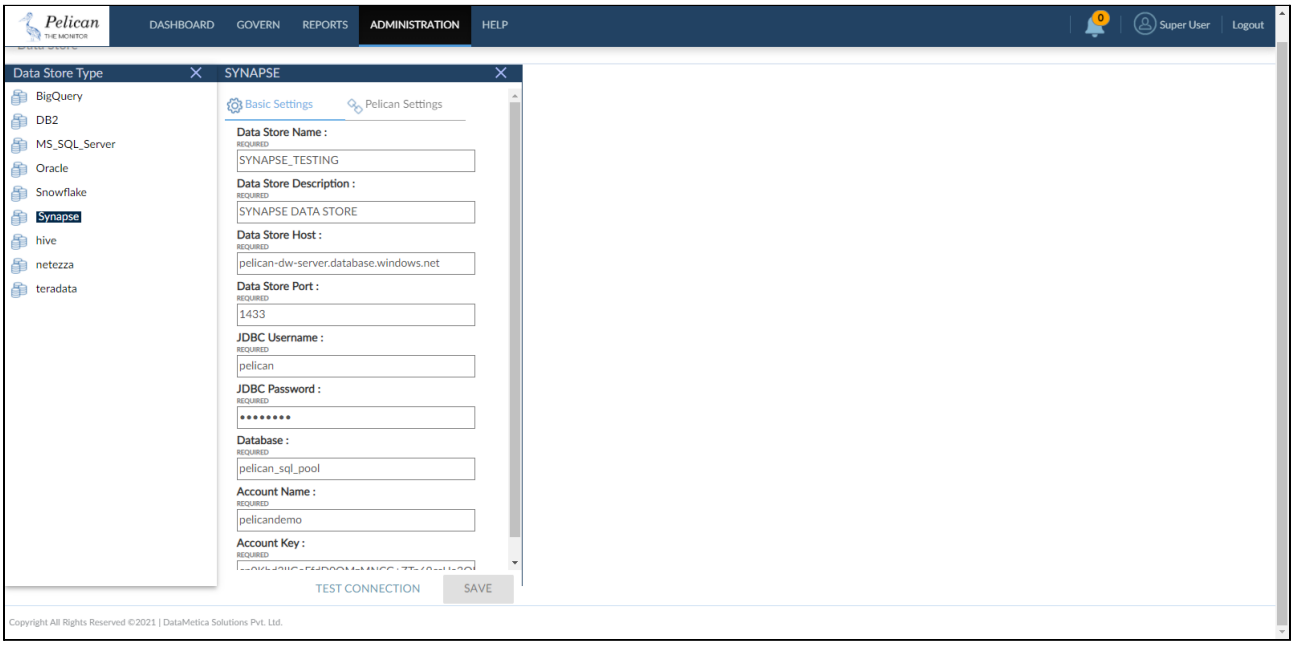
Field	Description
Data Store Name	The name the user wants given to the data store.
Data Store Description	Description for the data store.
Data Store Host	IP address of the data store host machine.
Data Store Port	JDBC port to connect to the data store.
JDBC Username	JDBC username using which Pelican will connect to the data store.
JDBC Password	JDBC password for the provided JDBC user.
Database	Enter the name of the database

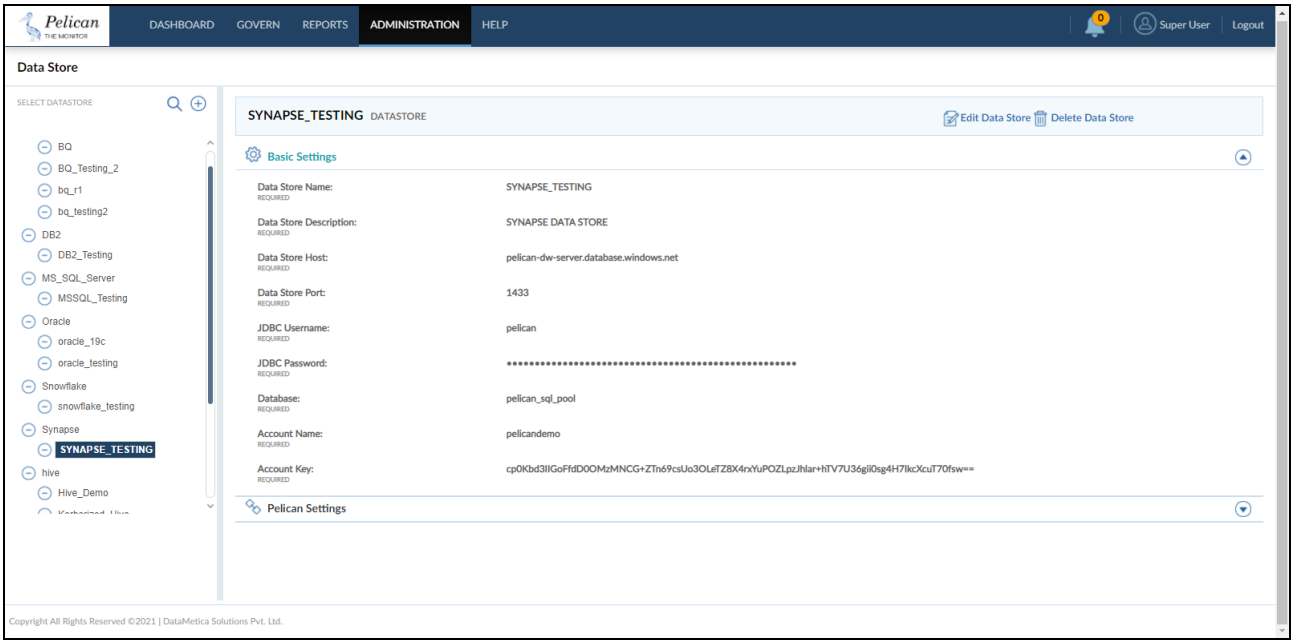
Pelican Settings Field Description

Field	Description
Temporary Database	Database used by Pelican for its functioning.
Supplementary DataStore Location	Writable Path on the machine where Pelican is installed.

6.8. I6.8. Steps for Synapse Configuration

4. Go to **Administration > Configure > Datastore**.
5. At the left-hand side pane a list of predefined data store list is visible so select the **Synapse** the data store field or enter the data store. Click on  icon.





6. Enter all the details in the: Basic Settings,Pelican Settings fields.Click **SAVE**. You may view the created data store in the list under the respective data store type. Similarly, you can configure any number of data store instances under the predefined data store type. The information to be entered while configuring the data store will vary from with different data stores.

Basic Settings Field Description

Field	Description
Data Store Name	The name the user wants given to the data store.
Data Store Description	Description for the data store.
Data Store Host	IP address of the data store host machine.
Data Store Port	JDBC port to connect to the data store.

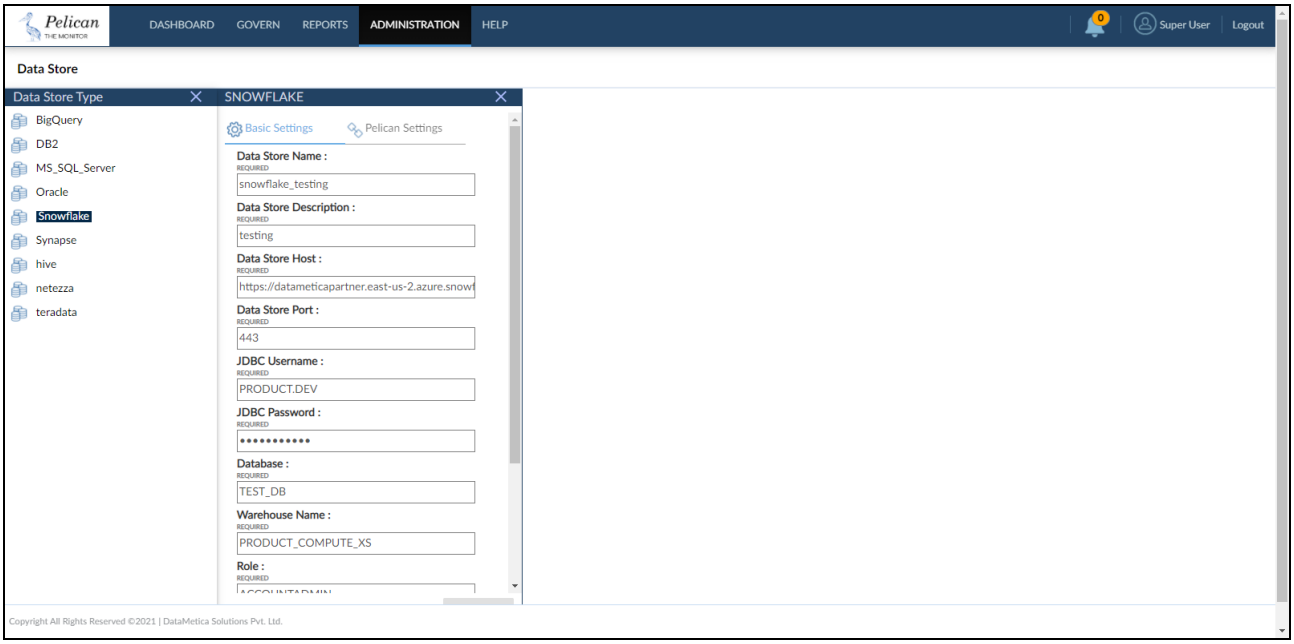
JDBC Username	JDBC username using which Pelican will connect to the data store.
JDBC Password	JDBC password for the provided JDBC user.
Database	
Account Name	
Account Key	

Pelican Settings Field Description

Field	Description
Temporary Database	Database used by Pelican for its functioning.
Supplementary DataStore Location	Writable Path on the machine where Pelican is installed.

6.9. Steps for snowflake Configuration

7. Go to **Administration > Configure > Datastore**.
8. At the left-hand side pane a list of predefined data store list is visible so select the **Snowflake** the data store field or enter the data store. Click on  icon.



9. Enter all the details in the: Basic Settings,Pelican Settings fields.Click **SAVE**. You may view the created data store in the list under the respective data store type. Similarly, you can configure any number of data store instances under the predefined data store type. The information to be entered while configuring the data store will vary from with different data stores.

Basic Settings Field Description

Field	Description
Data Store Name	The name the user wants given to the data store.
Data Store Description	Description for the data store.
Data Store Host	IP address of the data store host machine.
Data Store Port	JDBC port to connect to the data store.

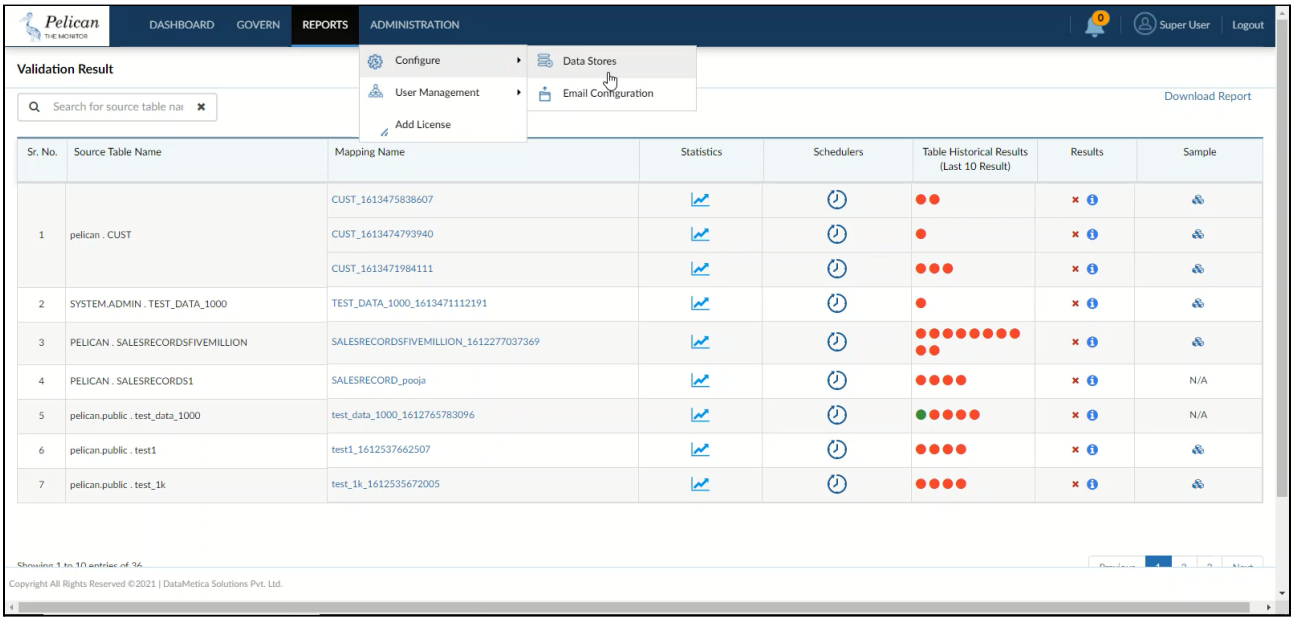
JDBC Username	JDBC username using which Pelican will connect to the data store.
JDBC Password	JDBC password for the provided JDBC user.
Database	
Warehouse Name	
Role	
OAuth Service Account Email	
Project Id	

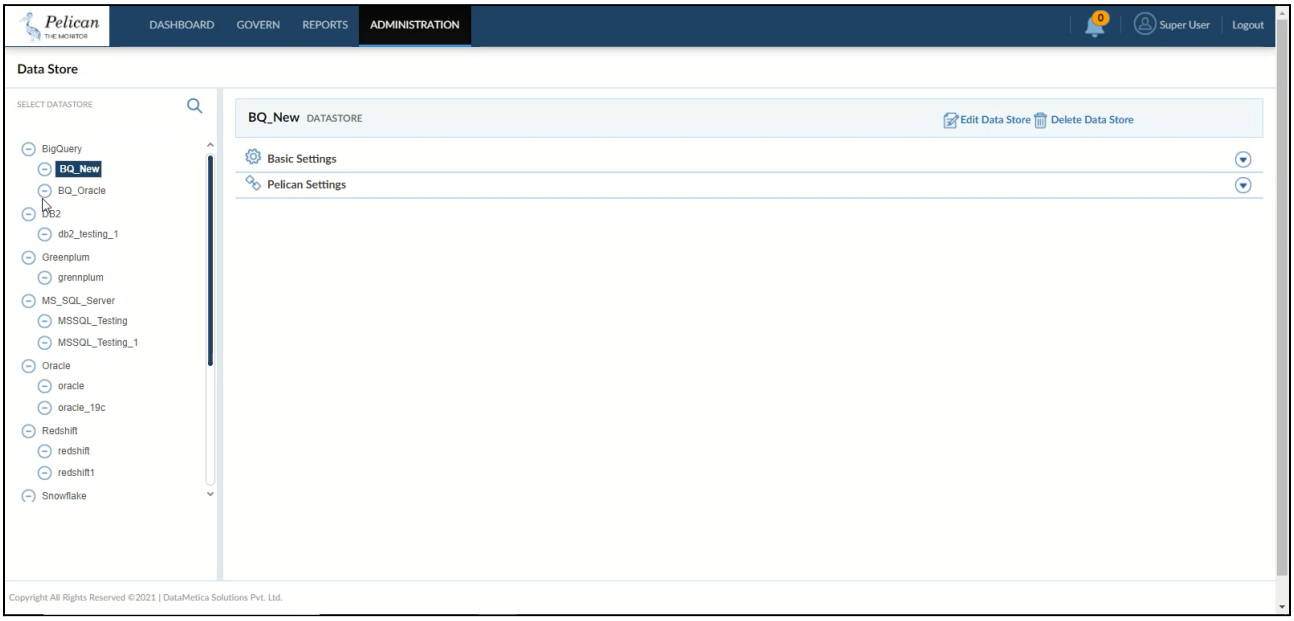
Pelican Settings Field Description

Field	Description
Temporary Database	Database used by Pelican for its functioning.
Cluster Zone	
Pipeline Temp Location	
Pipeline Stage Location	

6.10. How to edit or delete existing Datastore?

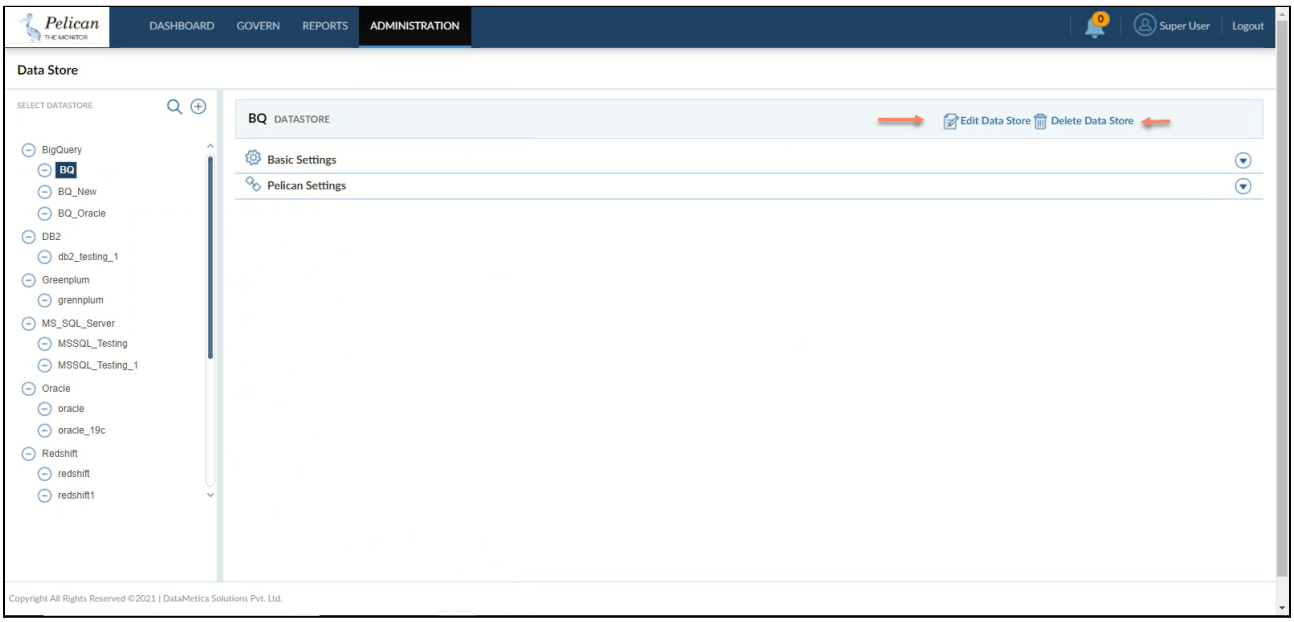
Step 1: Administration → Configure → Datastores



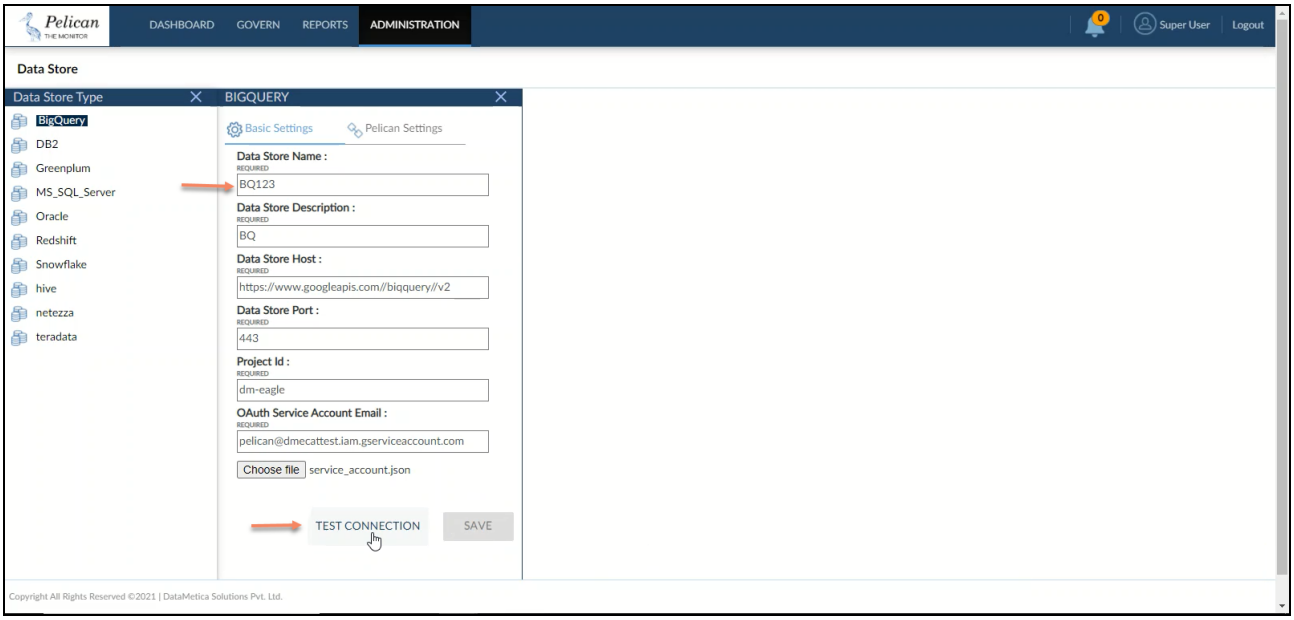


Additionally, you can edit or delete existing data stores as per requirement.

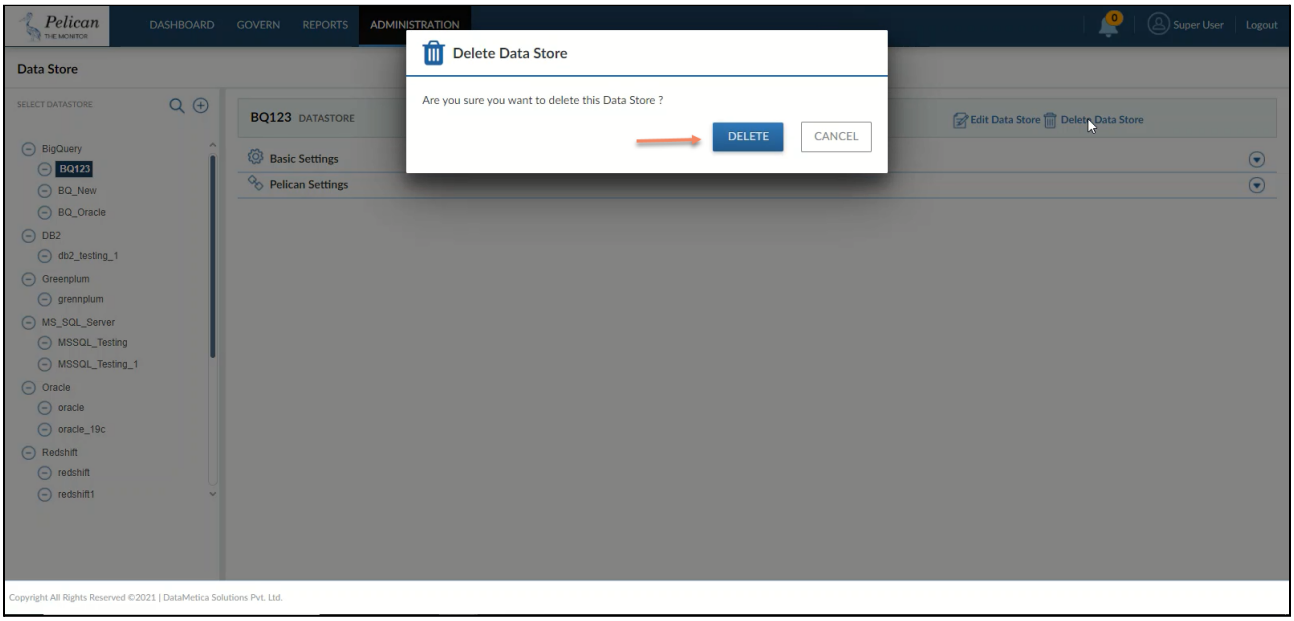
If you find the data entered is wrong or wanted to update some fields, just click on the Edit Data Store icon.



