DvSum Data-PARC User Guide

Edition **3**

Published **on July 25, 2019**

DvSum, Inc.
440 N. Wolfe Road
Sunnyvale, CA 94085
U.S.A.

Welcome

Thank you for signing up for DvSum DataPARC. You have started on a journey to data quality nirvana. If you are reading this user guide, that means you have already received an introduction email upon sign-up on www.dvsum.com.

This guide will help you get setup and start using DataPARC with your real production data in less than an hour.

Happy Data Validation!

Sincerely,

DvSum Team

Contacting DvSum

For any questions regarding Account Information, Technical Support or inquire about our additional services, you can contact us 24x7 on the following channels.

Help on DvSum	Help widget on bottom left https://prod.dvsum.com
Self-service Help Desk	https://support.dvsum.com
Email support (fastest response)	support@dvsum.com
Phone	+1.844.855.3282

1 Table of Contents

1 Configuration (Mandatory) for Account Owner

This chapter will go through the basic steps required for account owner to configure the DataPARC system for use. These steps need to complete before you start using the application or invite other user to use the application. If you are not the account owner, you can skip this section.

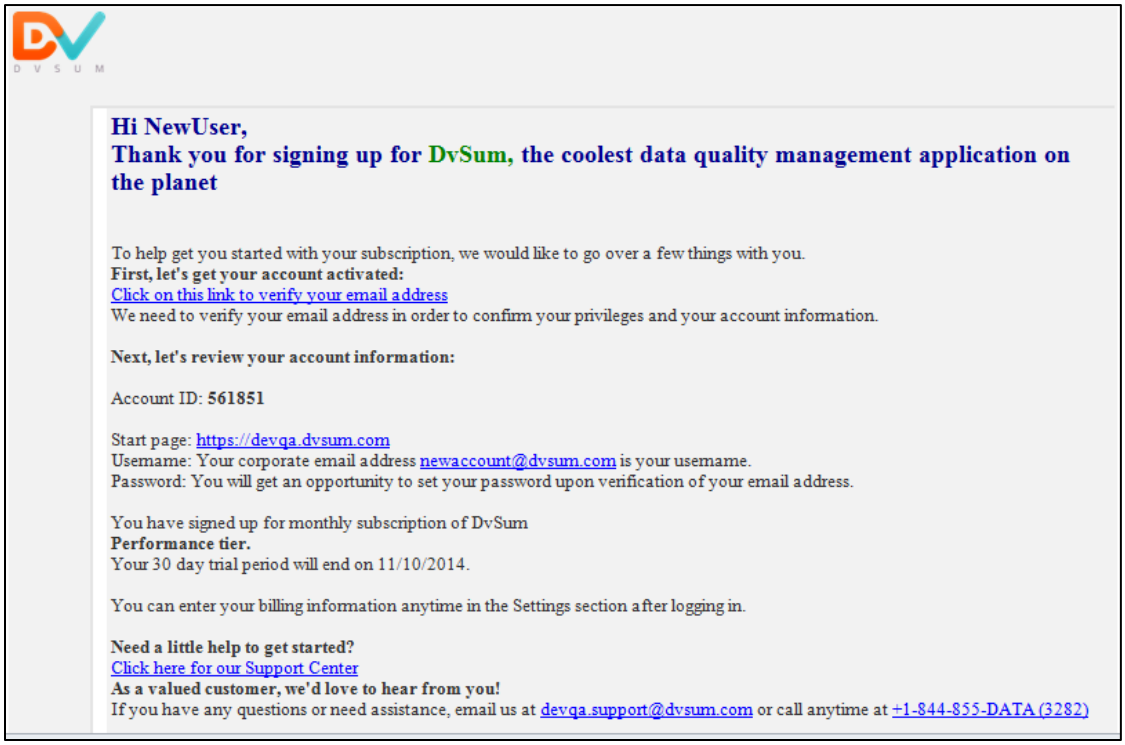
Applicability: Owner

System Requirements:

Operating System	Windows XP, 7, 8, 8.1, 10 or Windows Server 2008, 2012 and above.
Middleware	Java Runtime (JRE) 7.x (Java 7 dependency for EXCEL only) Java Runtime (JRE) latest version
Disk Space	1 GB (free)
Memory	4 GB
Browser	Internet Explorer 11 and higher Mozilla Firefox, Google Chrome, Apple Safari.