Edit the information of the desired field, upload the service_account.json file and click on the **Test Connection button**. If the connection is correct then the message will pop up as **Verified** in green color. Once the connection is verified, the **Save** button will be enabled.

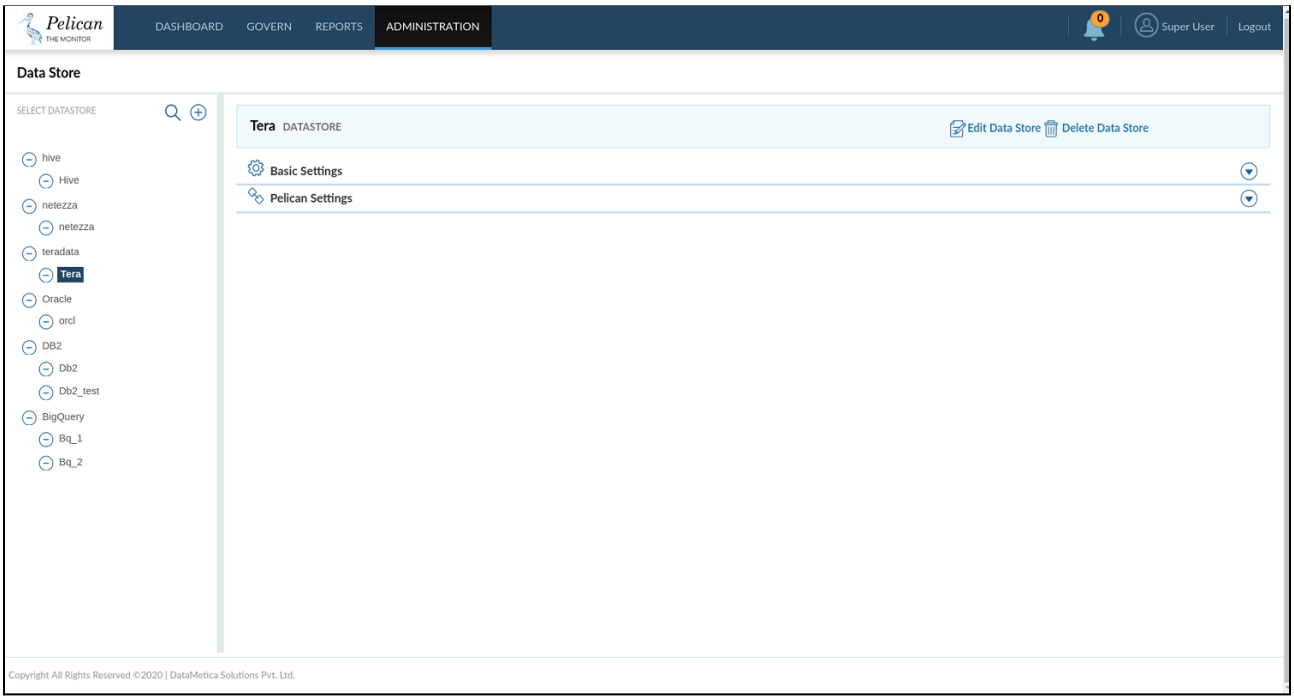


If you want to delete the data store, click on the **Delete Data Store** button. The confirmation message will pop up, if you are sure that you want to delete the data store, click on **Delete.**



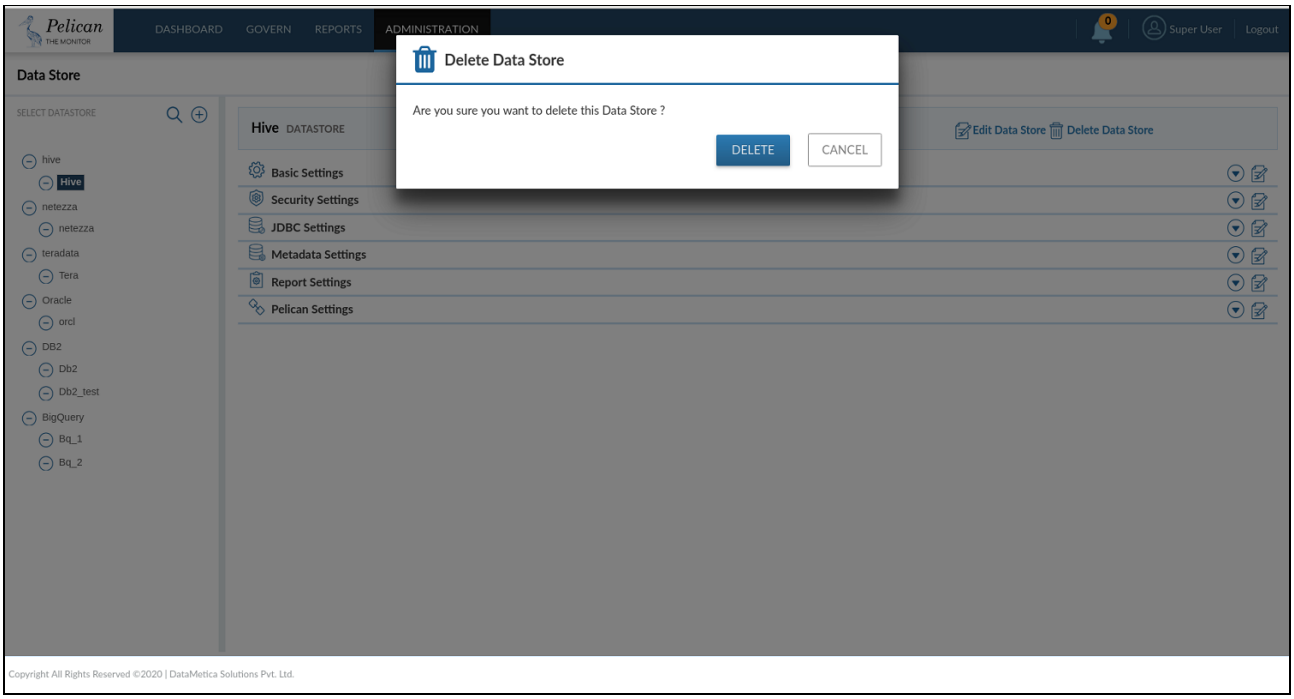
6.11. Edit existing database

You can edit an existing data store by selecting the required data store. Click the Edit Data Store icon and make required changes to the settings, then after you verify the connection click SAVE. The system pop-ups successful notification.



6.12. Deleting existing database

Select the data store you want to delete and click the Delete Data Store icon. The system displays a confirmation dialog box to ensure the deletion of the data store so click DELETE.



7. Validation Configuration

After the source and destination data store is created, the next step is validation configuration. Validation configuration allows you to select tables (source and destination) from the respective data stores and map them.

Validation configuration searches the table with similar column name and data type as in the source Datastore. The application provides various filters to search the desired data store, reference database, destination database, reference schema and destination schema. Additionally, the Pelican facilitates the following approximate mapping methods which allows you to map the tables based on their names and patterns:

- **Phonetic Matching**: A phonetic matching is an algorithm for **matching of words by their pronunciation**.

  For example, the words Principal and Principle are phonetically matching words. Means, table names (source and destination) Principal and Principle will be considered for mapping.

- **Approximate Matching**: Approximate matching is based on Levenshtein distance. It is a metric for measuring the difference between two words. The Levenshtein distance between two words is the minimum number of single-character edits (insertions, deletions or substitutions) required to change one word into the other. Levenshtein distance may also be referred to as edit distance. The maximum allowed edit distance is 2 words.

  For example, the Levenshtein distance between "kitten" and "sitting" is 3, since the following three edits change one into the other, and there is no way to do it with fewer than three edits:

  1. kitten → sitten (substitution of "s" for "k") → distance is 1.
  2. sitten → sittin (substitution of "i" for "e") → distance is 1.
  3. sittin → sitting (insertion of "g" at the end) → distance is 1.

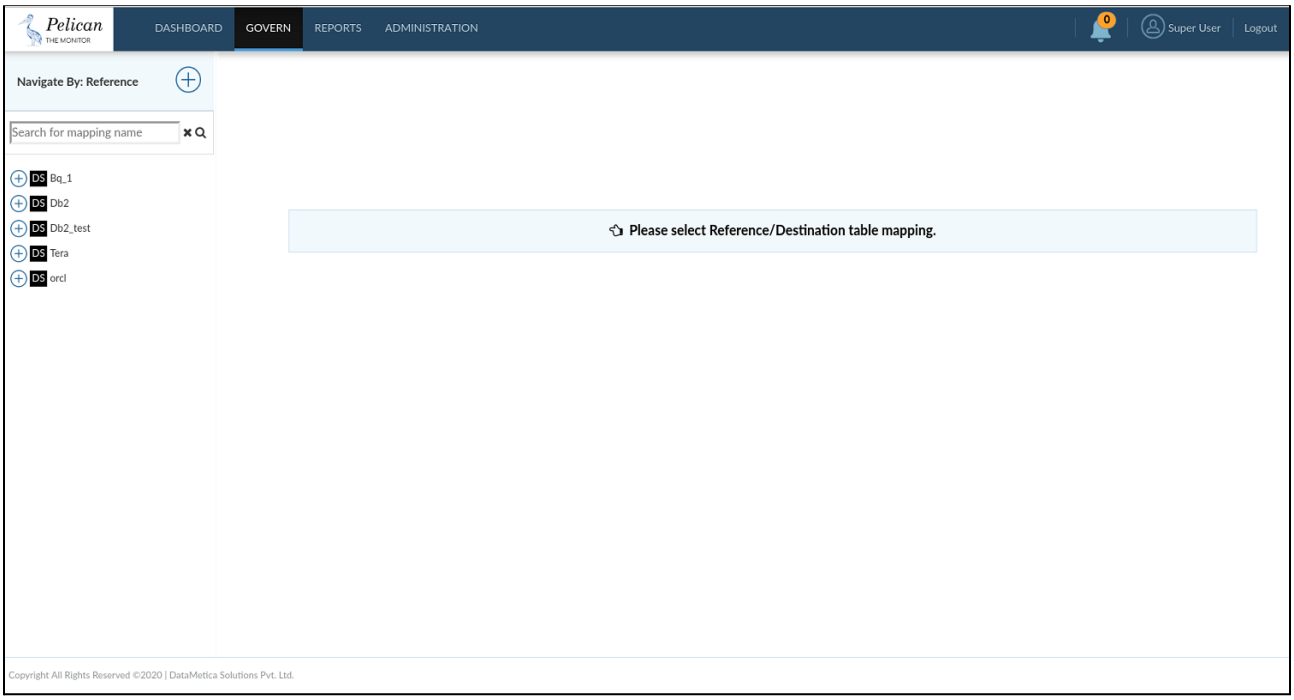
4. Employe → Employee (insertion of "e" at the end) → distance is 1. 5. Employ → Employee (insertion of "ee" at the end) → distance is 2.

The above-mentioned table pair (source and destination) is considered for the mapping as the Levenshtein distance between these table pairs is either 1 or 2.

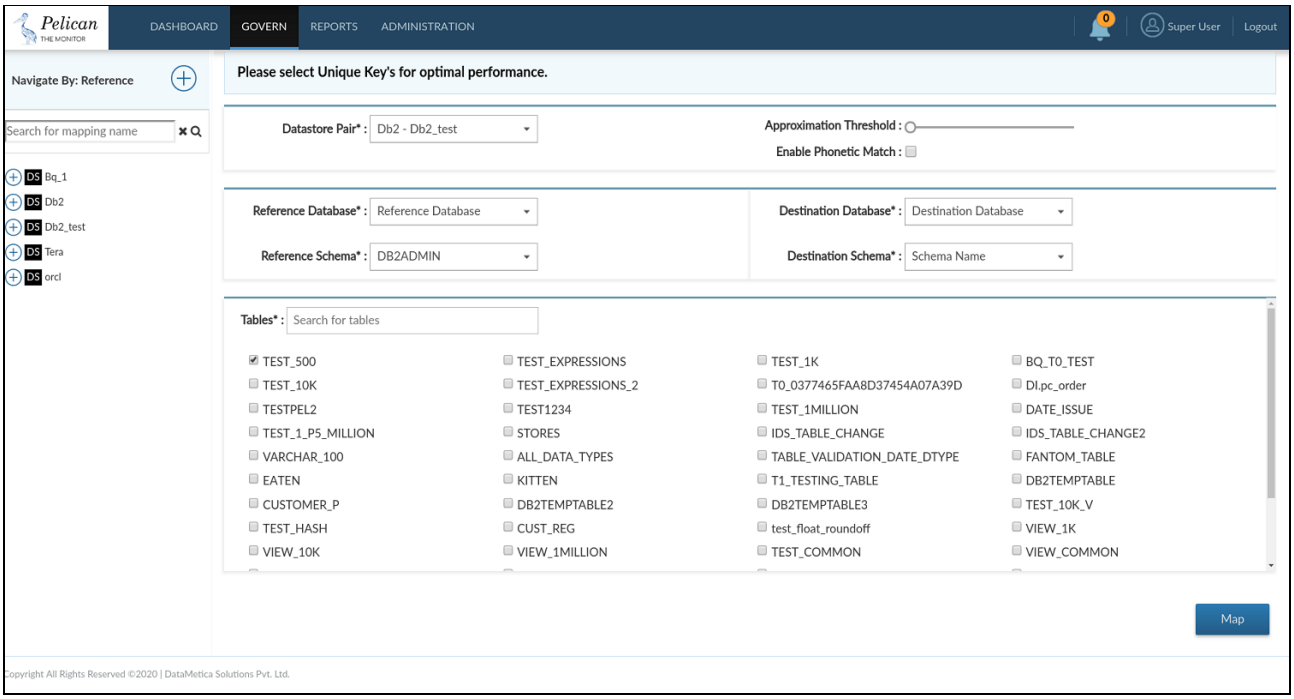
7.1. Table Validation

To validate tables between source and destination data store.

1. Go to **Govern** □ **Validation Configuration**. Click create new mapping icon



2. Click on create new mapping icon.



3. Select the following information.

a Select **Datastore Pair** from the drop-down list. This list displays all the data stores configured in the system.

b Select **Reference Database** from the drop-down list. This list displays all the databases of the corresponding source data store.

c Select **Destination Database** from the drop-down list. This list displays all the databases of the corresponding destination data store.

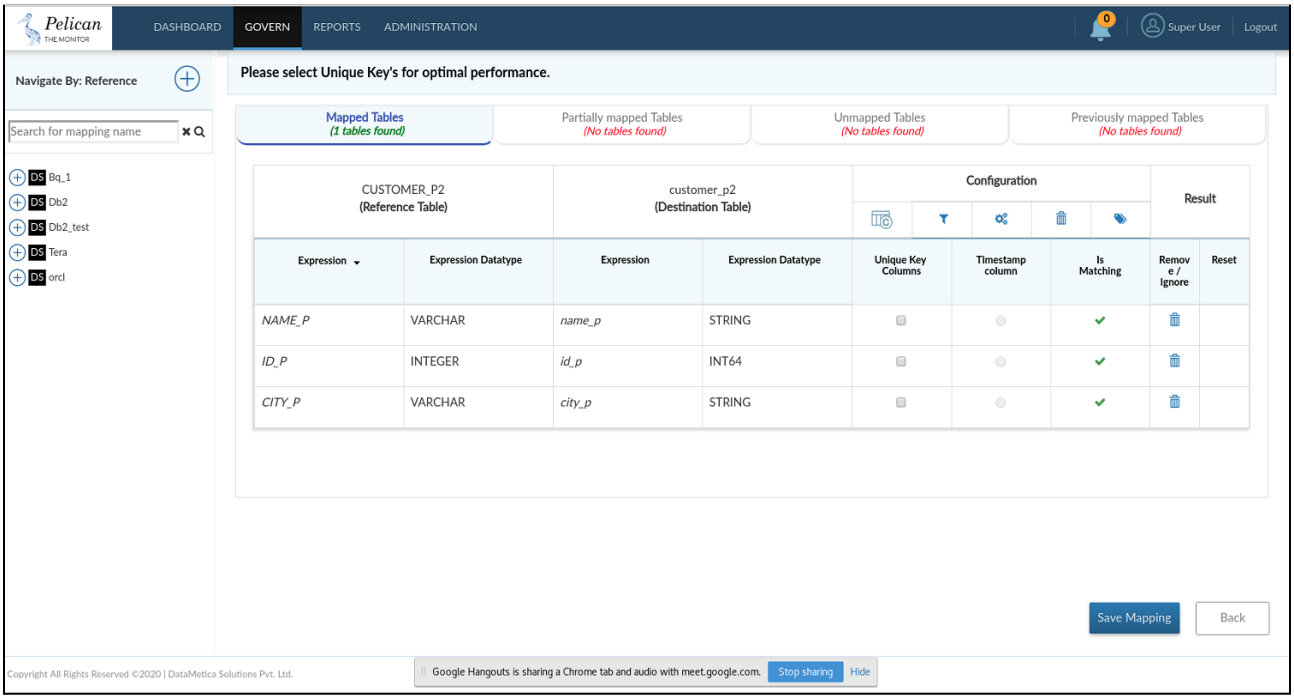
d Select tables in the table section.

This section displays all the tables of the selected reference database. You can select any number of tables. While mapping, the application searches the same table name in the destination database. Further it also searches similar column names and data types as in source Datastore. You can use the approximation matching parameters to search tables, columns, and data types of similar patterns as in the source data store.

e Click **Map**.

Note: After selecting the reference and destination database, you can also select the Schema from the Reference Schema and Destination Schema drop-down lists. Here you can select only one schema from both sides. While mapping, the application searches all the tables, columns, and data types as in the source data store's schema.

4. After you click on Map, the application displays the mapping result in four sections as shown below.



a **Mapped Tables**: This tab contains the table/s where every column of source table/s matches with the target table/s.

b **Partially Mapped Tables**: This tab contains the table/s where few columns of a source table/s match with the target table/s.

c **Unmapped Tables**: This tab contains the table/s where none of the columns of the source table matches with the target table/s.

d **Previously Mapped Tables**: This tab shows the previously mapped tables.

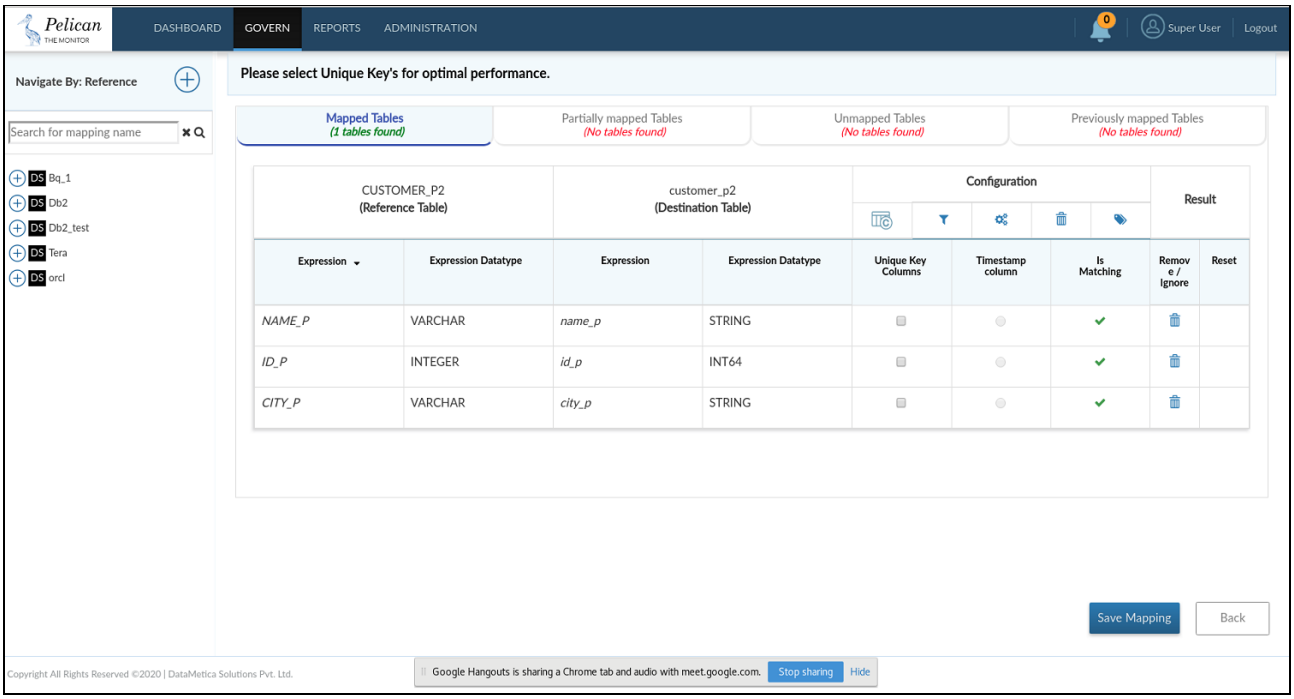
5. Click the respective tab (section) to see the results.
6. The user can perform following operations on the result page:
    - Show Column
    - Edit Configuration
        - Edit Column
        - Edit Expression
        - Edit Data Type
        - Add New Columns
    - Delete table and column
    - Override Unmapped Columns

7.1.1. View columns of mapped, partially mapped, and previously mapped tables

To view columns of mapped, partially mapped, and previously mapped tables,

first open the respective section and then click the **Show Column**  icon.

The application displays the list columns of the reference and destination table as shown below.



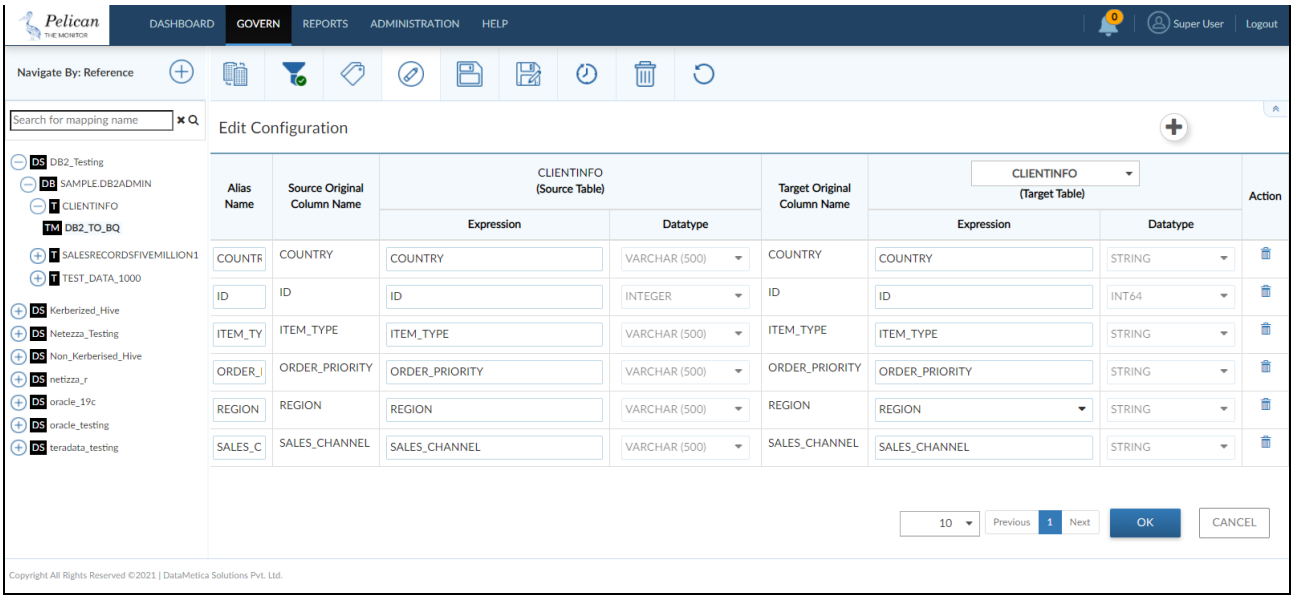
7.1.2. Edit the existing configuration

How to edit an existing column, expression, datatype, and add new columns?

This functionality allows you to edit existing columns, expression, data type, and add new columns as per the business requirement.

To edit the existing configuration:

Click the **Edit Configuration**  icon on the screen. The application displays the Edit Configuration pop-up window as shown below.



In the Edit Configuration window, you can edit existing columns, existing expressions, existing data type and add new columns.

- **Edit existing column name**

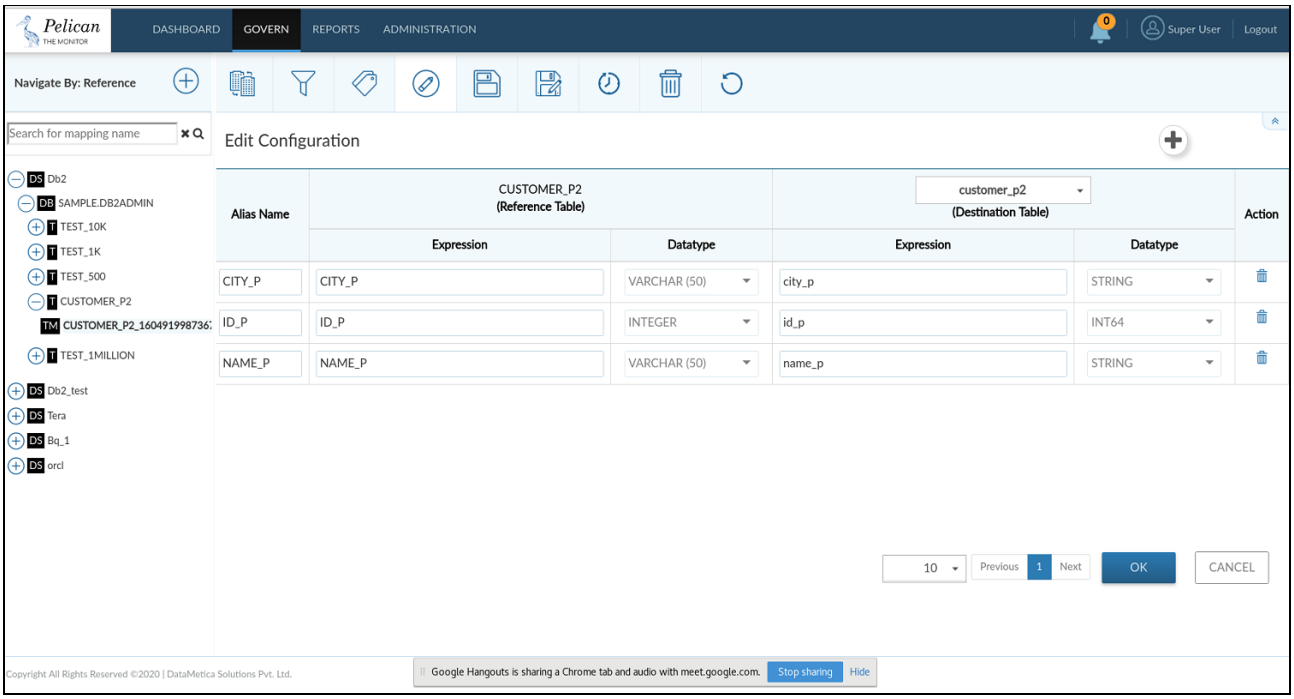
Pelican allows the user to edit the existing column name (Alias Name) to make it uniform on both sides (Reference and Destination).

For example:

Suppose, two columns, namely, ***order_name*** and ***ordername*** exist in the reference and destination tables respectively. Both the columns display the order name hence logically they are the same. So, instead of having two different column names you can edit it to either ***order_name*** or ***ordername*** as an alias name.

To edit existing column name:

Enter the desired column name in the **Alias Name** field of the respective column and click **OK**.



the changes made in the Edit Configuration screen will be saved only in Pelican UI and not in the original database.

- **Edit existing expression**

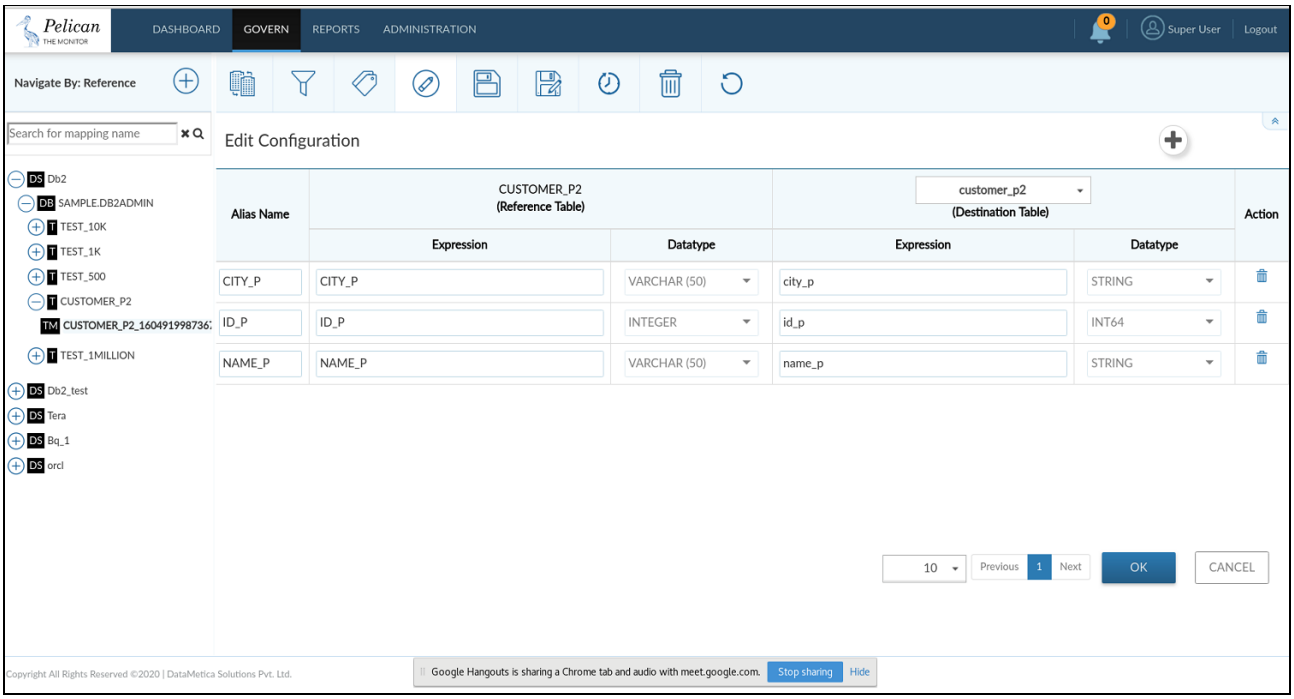
After mapping, the expressions are automatically created. The expressions are created based on the datatype of the respective column. However, the user can edit these existing expressions as per the requirement.

For example,

Suppose the ***date*** column exists on the both sides (Reference and Destination tables), the data types of these columns are **DD:MM: YYYY HH:MM: SS** and **DD:MM: YYYY** respectively. the **HH:MM: SS** is missing in the destination table. So, using edit expression, you can match the mismatched expression.

To edit the existing expression:

Click the desired expression row, edit the expression and click **OK**.

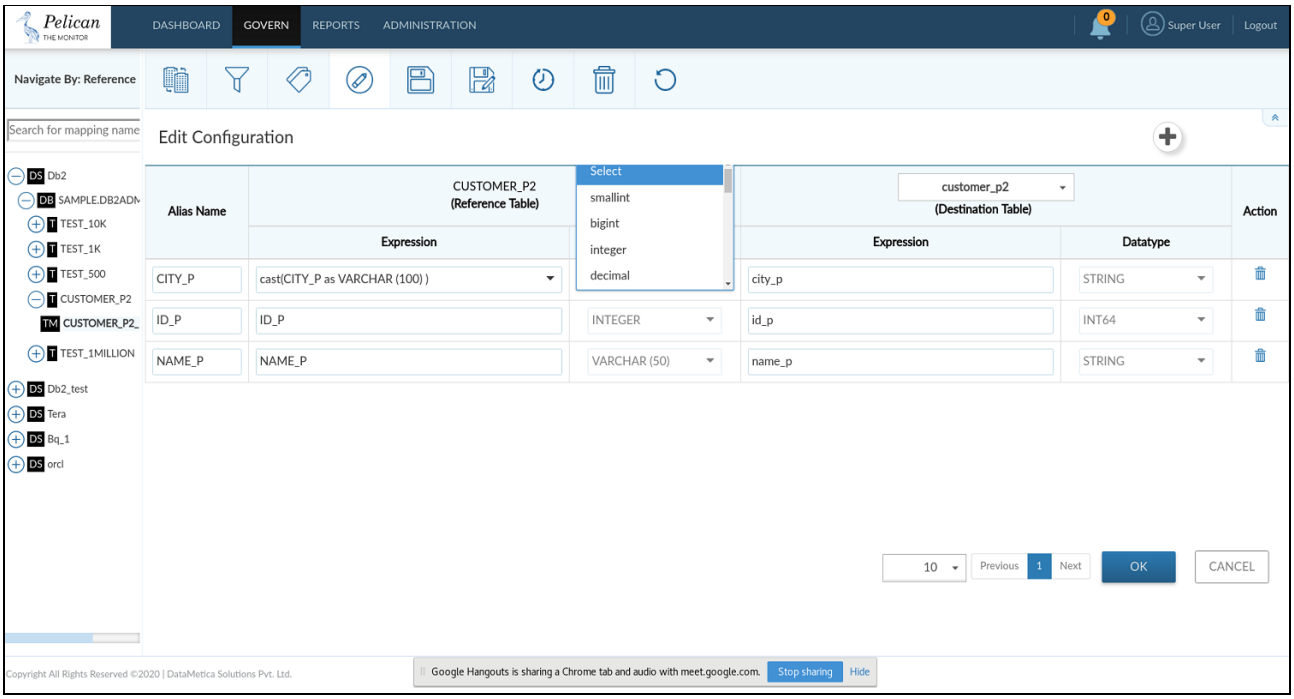


Edit existing data type

The user can edit the existing data type of any column in the reference and destination table. Suppose there is a column that exists on the both sides (Reference and Destination tables), but the data type of these columns is ***Varchar*** and ***String*** respectively. Logically these two data types are similar in nature. Hence, instead of two data types, the user can change it to either ***String*** or ***Varchar***.

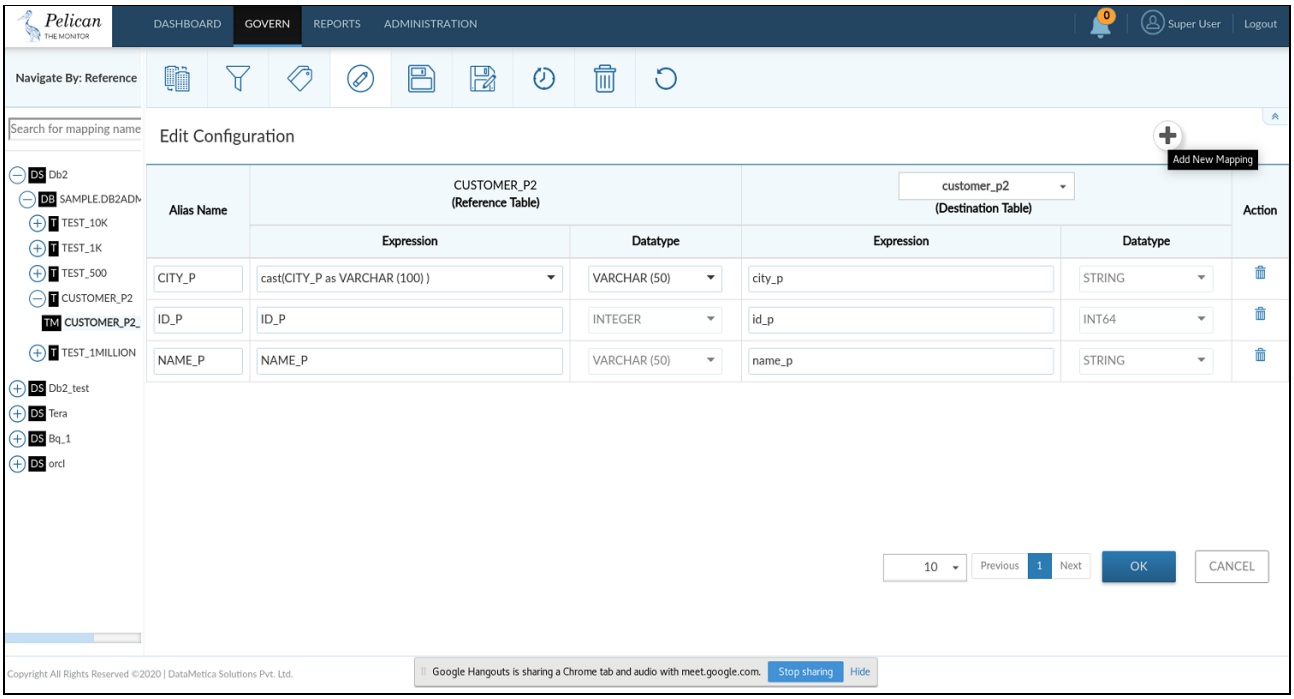
To edit the existing data type:

Select the new data type from the **Datatype** drop-down list and click **OK**.

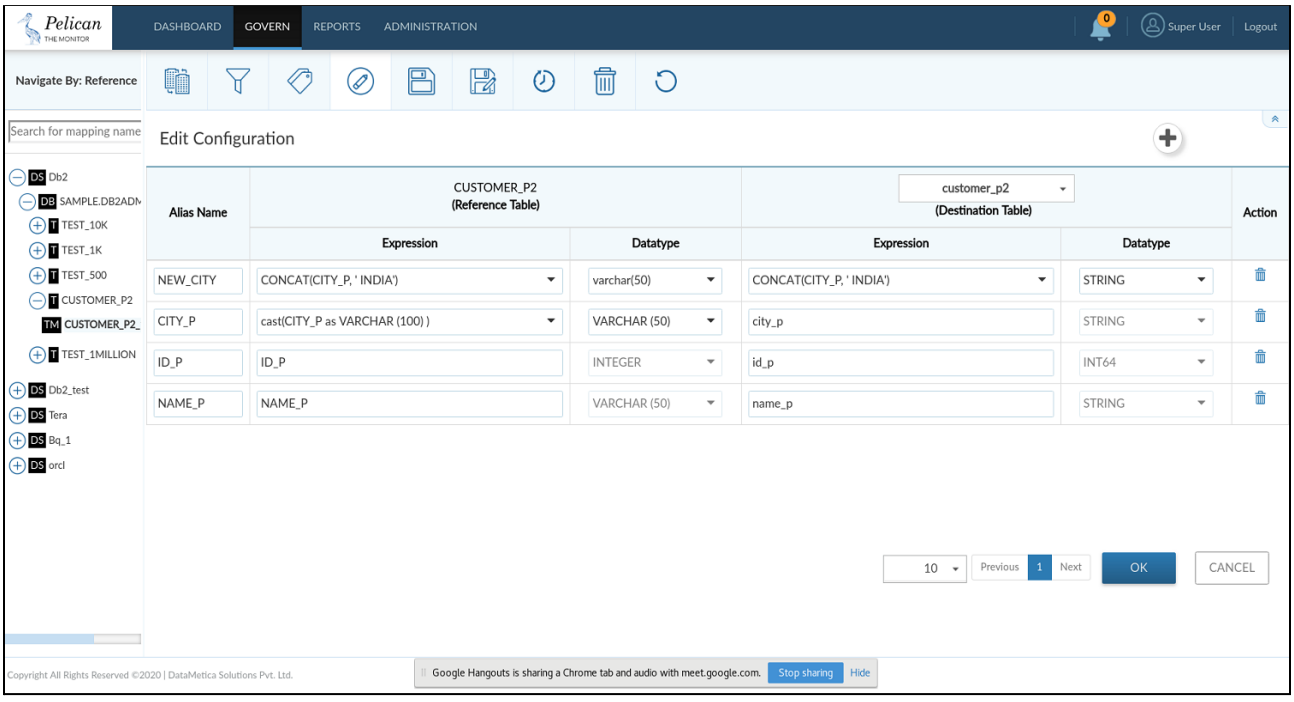


Add new column

Pelican allows the user to add new columns as per the business requirement, the user can add any number of columns.



2. The application adds the new column as shown below.



3. Enter column name (Alias Name).

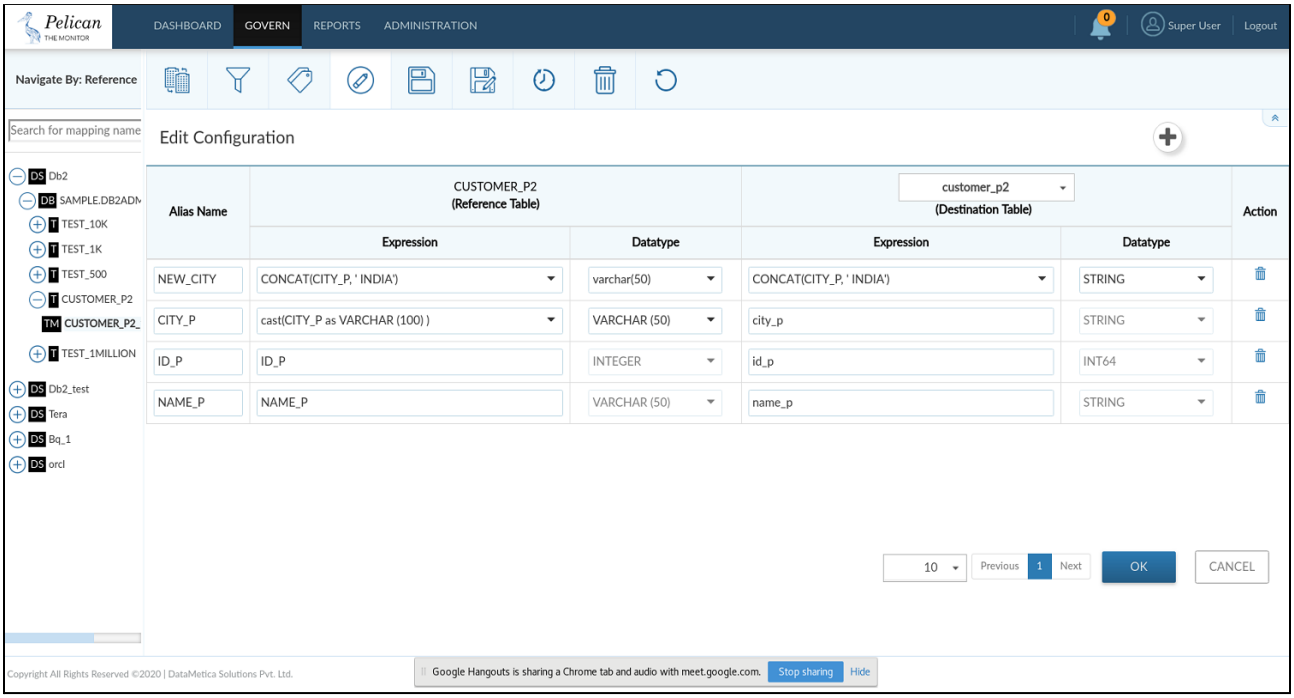
4. Enter expression, datatype on the both sides (Reference and Destination tables) and click Save to save the newly added column.

Delete Tables

The user can delete tables from the respective sections if they are not required. Users can delete the mapped, partially mapped, unmapped and previously mapped table sections. However, the deleted tables will be removed only from the Pelican UI and not from the original database.

To delete a table:

1. Click Delete Table  icon as shown below.



Delete columns

The user can delete the existing columns of the reference and destination table if they are not required. Users can delete columns from the mapped, partially mapped, unmapped and previously mapped table sections as well. However, the deleted columns will be removed only from the Pelican UI and not from the original database.

7.1.3. Add Tags

Tagging is a feature associated with table mappings. Now a user can set one or more tags in a table mapping. For example, a user can create a table mapping M1 with multiple tags T1 and T2.

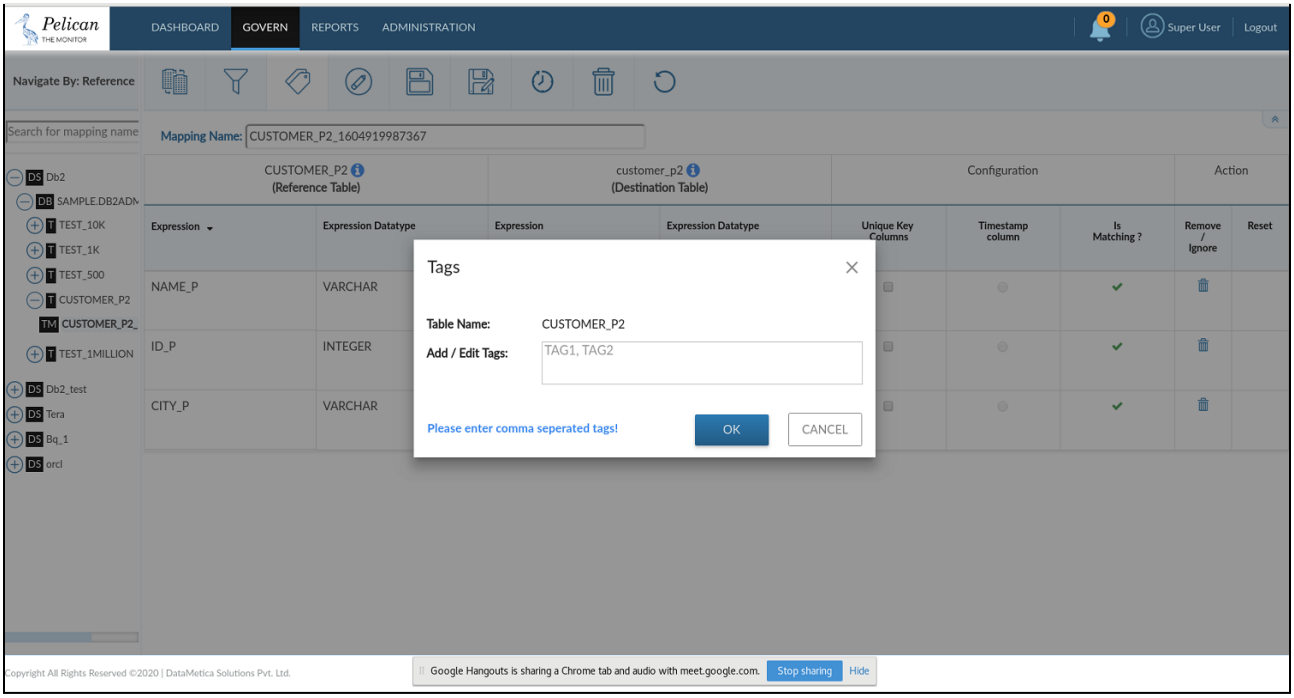
Number of tags in a mapping could be zero. Same tags can be used with multiple mappings. Users can edit tags any number of times from the Edit Mapping page.

Steps to Add tags

1) Create new mapping (Administration → Govern → Validation Configuration → **+** sign)

2) Select databases and tables to map and click to **MAP.**

3) Add new Tags from the new mapping page by clicking 5th option in Configuration.



4) Save the new mapping.

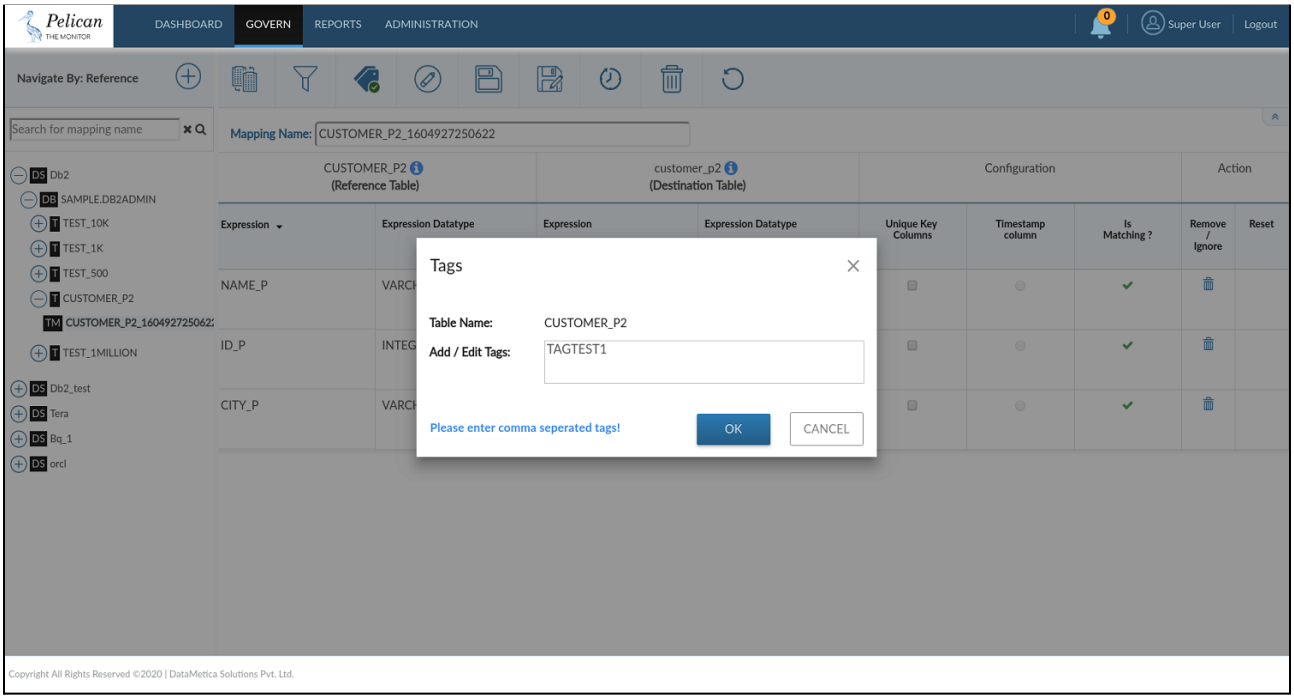
Steps to Edit tags

1) Open mapping (Administration → Govern → Validation Configuration → Select data store and mapping from left)

2) Click on the add tags option  from the menu.

3) Update the tags in the pop-up. Already saved Tags will display in pop-up. Click **OK**.

4) Click on Save Mapping.



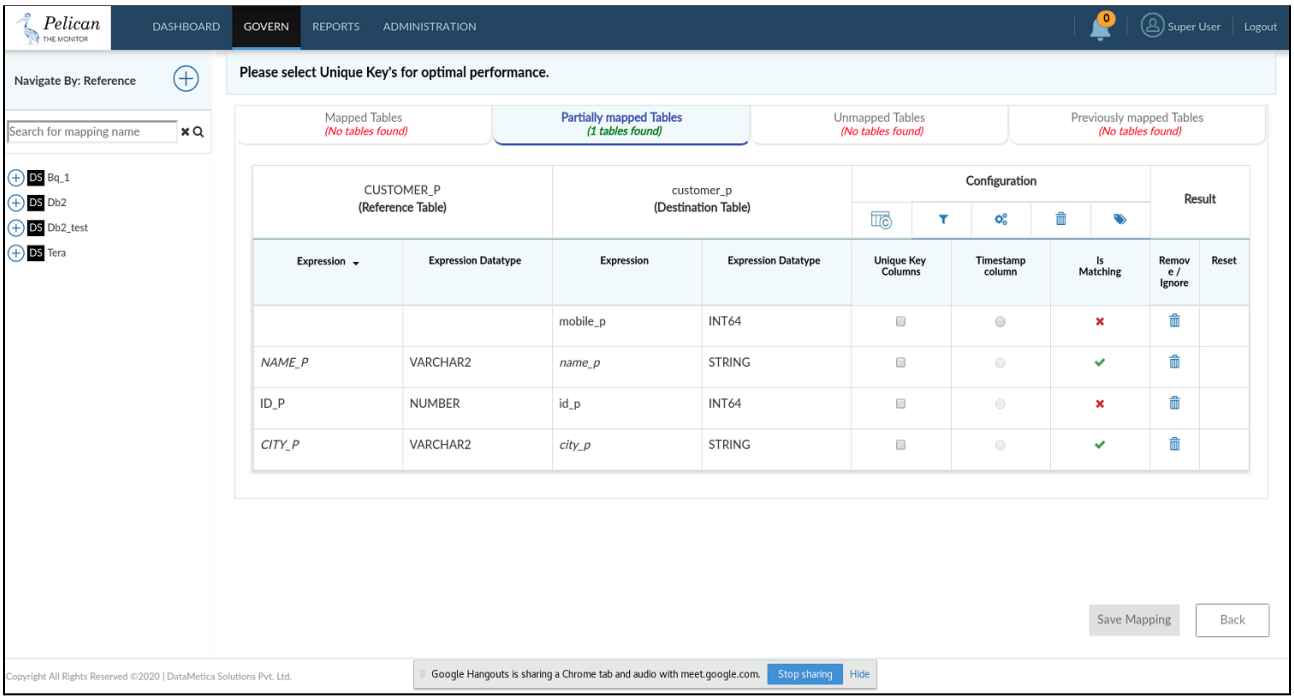
7.1.4. Override unmapped columns

Pelican allows the user to override the unmapped column. For example,

Suppose column ***stud_id*** and ***id*** exist on the both sides (Reference and Destination tables), logically these two columns are identical as both represent student identification numbers in the respective databases. However, while mapping, the application won't map ***stud_id*** and ***id*** due to variation in name and mark then under the unmapped columns section. In such cases, the user can override the unmapped columns and into mapped columns.

To override unmapped columns:

1. Locate unmapped column row on the screen and click the symbol as shown below.



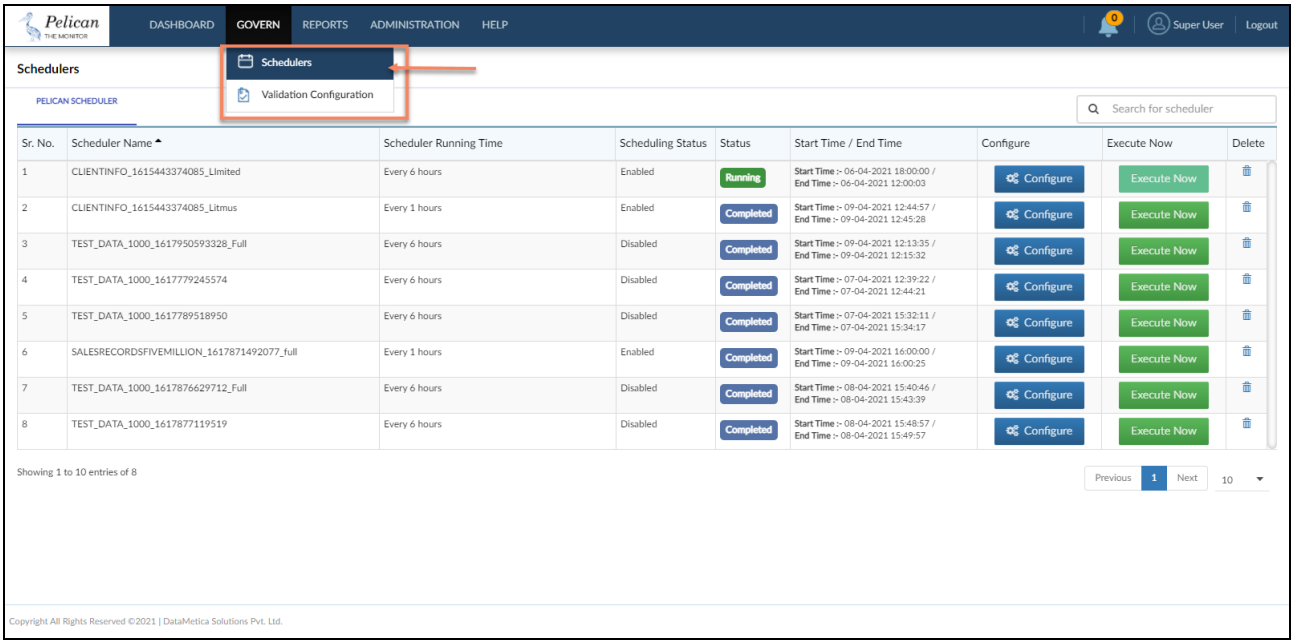
2. Click the **Save Mapping** button to save the changes.

8. Scheduler Configuration

The Scheduler allows the user to execute processes at a regular time interval. In Pelican, the user can create a scheduler for a saved mapping; so that, after a specific time period the scheduler executes the process and it validates the source table with destination tables.

8.1. Modes in Scheduler Configuration

The user can execute processes at a regular time interval. In Pelican, the user can create a scheduler for a saved mapping. The scheduler executes the process and validates the source table with destination tables. Once you complete the table mapping process, the application saves the respective mapping in the Pelican and then you can configure the scheduler against it. In the Scheduler you can set two different modes Litmus and Full.



8.1.1. LITMUS and FULL for Scheduler Configuration

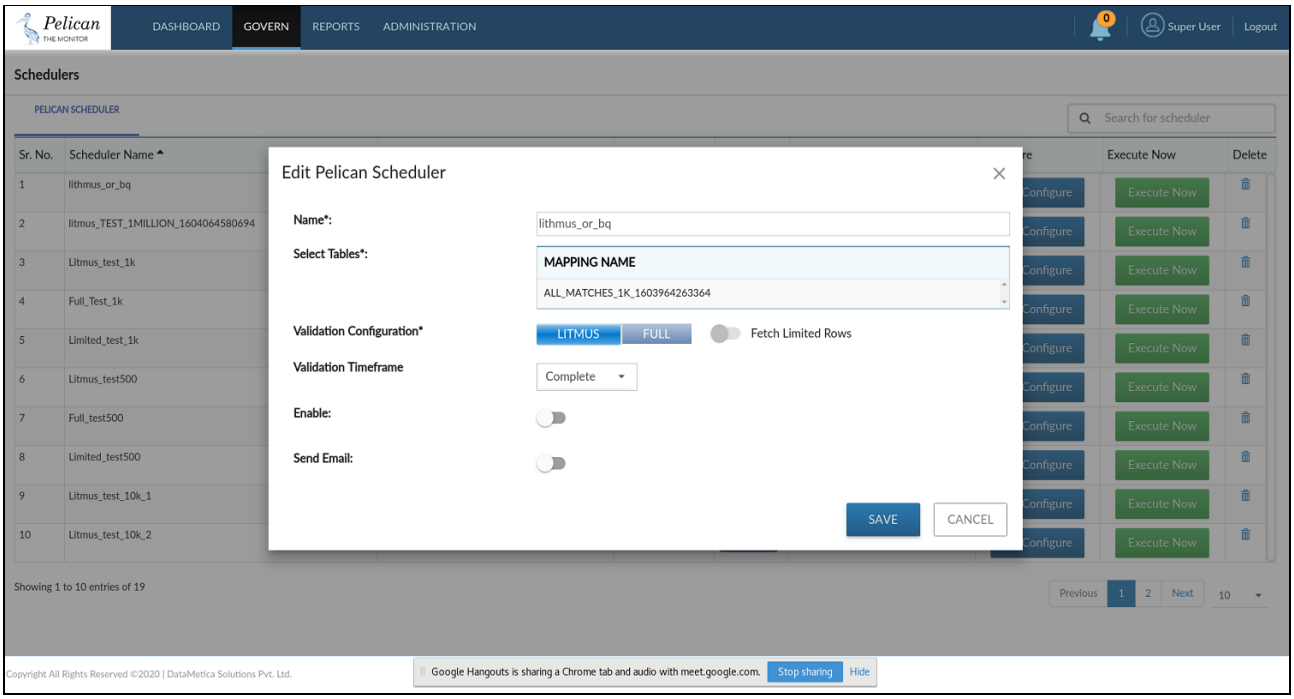
Once you complete the validation mapping process, the application saves the respective mapping in the Pelican UI and it allows you to create and configure the scheduler against it.

LITMUS Scheduler

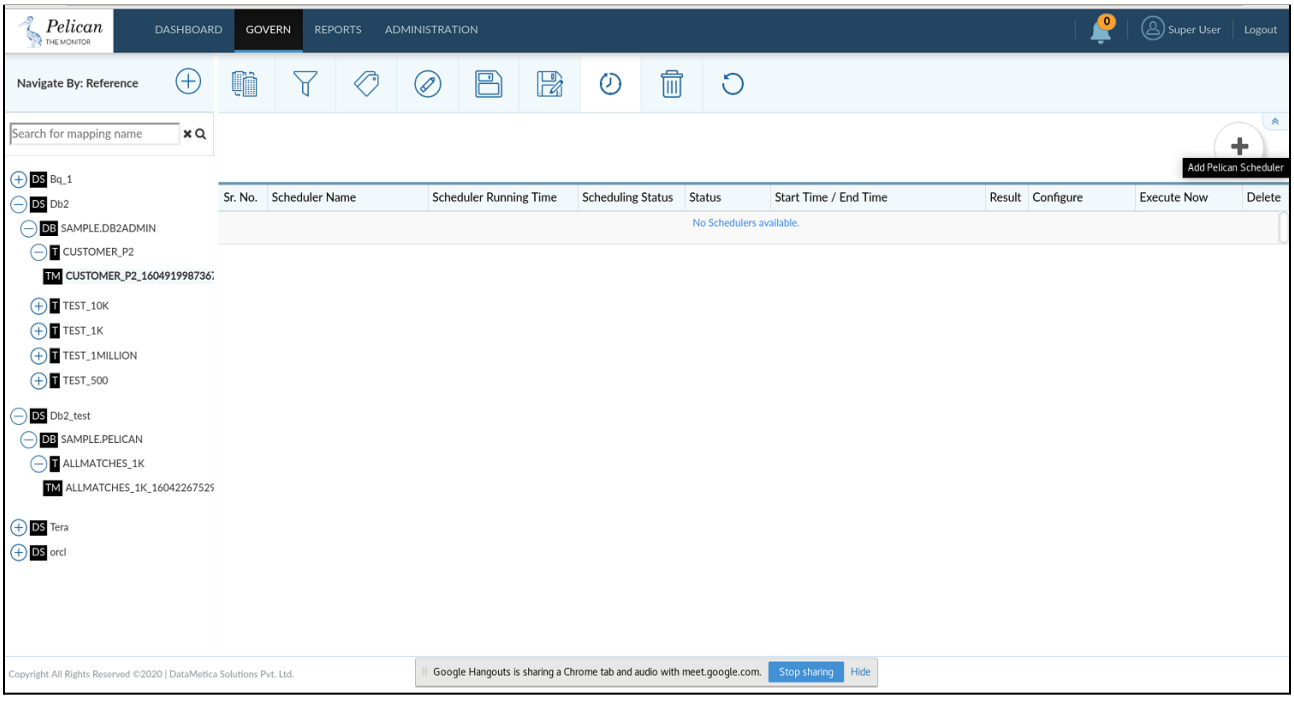
Litmus supports analysing if the tables at source and target are matching or mismatching. It does not give us the cell level difference.

Steps

1. Navigate **Govern** -> **Validation Configuration**.

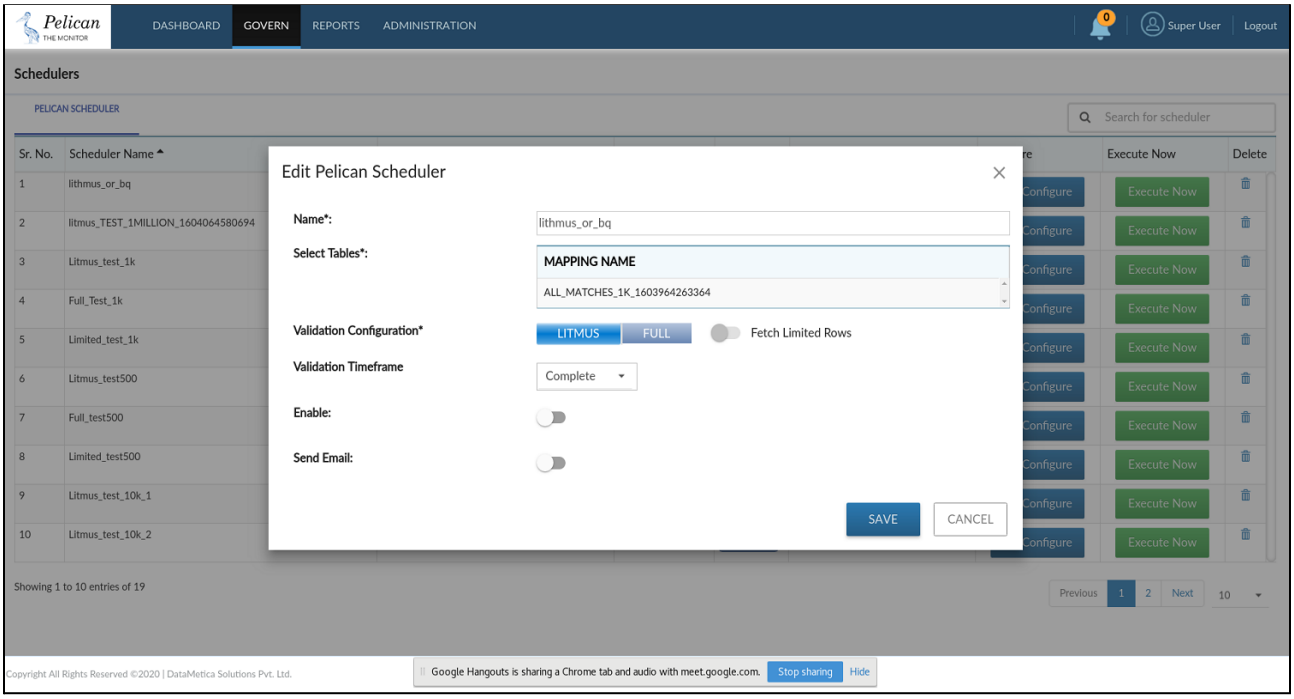


2. Select Table Mapping from the hierarchy and click **View Scheduler** icon on the toolbar.
3. Click the **Add Pelican Scheduler** icon.



4. Enter Scheduler Name in the field provided and select Tables.
5. For Validation Configuration either you select LITMUS or FULL as per requirement.
6. Select Validation Time Frame from the drop-down list and then click the Enable toggle button to set Recurrence Pattern.
7. Select Email toggle button and enter email address if user wants reports to be sent for specific scheduler
8. Click **SAVE**.

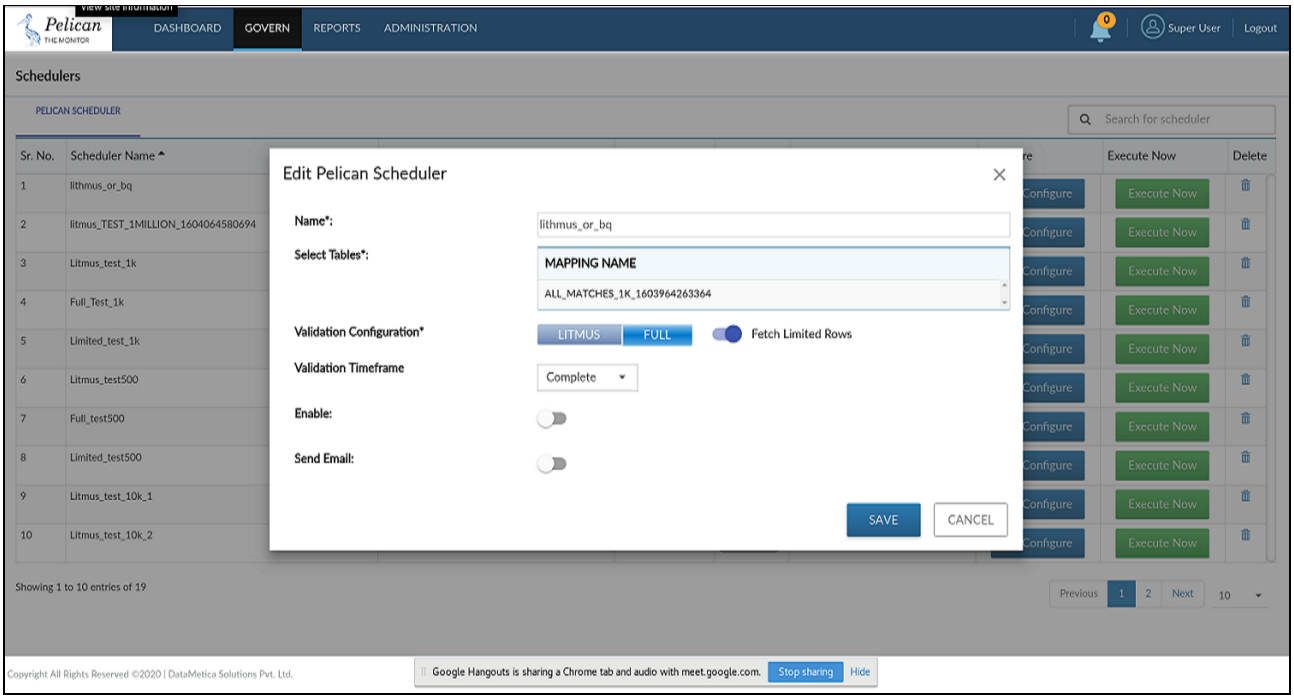
The newly created scheduler is added to the **Schedulers** screen as shown below.



9. Click **Execute Now** to execute the respective scheduler. The application automatically executes the corresponding scheduler as per the set configurations and validation is done accordingly for the saved mappings.

FULL Scheduler

Full mode provides a page of differences between the source and target that displays the mismatches in the tables if present else sample data from both sides is shown.



To create FULL mode select FULL from Validation Configuration and then follow steps specified above for LITMUS Scheduler.

Fetch Limited Rows

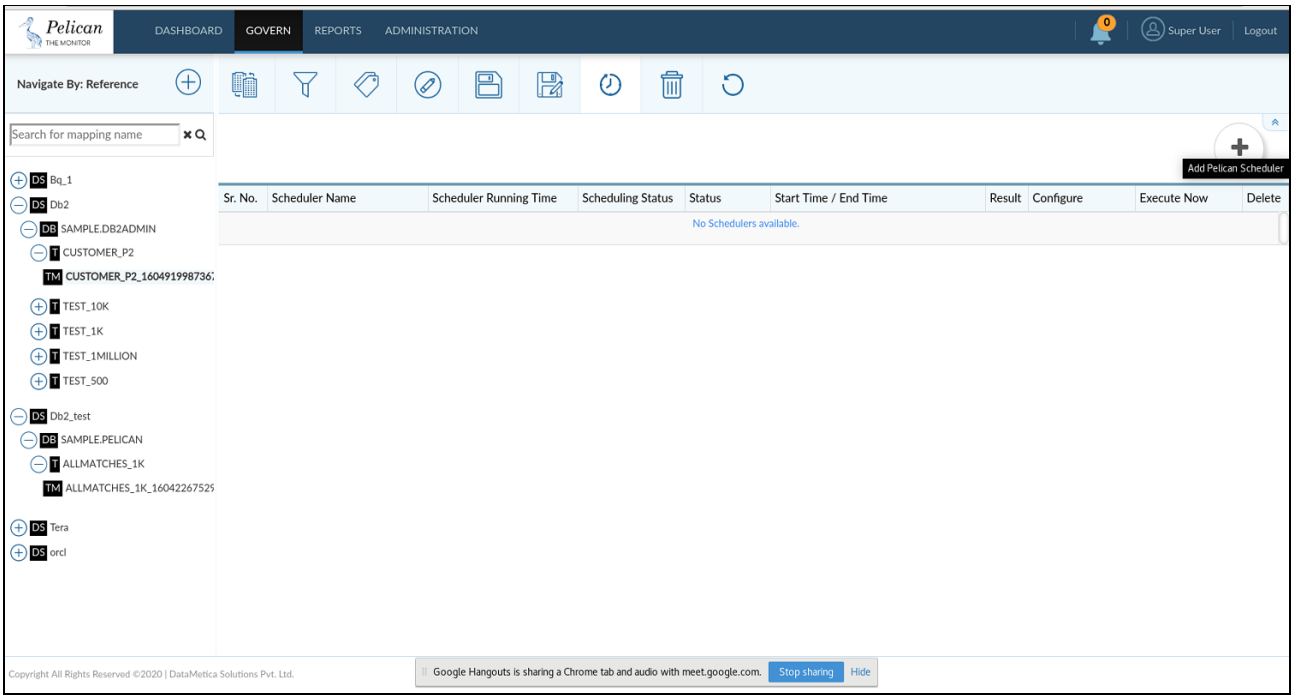
When we select FULL mode with “Fetch Limited Rows” options then the user will get a sample according to the configured number of rows. If the Fetch Limited Rows is not enabled then the user will get complete mismatch.

8.1.2. Configure Pelican Scheduler and enter the recipient information

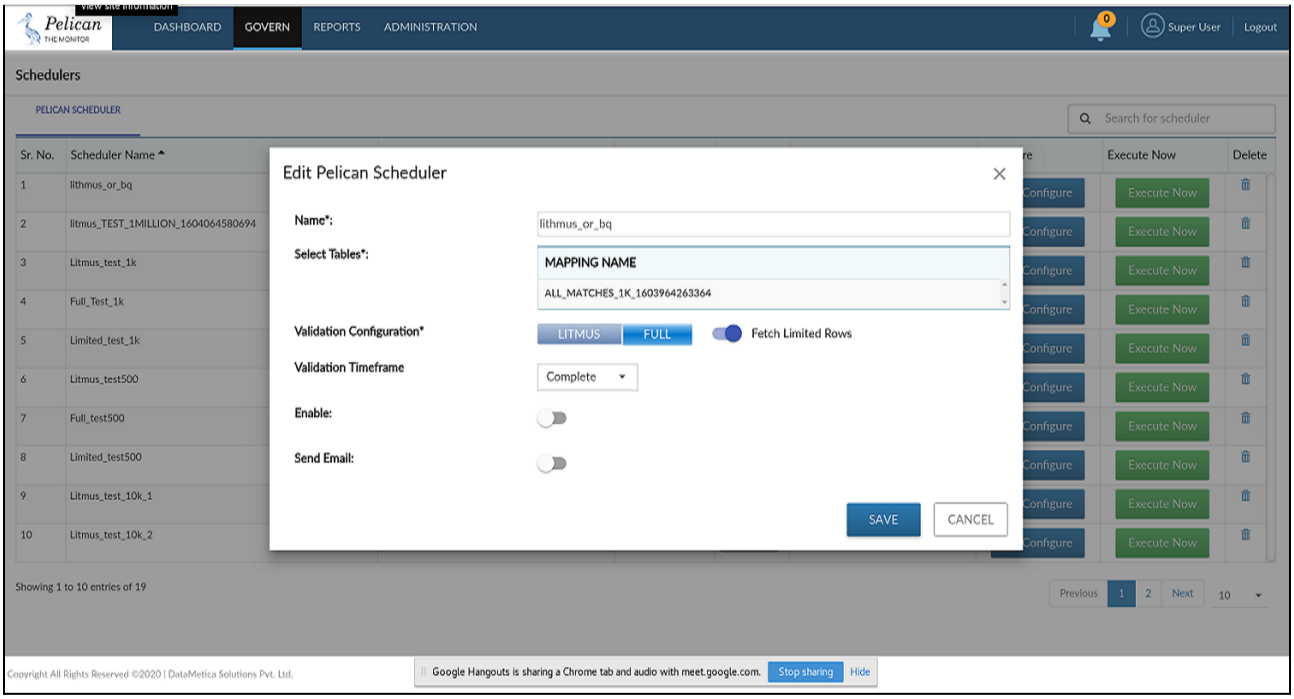
After selecting Pelican reports from the Pelican Status report screen the user needs to configure the Pelican scheduler and enter the recipient's information.

To configure Pelican scheduler and enter recipient information:

1. Go to the Administration menu.
2. Select the **Govern** and click **Schedulers**. This displays the Scheduler screen as shown below.



3. Select the **Pelican Scheduler** tab and click **Configure**, this displays the **Edit Pelican Scheduler** pop-up window as shown below.



4. Click the **Send Email** toggle button, this enables the text field which allows the user to enter the recipients email address.
5. Enter recipients email address in the field provided. Here you can add more than one recipient by using a comma (,).
6. Click **Save** to save the configuration

8

9. Validation Result

After scheduler execution, the application automatically generates the validation result in the form of a report. The validation result can be viewed in two ways:

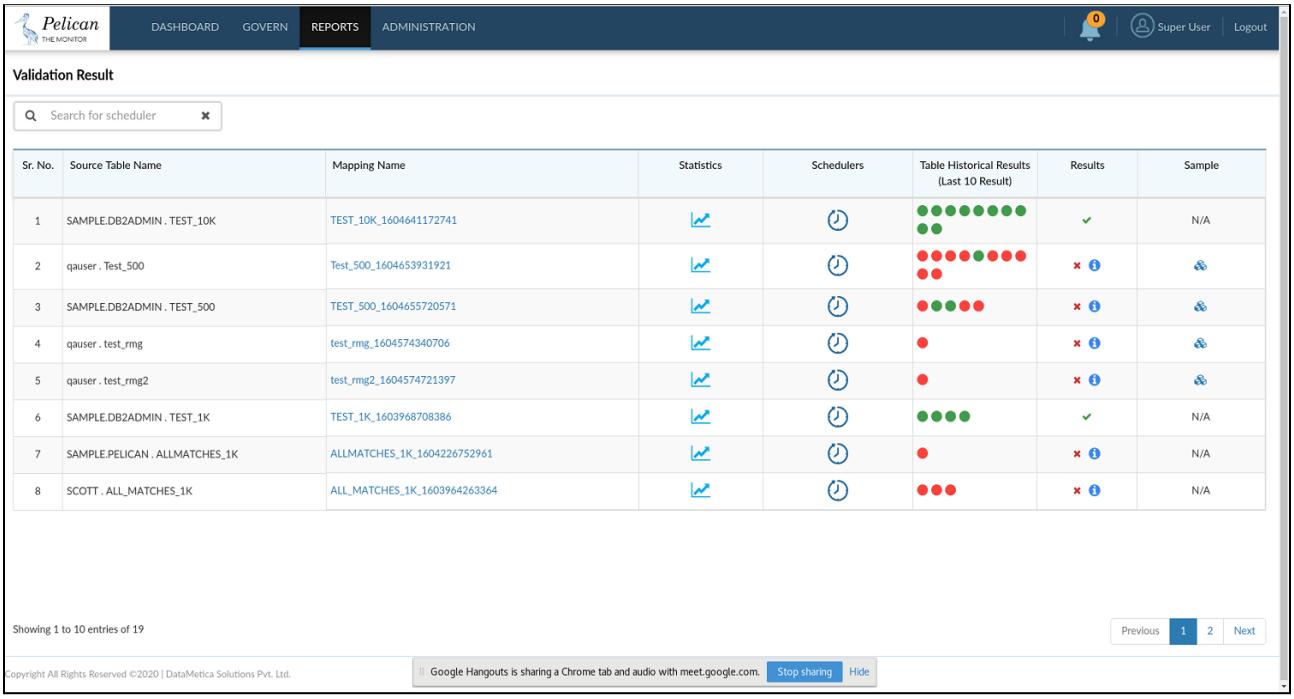
- Common (Validation result for all schedulers): Here the user can view scheduler results for all schedulers that have executed in the system.
- Scheduler Wise: Here the user can view validation results for a specific scheduler against the mapping.

9.1. Results from Report page

Here the application displays the list of all the schedulers along with the validation results.

To view validation result:

Go to **Reports** and click on **Validation Report**. The application displays the Validation Result screen as shown below.

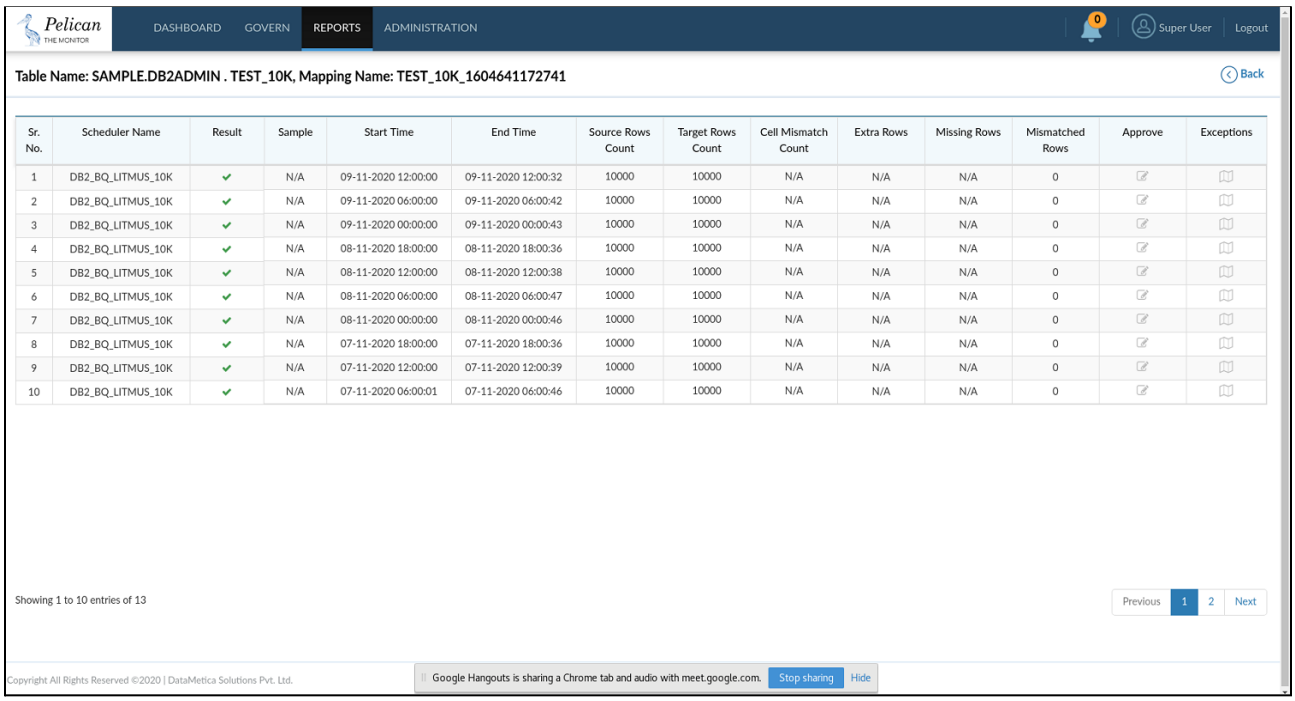


The Validation Result screen displays the following information.

1. **Source Table Name** : This column displays the list of source tables for which mapping has been executed in the system.
2. **Mapping Name**: This column displays the mapping which has been executed for the particular table.

3. **Statistics**: This column displays the scheduler history of the respective table mapping.

   Click the **View Mapping Historical Results** icon to view detailed history. This displays the detailed history of the respective scheduler as shown below.



Sr. No.	Scheduler Name	Result	Sample	Start Time	End Time	Source Rows Count	Target Rows Count	Cell Mismatch Count	Extra Rows	Missing Rows	Mismatched Rows	Approve	Exceptions
1	DB2_BQ_LITMUS_10K	✔	N/A	09-11-2020 12:00:00	09-11-2020 12:00:32	10000	10000	N/A	N/A	N/A	0		
2	DB2_BQ_LITMUS_10K	✔	N/A	09-11-2020 06:00:00	09-11-2020 06:00:42	10000	10000	N/A	N/A	N/A	0		
3	DB2_BQ_LITMUS_10K	✔	N/A	09-11-2020 00:00:00	09-11-2020 00:00:43	10000	10000	N/A	N/A	N/A	0		
4	DB2_BQ_LITMUS_10K	✔	N/A	08-11-2020 18:00:00	08-11-2020 18:00:36	10000	10000	N/A	N/A	N/A	0		
5	DB2_BQ_LITMUS_10K	✔	N/A	08-11-2020 12:00:00	08-11-2020 12:00:38	10000	10000	N/A	N/A	N/A	0		
6	DB2_BQ_LITMUS_10K	✔	N/A	08-11-2020 06:00:00	08-11-2020 06:00:47	10000	10000	N/A	N/A	N/A	0		
7	DB2_BQ_LITMUS_10K	✔	N/A	08-11-2020 00:00:00	08-11-2020 00:00:46	10000	10000	N/A	N/A	N/A	0		
8	DB2_BQ_LITMUS_10K	✔	N/A	07-11-2020 18:00:00	07-11-2020 18:00:36	10000	10000	N/A	N/A	N/A	0		
9	DB2_BQ_LITMUS_10K	✔	N/A	07-11-2020 12:00:00	07-11-2020 12:00:39	10000	10000	N/A	N/A	N/A	0		
10	DB2_BQ_LITMUS_10K	✔	N/A	07-11-2020 06:00:01	07-11-2020 06:00:46	10000	10000	N/A	N/A	N/A	0		

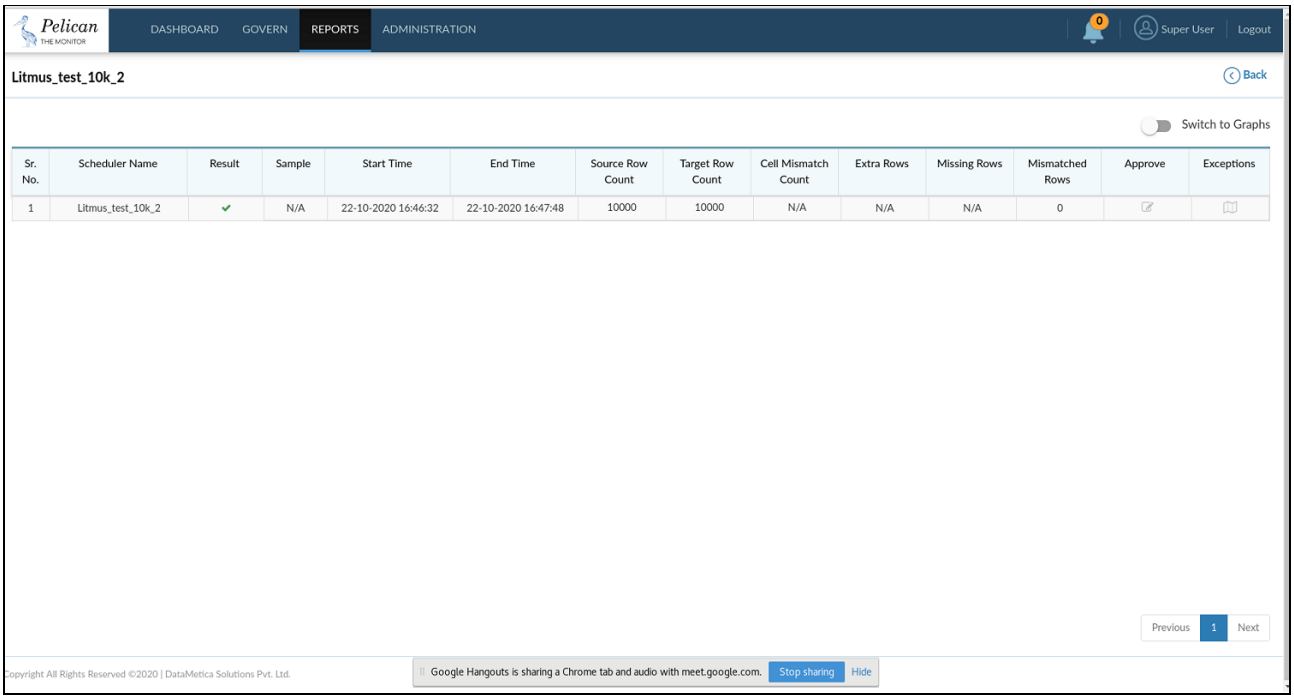
3.1. The scheduler history page displays the detailed information of the scheduler, such as:

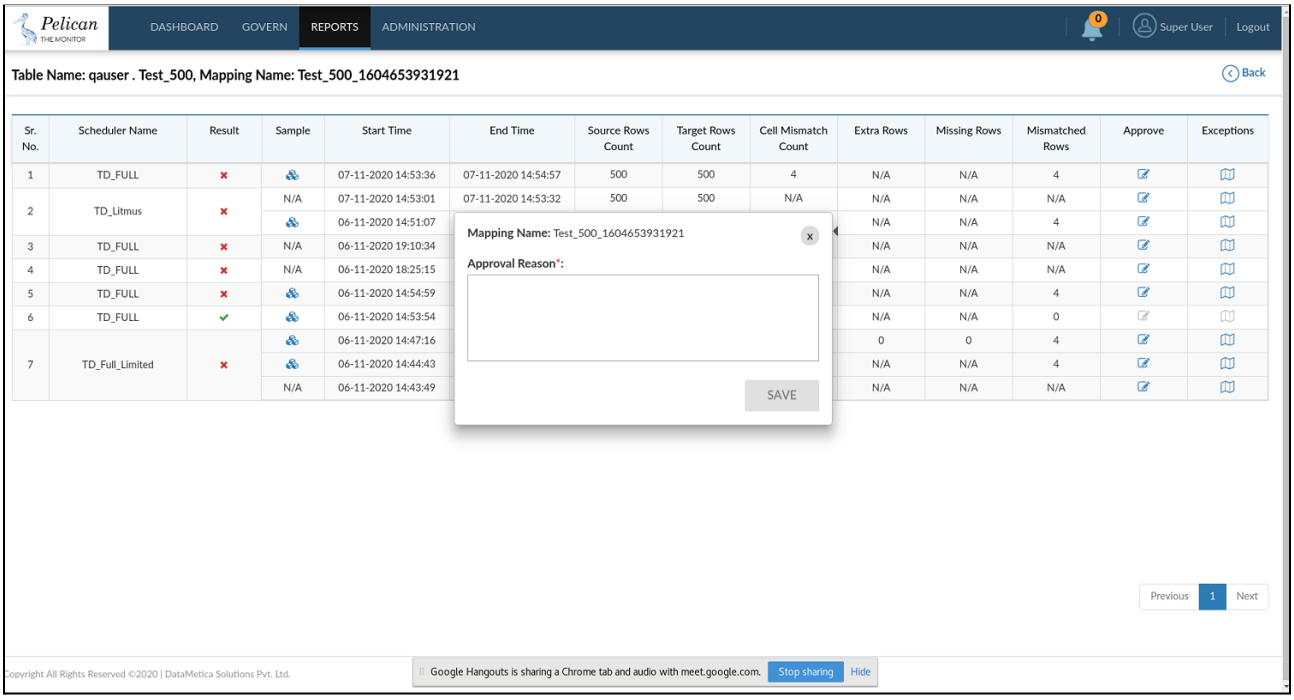
- Scheduler name
- Result
- Sample
- Start time
- End time
- Source Rows Count
- Target Rows Count
- Cell Mismatch Cout
- Extra Rows Count
- Missing Rows Count
- Total Mismatch RowsCount
- Approve
- Exceptions

Among these following are the some of the major features: -

Approve

- It is a feature to approve a mapping if a mapping fails.
- This feature has been added for the cases where the cause of failure is a known issue and can be considered as an exceptional case.
- As shown below we can approve a mapping by clicking on the icon in the approve column and click save to approve a mapping.

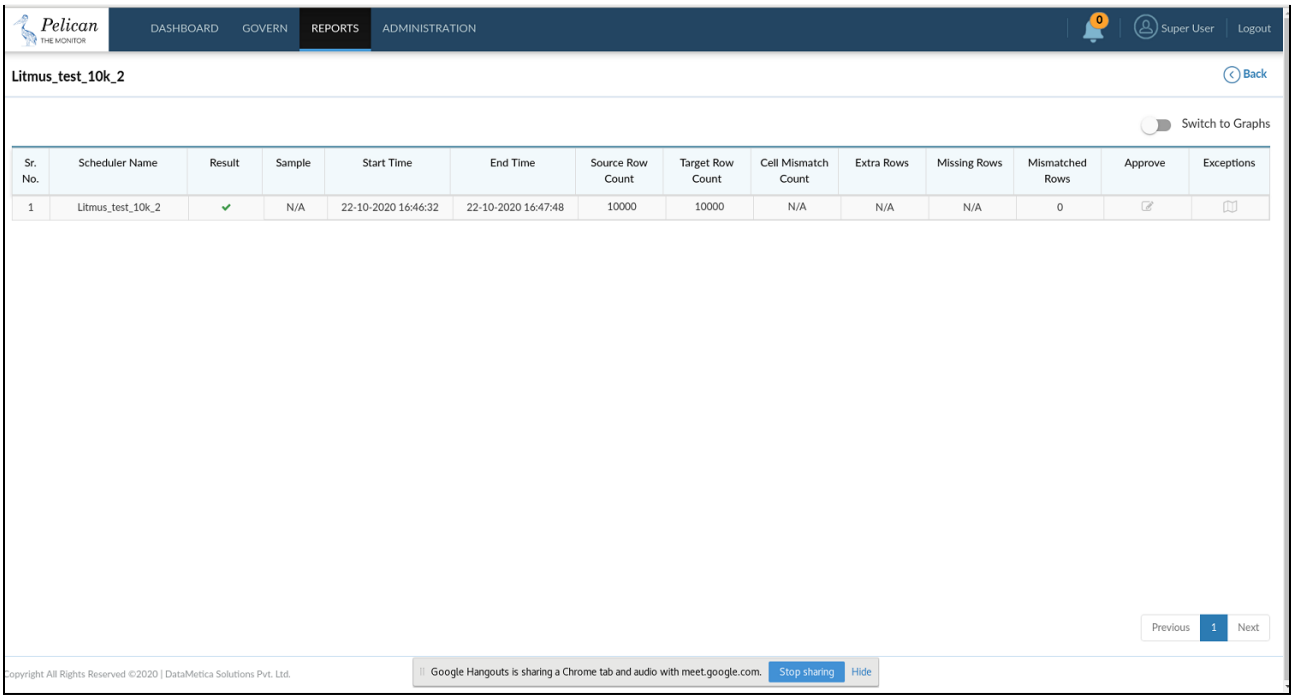


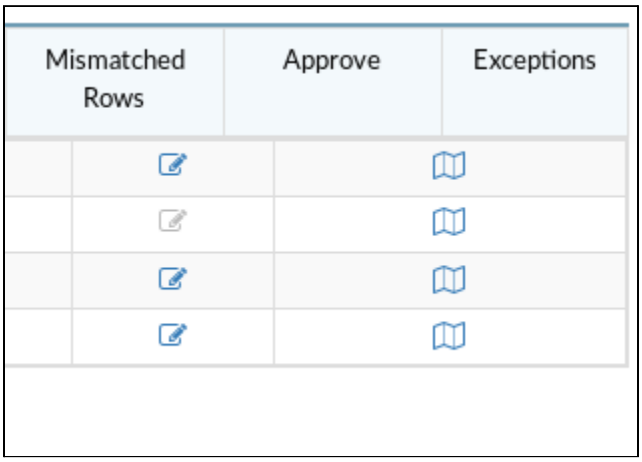


- If the result is either approved or passed, the icon to open approve a mapping modal will be disabled.

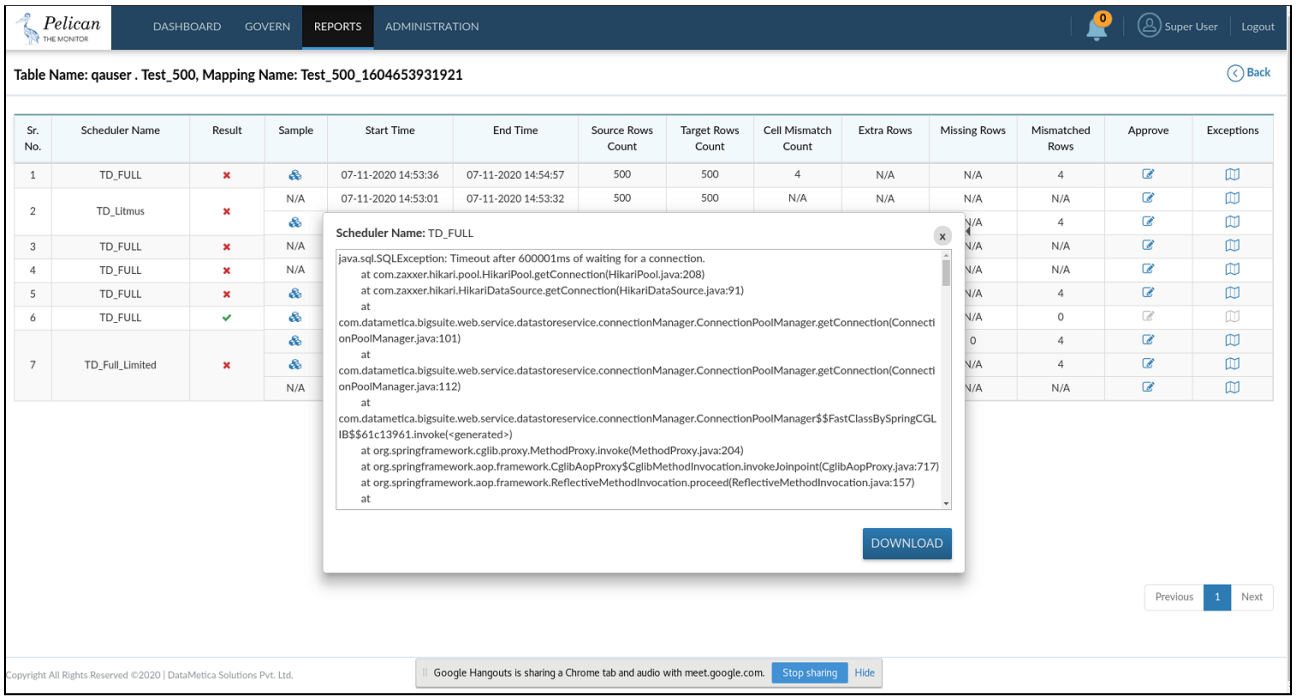
Exceptions

- If a scheduler failed due to an exception, then log will display its exception stack trace. One can click on the logs icon to exception view logs.



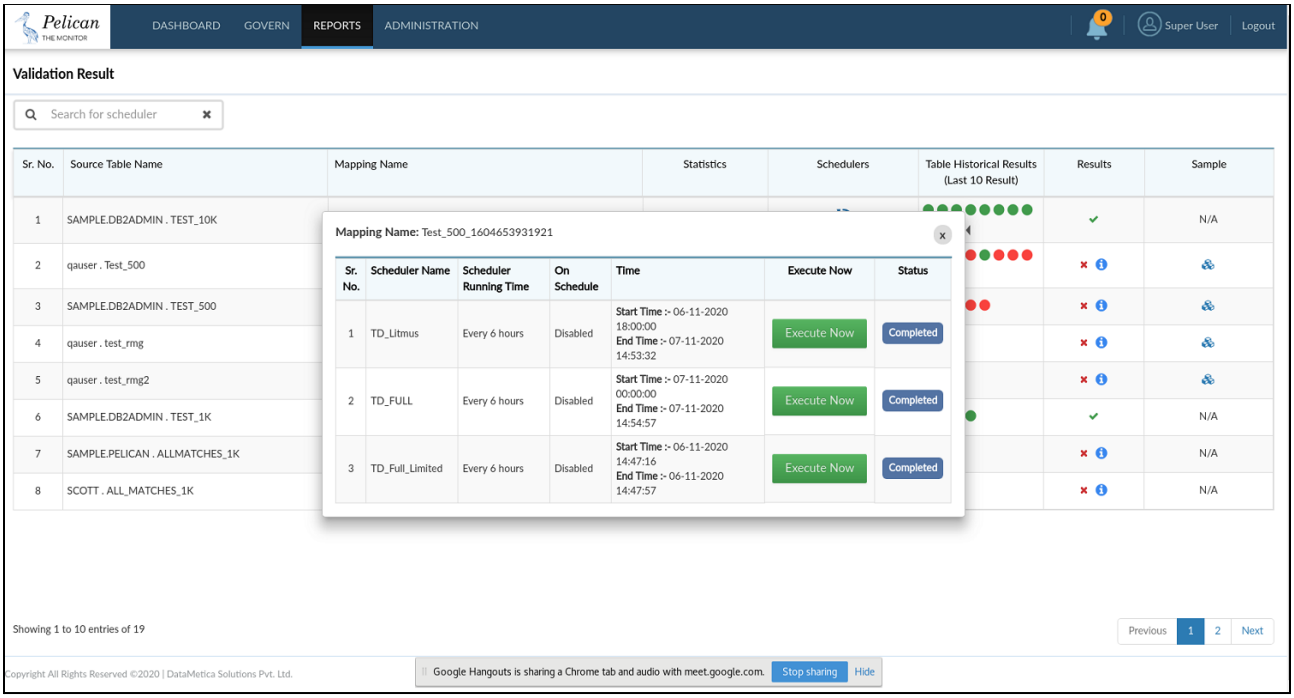


- Once we click the icon on the logs column, the new pop-up window will open for the logs.



- If Scheduler runs without Exception, then the Log option will be disabled.

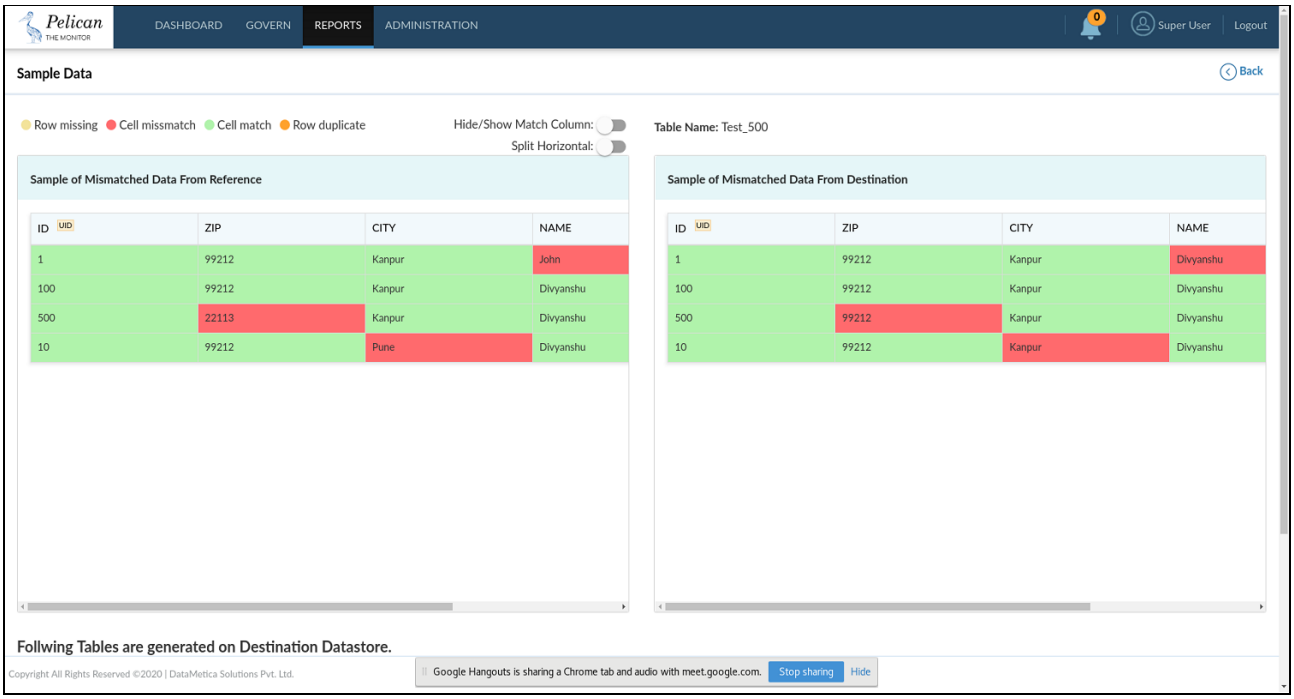
4. **Schedulers:** Click on the  icon to view scheduler details.



The screen will pop up and will show information and properties of the scheduler such as scheduler name, scheduler running time, on schedule, starting and ending time of last execution of scheduler, execution option of scheduler using Execute Now button and status of the execution.

5. **Table Historical Result**: This column displays the execution history of the schedulers of the respective mapping. This information is represented by **Green**, **Red** and **Blue** circles. The Green circle indicates the **Success**, the Red circle indicates the **Failure,** the Blue circle indicates the **Approve**. The user can hover mouse over these circles to view execution start date and end date of the respective scheduler as shown below.
6. **Results:** This column displays the data validation result. The Result is represented by two flags, namely:
   - Success
   - Failure

   Hover mouse over icon to know the reason for failure.
7. **Sample:** This column displays the sample of mismatch data in case if total mismatch count is greater than zero. Click the symbol to view samples of mismatched data. This displays the **Sample Data** screen as shown below.



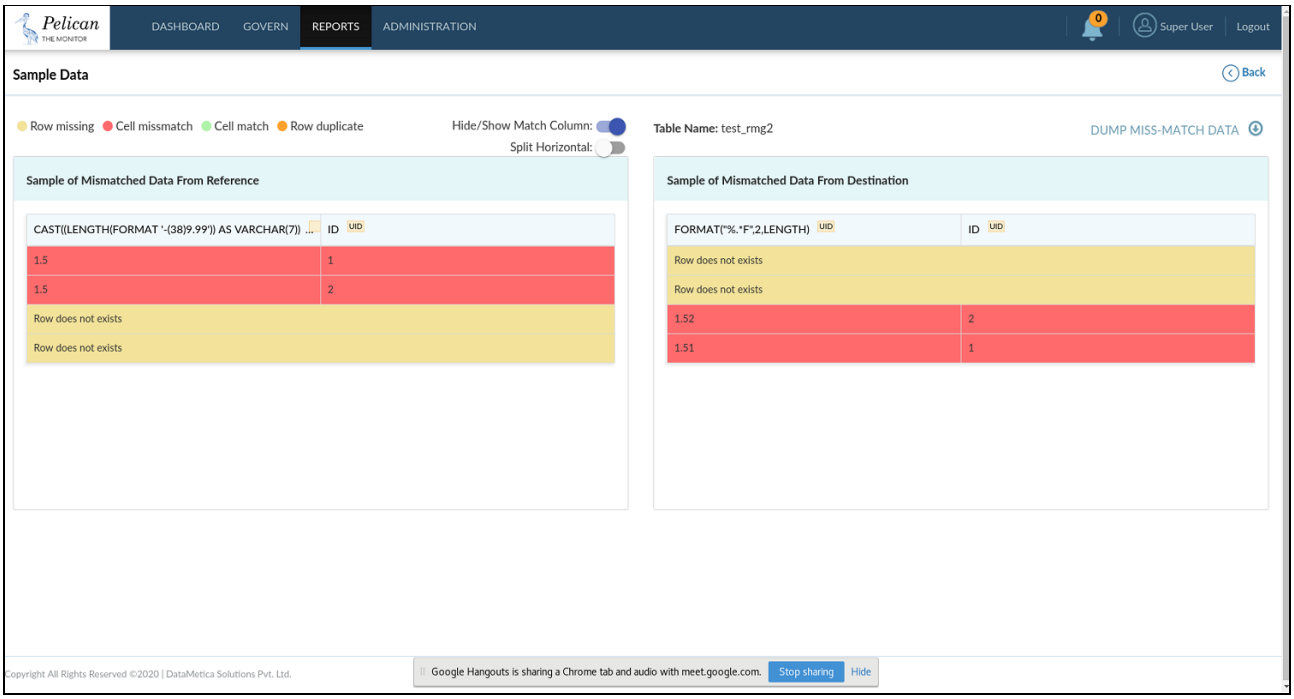
7.1. The Sample Data report displays sample mismatch data from source and target table with different colours. This report also provides the following filters which helps the user to view the desired information.

🎬 Hide/Show Match Column 🎬 Split Horizontal

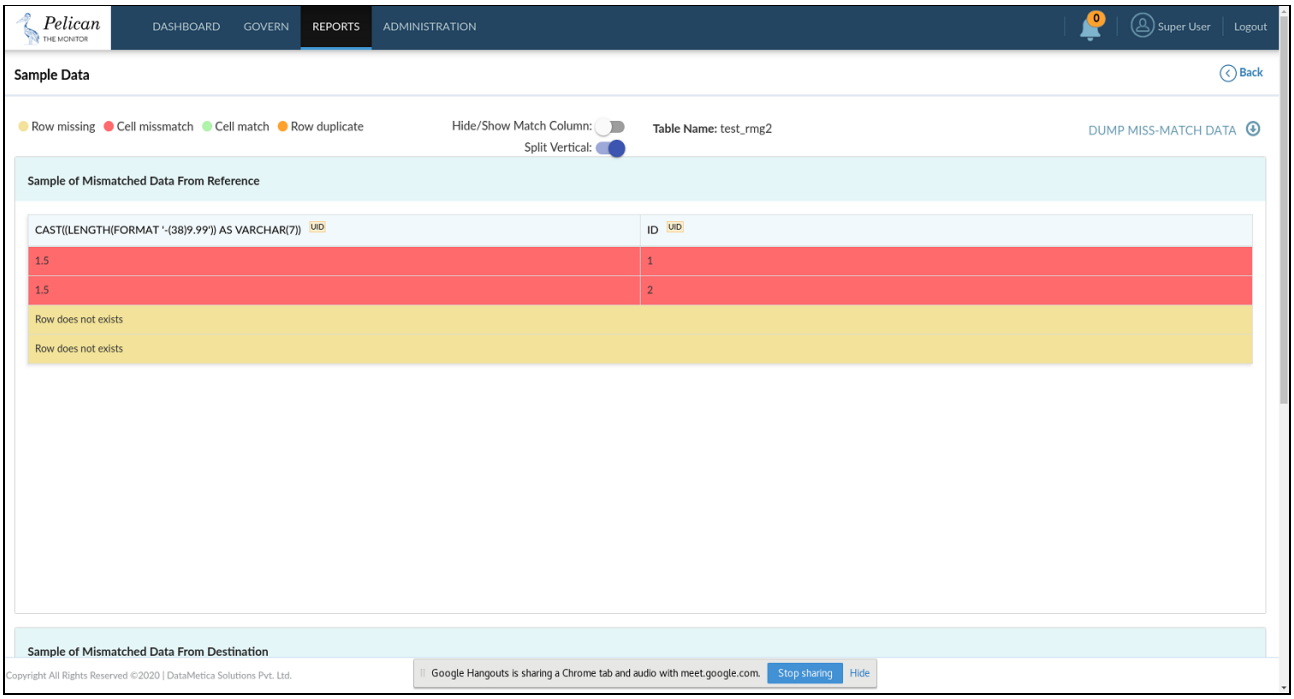
7.1.1. **Hide/Show Match Column:** The Hide/Show Match Column is a toggle button which shows two States ON and OFF.

When the **Hide/Show Match Column** toggle button is set to **OFF**, then the application displays both matched and unmatched columns as shown above.

When the **Hide/Show Match Column** toggle button is set to **ON**, then the application displays only unmatched columns as shown below.



7.1.2. **Split Horizontal:** Set **Split Horizontal** toggle button to **ON** to view mismatch data in the horizontal format, for more information, refer below image.



8. **Backtrack**: Backtracking functionality helps users to see the execution history and its results for the particular table.

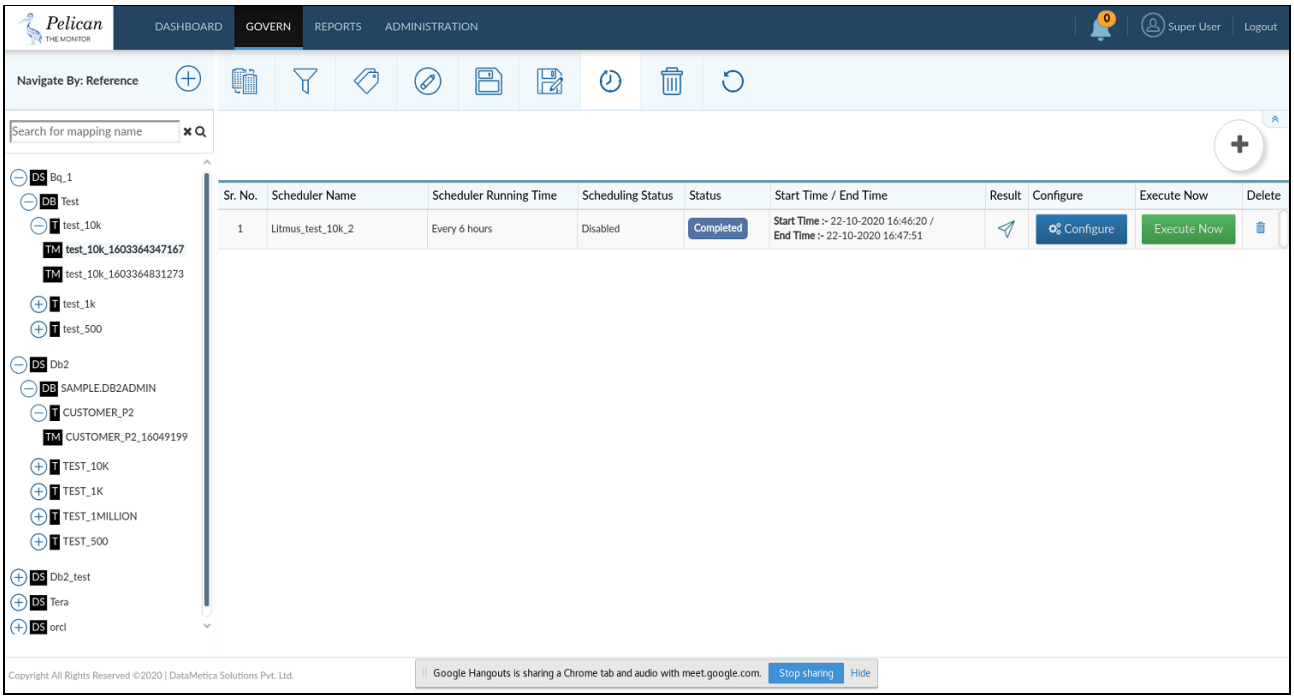
   It shows the graph containing lineage for the selected table. Also, it shows the execution details of all the tables present in the graph.

9.2. Results from Scheduler page

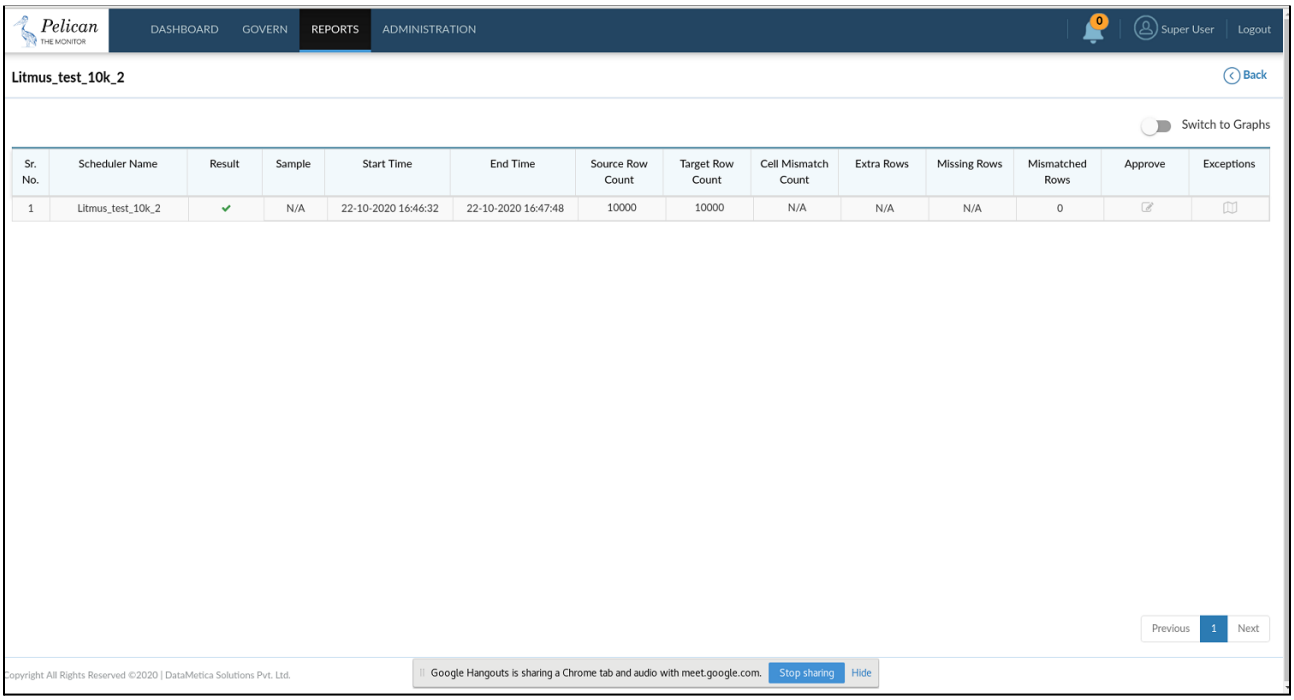
Here, the user can view validation results for the saved mapping.

To view validation result:

1. Go to the **Validation Configuration** page and select the saved mapping for which you want to see the validation result and then click on the view scheduler icon  .The following screen will pop up.



2. Click on the icon  to see the statistics of the validation result page for the mapping.



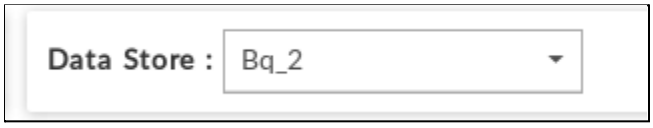
The fields (columns of the validation result page are the same as explained above).

10. Dashboard

Dashboard screen displays result, trends, graph of the scheduled mappings based on datastore, database, tags, and date. When for the first time the dashboard screen is loaded it will display the dashboard screen for the default datastore.User can select datastore and databases to see the corresponding graphs.

10.1 Filter option for the dashboard Screen

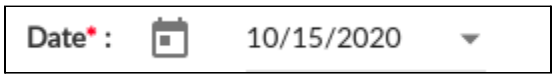
1. Datastore
    - We can select the datastore for which we want to see the dashboard screen.



2. Database
    - We can select the databases for which we want to see the dashboard screen of that particular datastore.
    - By default, all databases are selected for the particular datastore.

3. Date

- Result trend has an option to select duration ,for Pelican db result - db wise and donut screen we do have an option to select a date for which we want to see the result

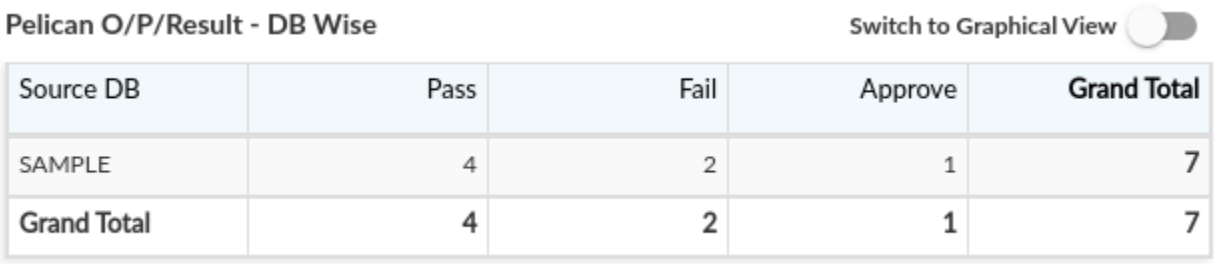


10.1. Different Results on the dashboard screen

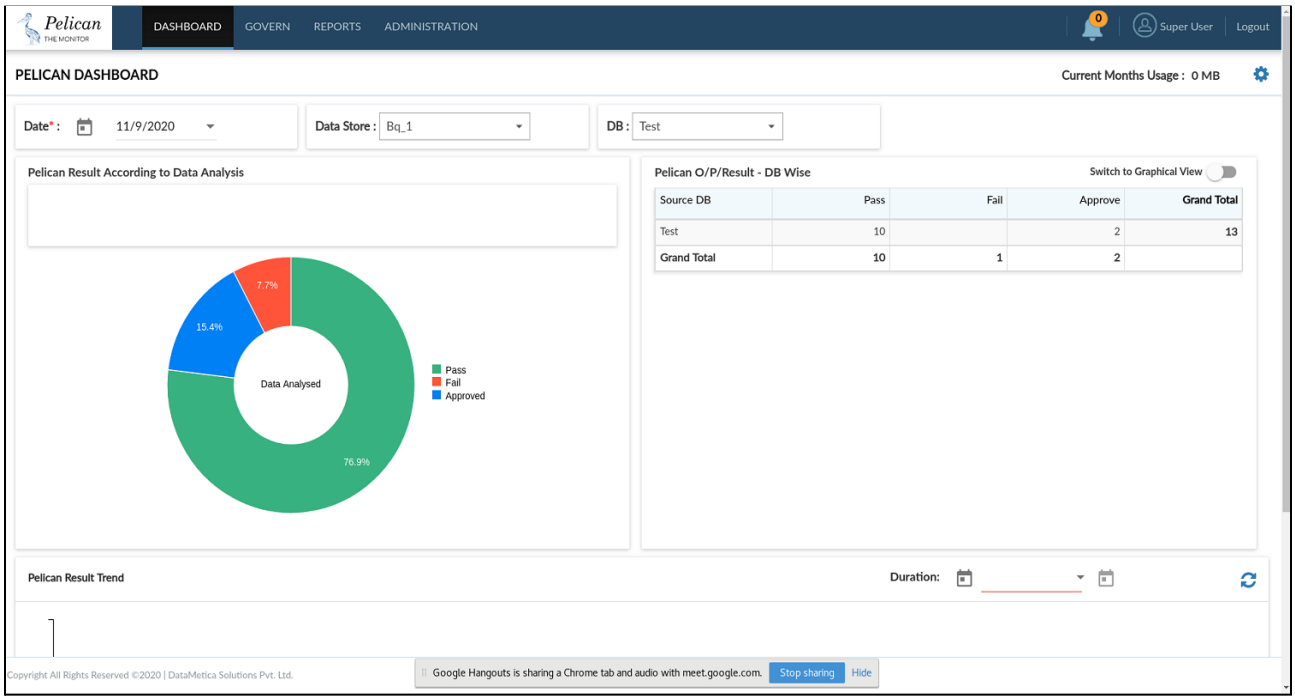
We have following graph result trends that is being displayed on the pelican screen:

10.1.1 10.1.1. Pelican output result db wise

- It displays total pass/ fail/ approve count of the scheduled mappings for the selected database.
- This info can be seen in two formats shown below, users can switch between graphical and tabular format using the toggle option present on the upper right corner.
- In graphical format Green represents total pass count, Red represents total fail count and Blue represents total approve count.

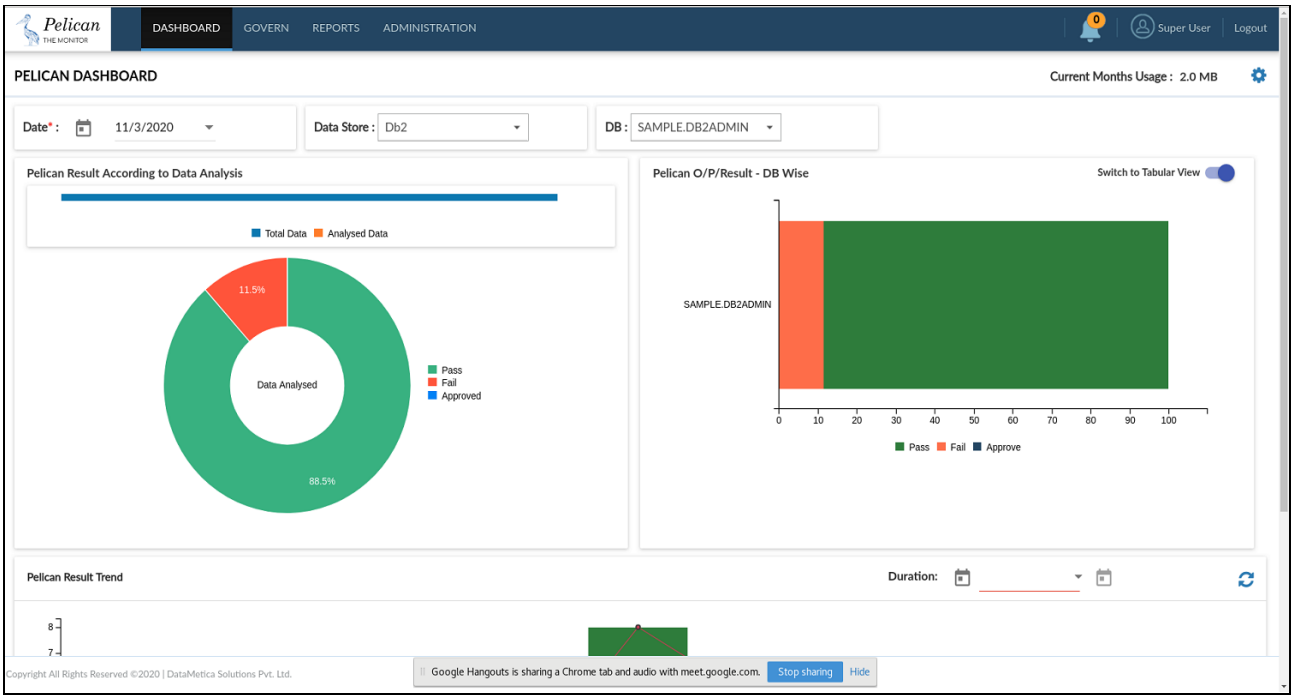


Source DB	Pass	Fail	Approve	Grand Total
SAMPLE	4	2	1	7
Grand Total	4	2	1	7



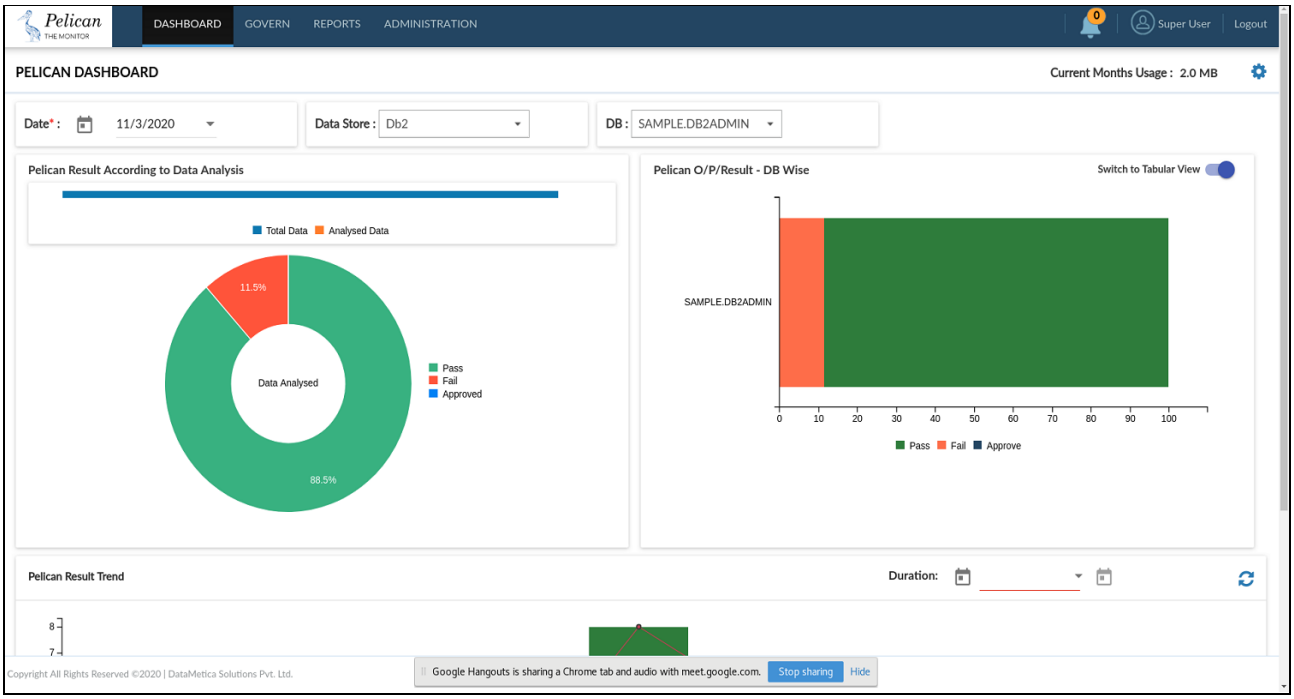
10.1.2 10.1.2. Pelican Pass/Fail/Approve Percentage

- It represents the total pass/fail/approved percentage of the total scheduled mappings for the particular datastore and databases.



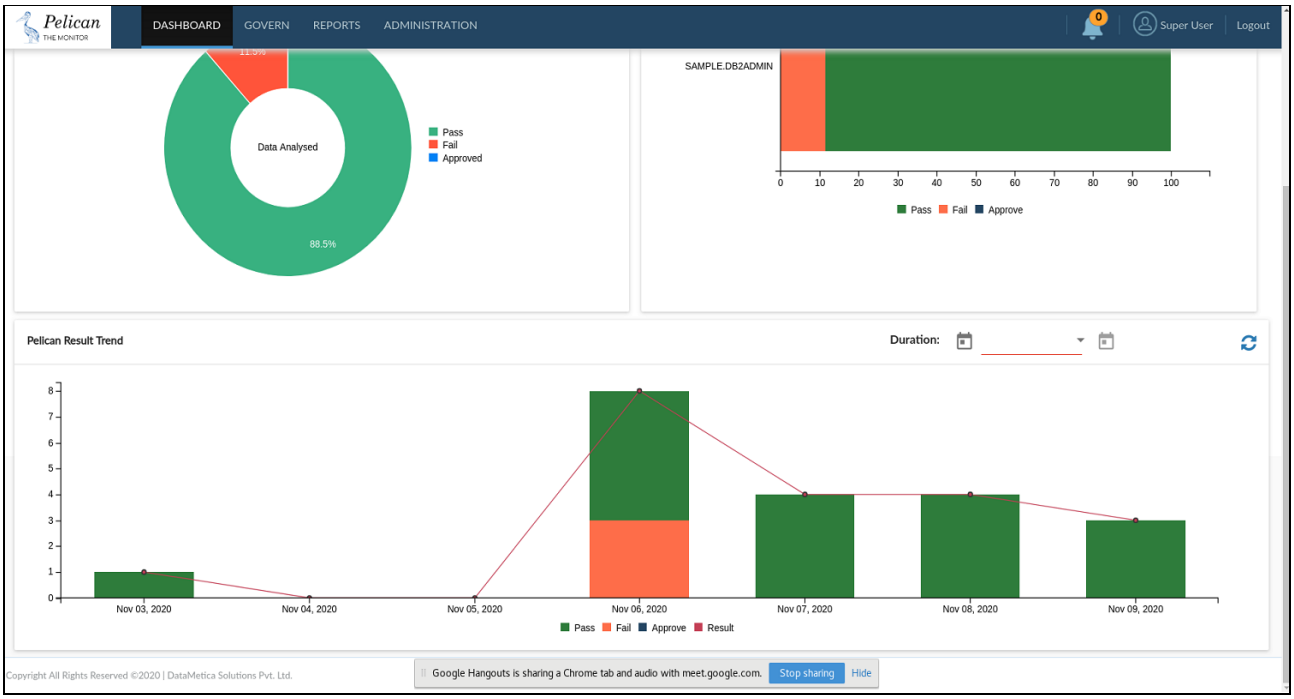
10.1.3 10.1.3. Pelican Result trend

- It represents the total pass/ fail/ approve counts date wise for the selected duration, databases, and tags for particular datastore.
- By default, duration is seven days from the first time the mapping containing the datastore has been executed.
- Green represents total pass count; Red represents total fail count and Blue represents total approve count.



10.1.4 10.1.4. Pelican table usage

It represents the number of tables for which at least a single scheduler has been run and the total number of tables in a particular data store. Blue represents tables for which at least a single scheduler has been run whereas Orange represents the total number of tables in a particular data store.



11. Add Licence - Buy Your Own license (BYOL)

Buy Your Own license (BYOL) notifies you when your application will expire. Pelican has introduced a new feature, i.e Add licence. This new feature notifies you, the date of the renewal.

11.1. Adding /updating licence

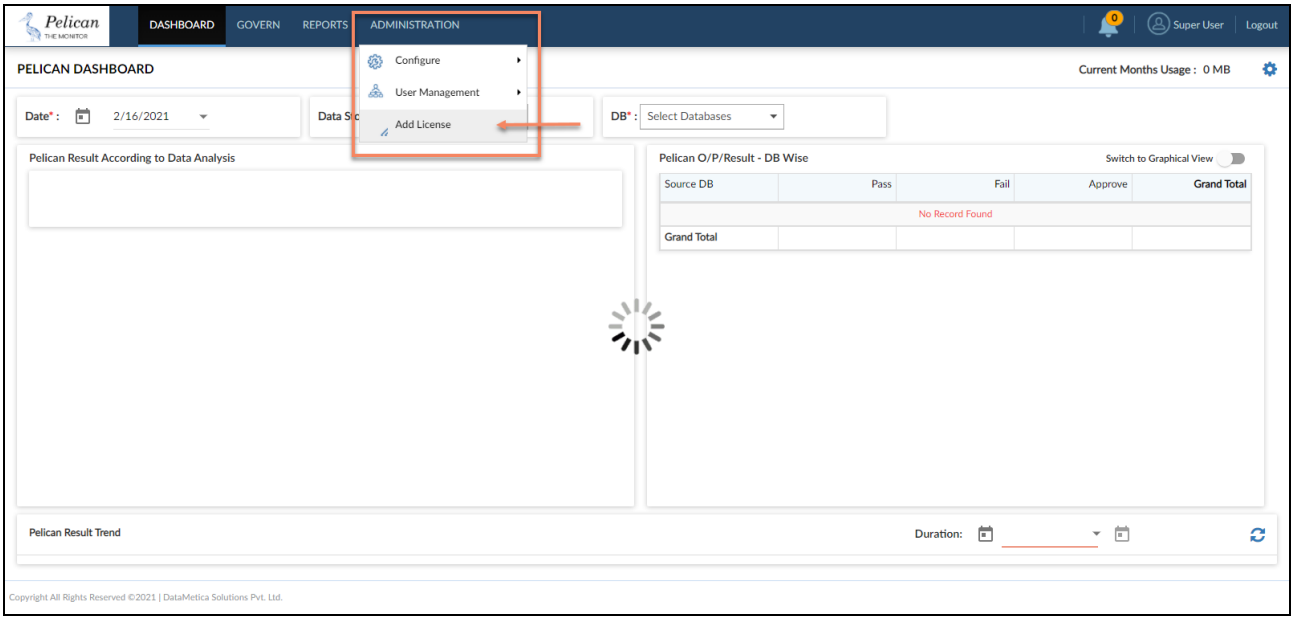
Once you received the pelican licence file, i.e. licence.pel follow the following steps.

Steps 1; log in to the Pelican

Step 2: Administration -> Add Licence

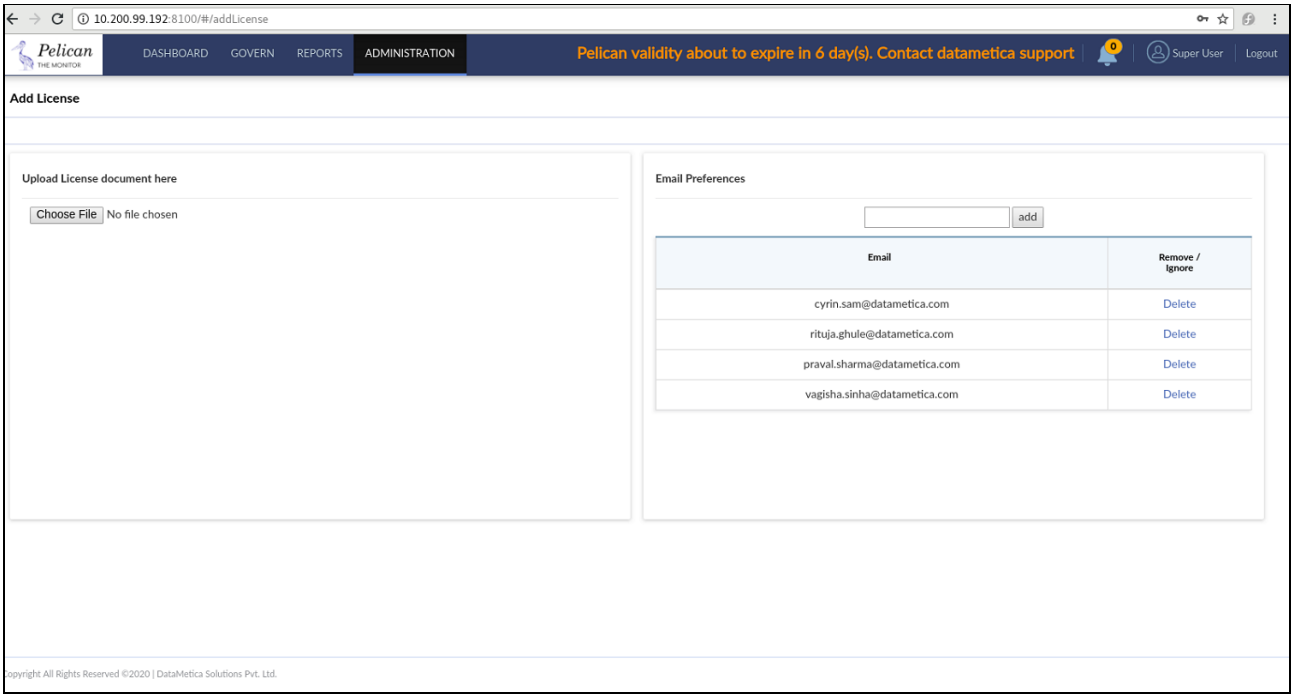
Step 3: Click on the upload and select the Pelican.pel from the file system.

Step 4: Again login to continoue the usage.



11.2. You will be notified in three ways.

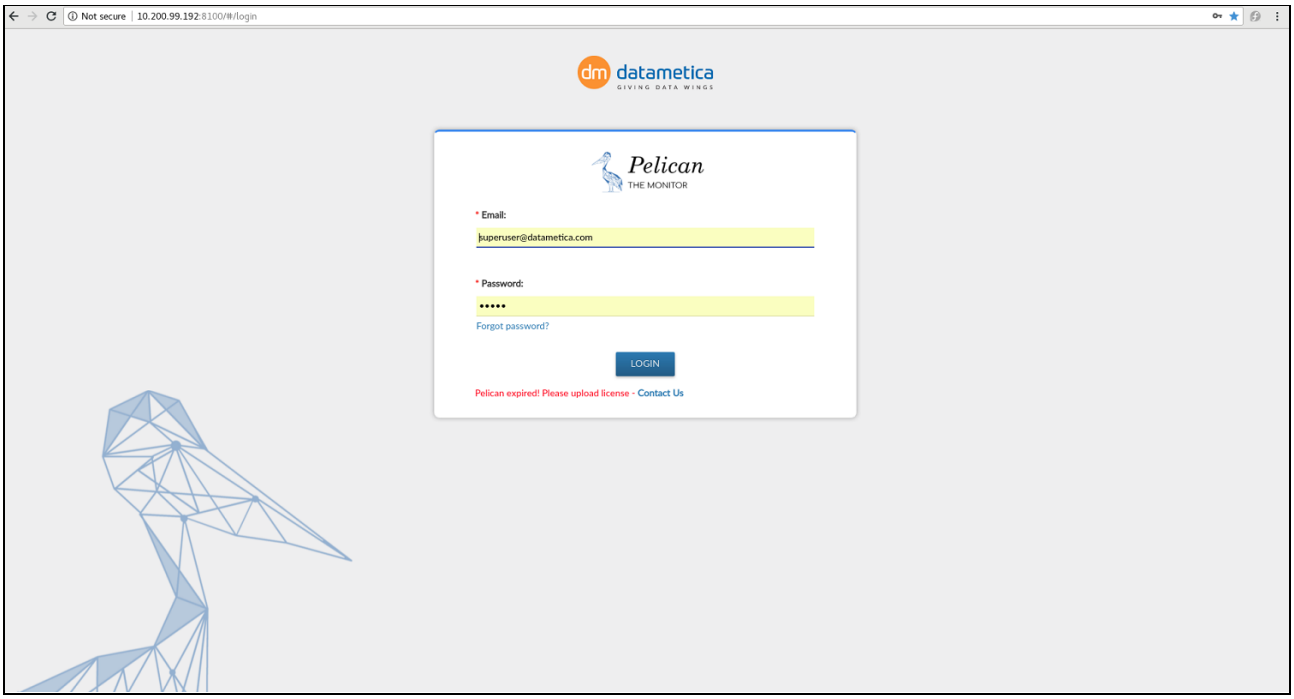
1. Detailed email is sent to your registered email address prior to 15 days of renewal , mentioning your product ID and date of renewal. All the registered email addresses will receive the email of notification.



2. On the login screen, below the login tab,

Eg.

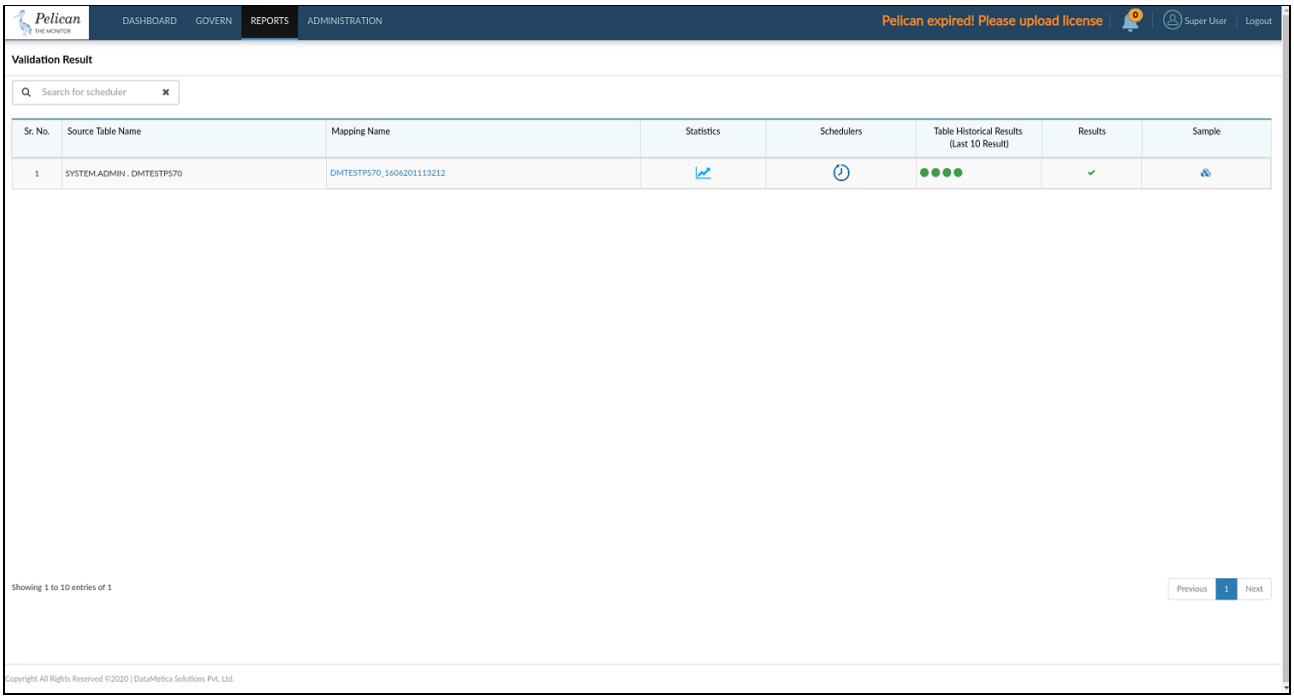
Pelican validity is expired. Contact Datametica support, Contact Us.



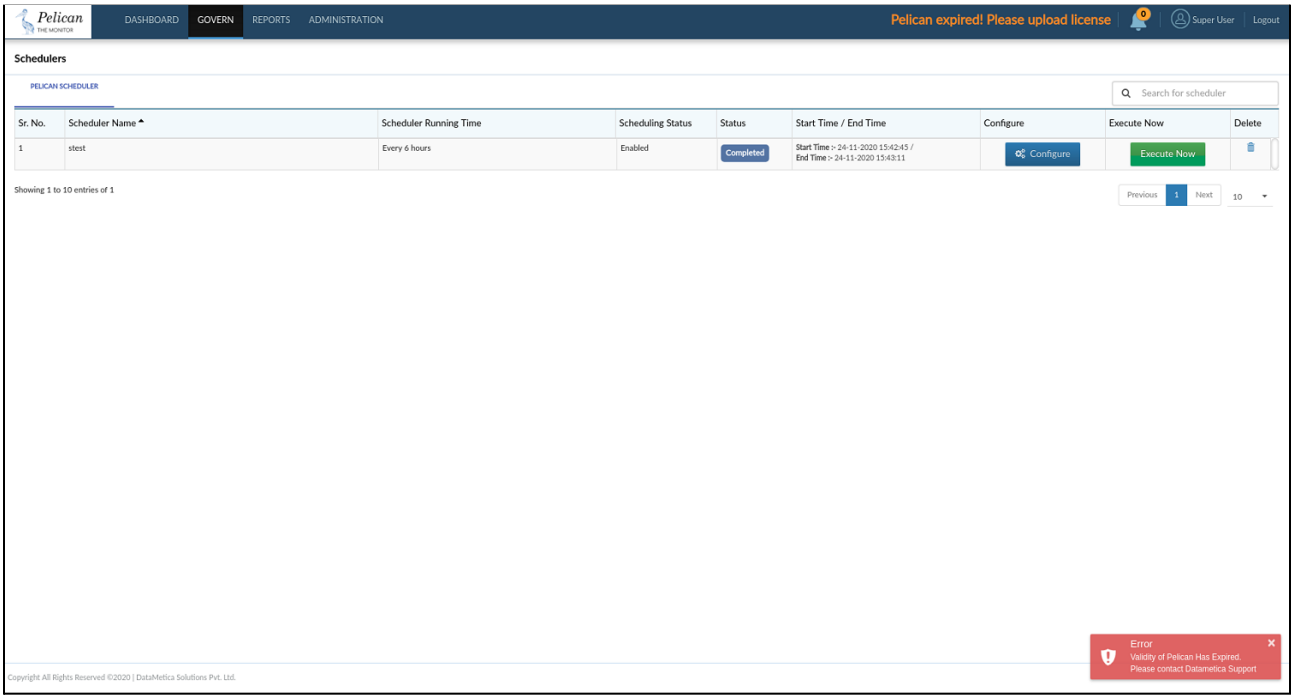
3. On the main screen, on the menu bar,

Eg.

Pelican validity is expired. Contact Datametica support, Contact Us.

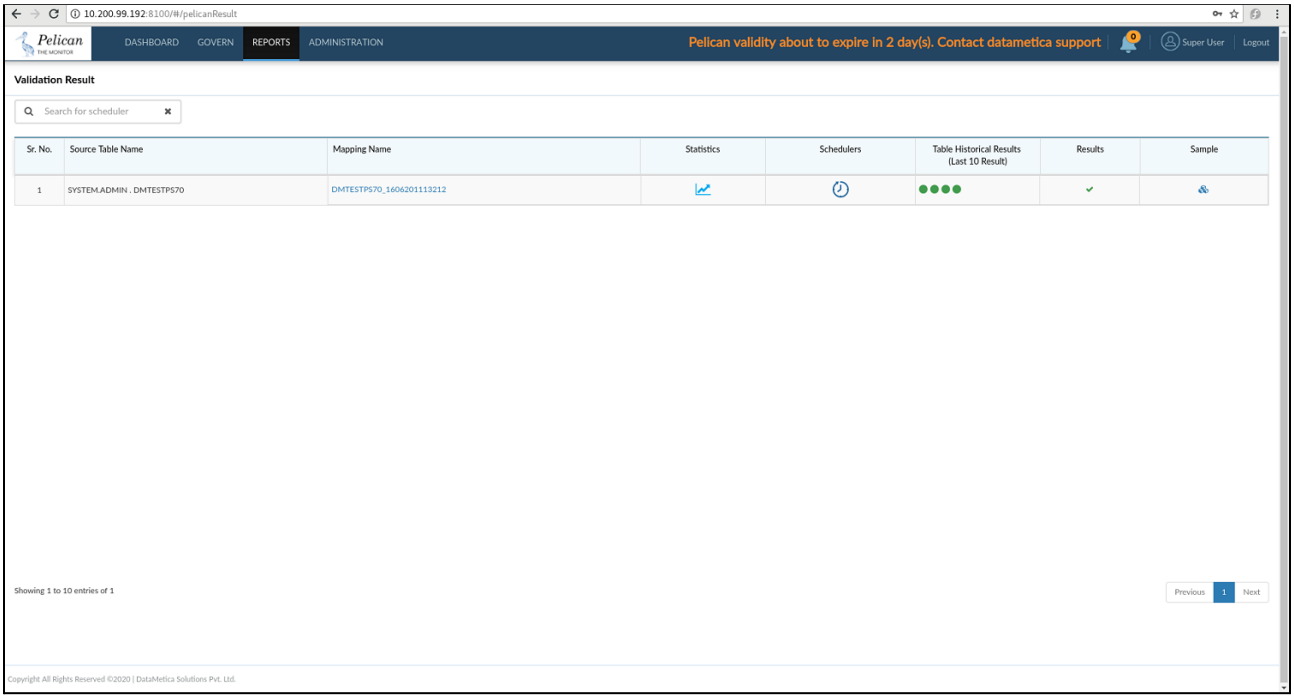


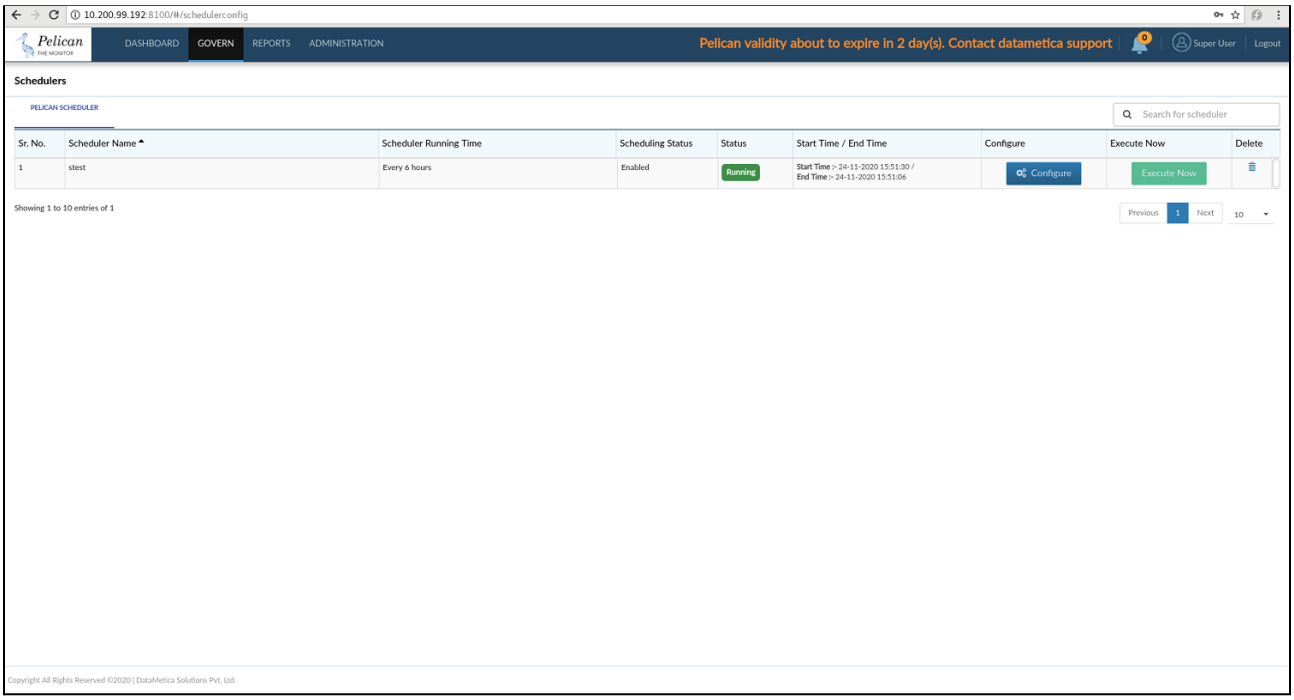
Once your license is expired, you are able to create the mappings, but cannot run the schedulers. To continue, you need to contact Datametica..



Once you renew the licence, all the functionalities are resumed.

Now, you can run schedulers.





12. Frequently Asked Questions

1. 12.1. How to view the validation result in Pelican?

The validation result can be viewed only for those tables whose table mapping is done. So, if you are a new user and the validation mapping has not been done yet then you need to follow the below given steps:

Step 1: First, you need to configure a data store.

Step 2: Followed by datastore configuration, you need to validate tables between source and destination datastore.

Step 3: Next, you need to configure the scheduler for the saved validation mapping

Step 4: Finally, go to Reports and click Validation Result. For more information, refer to the *Validation Result* topic.

12.2. Which are the various data stores the Pelican supports?

Pelican supports comparison between various data stores. For more information, refer to the *Data Store Support* topic.

2. How to validate tables between source and destination datastore?

You can validate tables between source and destination Datastore using Validation Configuration functionality. For more information, refer to the *Validation Mapping* topic.

3. 12.3. What is the scheduler and how to configure it?

The Scheduler allows the user to execute processes at a regular time interval. In Pelican, the user can create a scheduler for a saved mapping; so that, after a time period the scheduler executes the process and it validates the source table with destination tables. For more information, refer to the Scheduler Configuration topic.

4. 12.4. How to configure the email notification?

Email notification functionality allows the user to send an email automatically to various users
on the execution of the scheduler. For more information, refer to the *Email Notification* topic.

12.5. How to update the pelican licence?

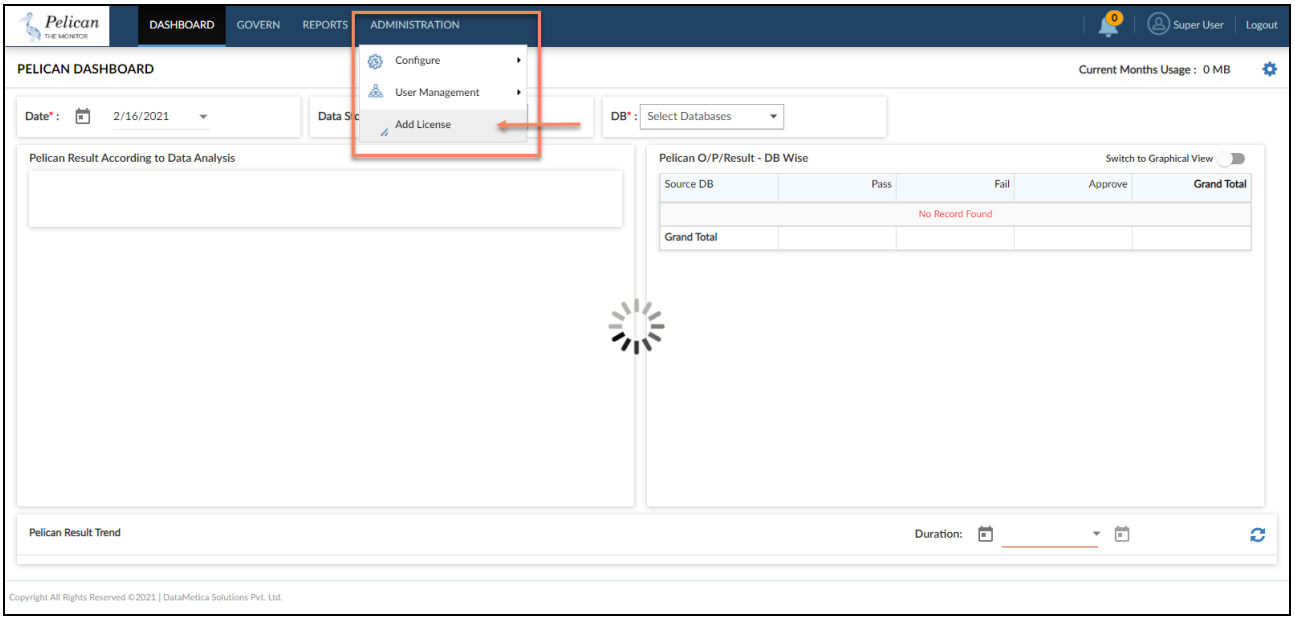
Once you received the pelican licence file, i.e. licence.pel follow the following steps.

Steps 1; log in to the Pelican

Step 2: Administration -> Add Licence

Step 3: Click on the upload and select the Pelican.pel from the file system.

Step 4: Again login to continue the usage.

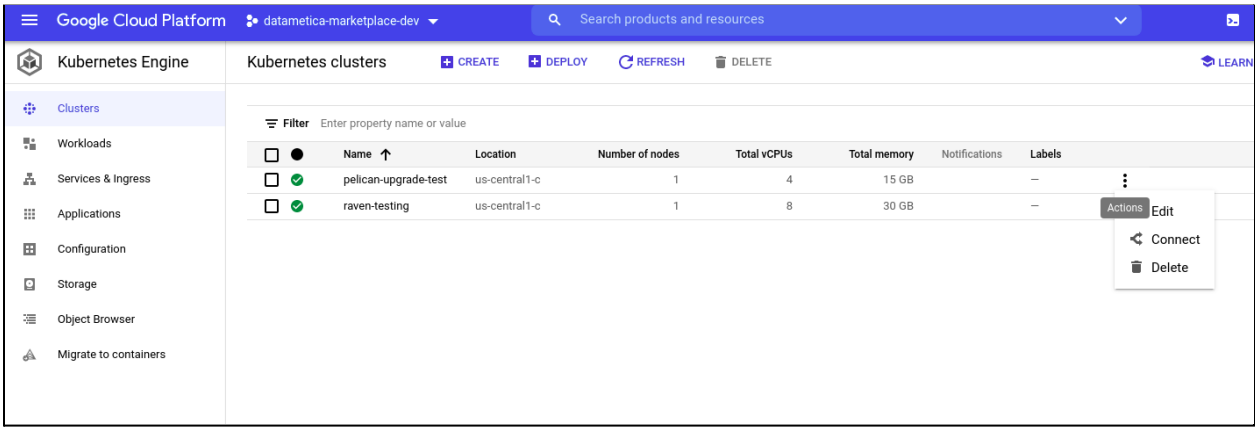


12.6. Who can create LDAP users?

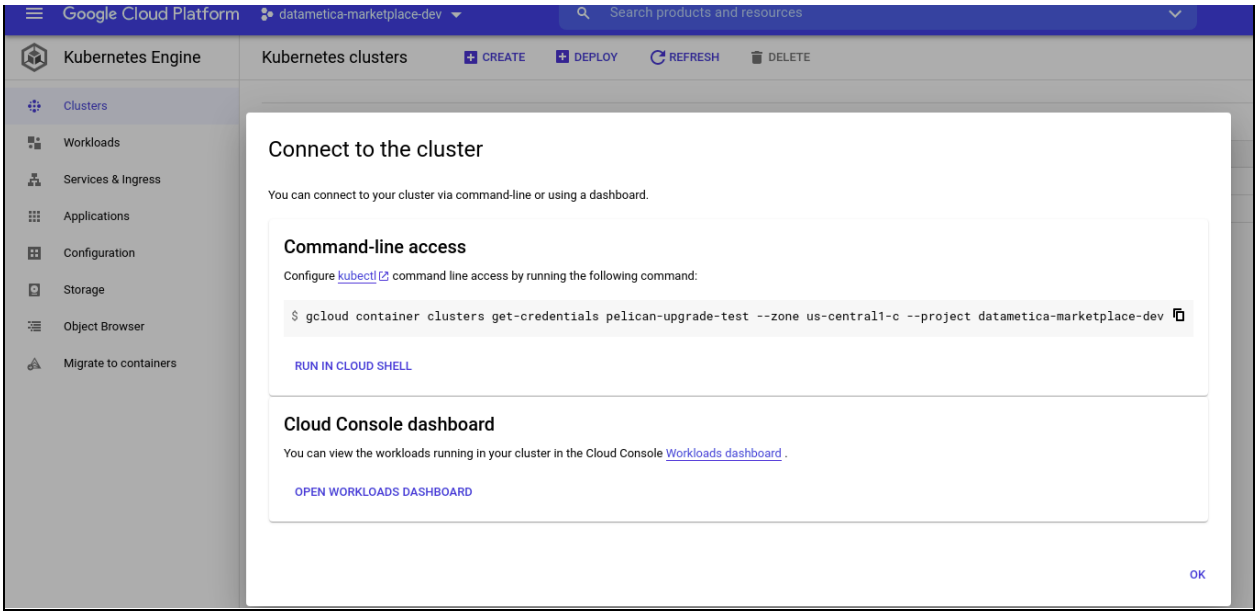
Only the superuser has the right to create LDAP users. LDAP users can't create any new users. Only users with Admin access can create new LDAP users.

12.7. How to get GKE server(Pelican services) start-stop access permissions to QA Lead?

Step 1 - Go to GCP console > Search for GKE
Step 2 - Click on Connect your cluster



Step 3 - Click on Run in cloud shell



Step 4 - Run this command after login on cluster - **kubectl get pods**
Step 5 - Copy the web pod name and run this command to restart the service - **kubectl delete podname**

Note - Pelican service pod will be automatically started and old one will be deleted.

12.8. How to get Application log file url along with access permissions to QA users (catalinaout.log and application.log) ?

Step 1 - To see the application logs, we need to go inside the pelican pod. Run this command to go inside the pod - **kubectl exec -it podname bash**

Step 2 - Now you can view the logs on this file path - **/usr/local/apache-tomcat-8.5.65/logs/application_logs/log.log**