1 Verification of email and Setting up your security credentials

The welcome email that you received should look something like this.



Hi NewUser,
Thank you for signing up for DvSum, the coolest data quality management application on the planet

To help get you started with your subscription, we would like to go over a few things with you.
First, let's get your account activated:
Click on this link to verify your email address
We need to verify your email address in order to confirm your privileges and your account information.

Next, let's review your account information:

Account ID: **561851**

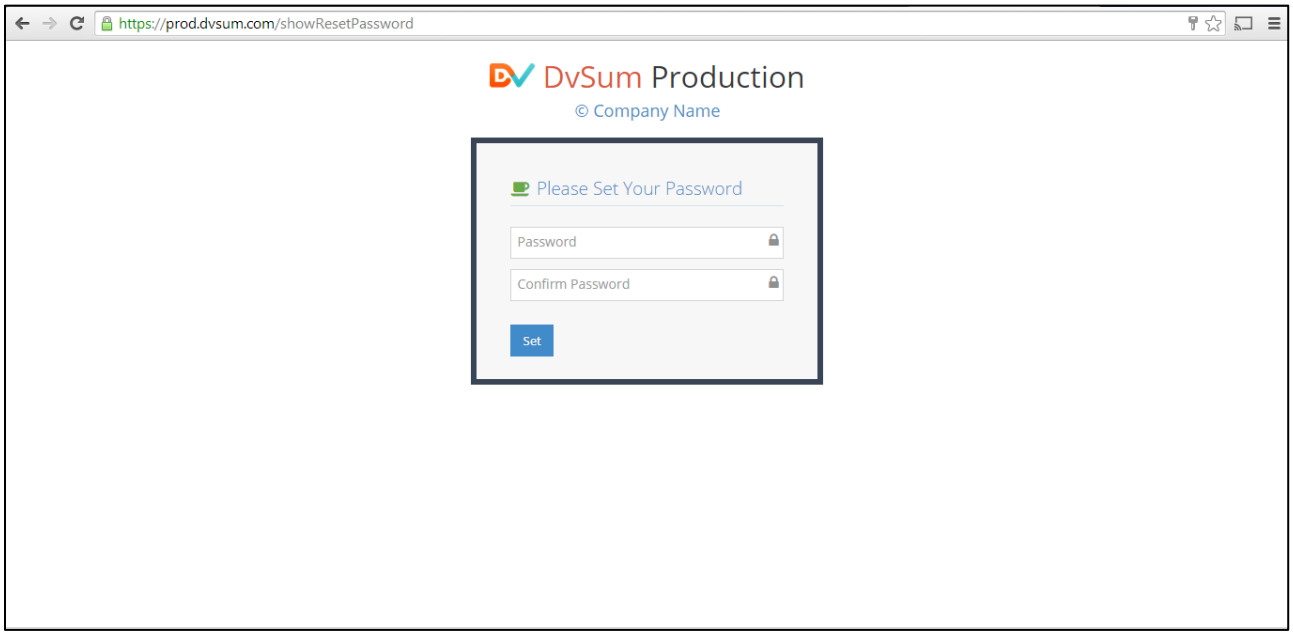
Start page: https://devqa.dvsum.com
Username: Your corporate email address newaccount@dvsum.com is your username.
Password: You will get an opportunity to set your password upon verification of your email address.

You have signed up for monthly subscription of DvSum
Performance tier.
Your 30 day trial period will end on 11/10/2014.

You can enter your billing information anytime in the Settings section after logging in.

Need a little help to get started?
Click here for our Support Center
As a valued customer, we'd love to hear from you!
If you have any questions or need assistance, email us at devqa.support@dvsum.com or call anytime at +1-844-855-DATA (3282)

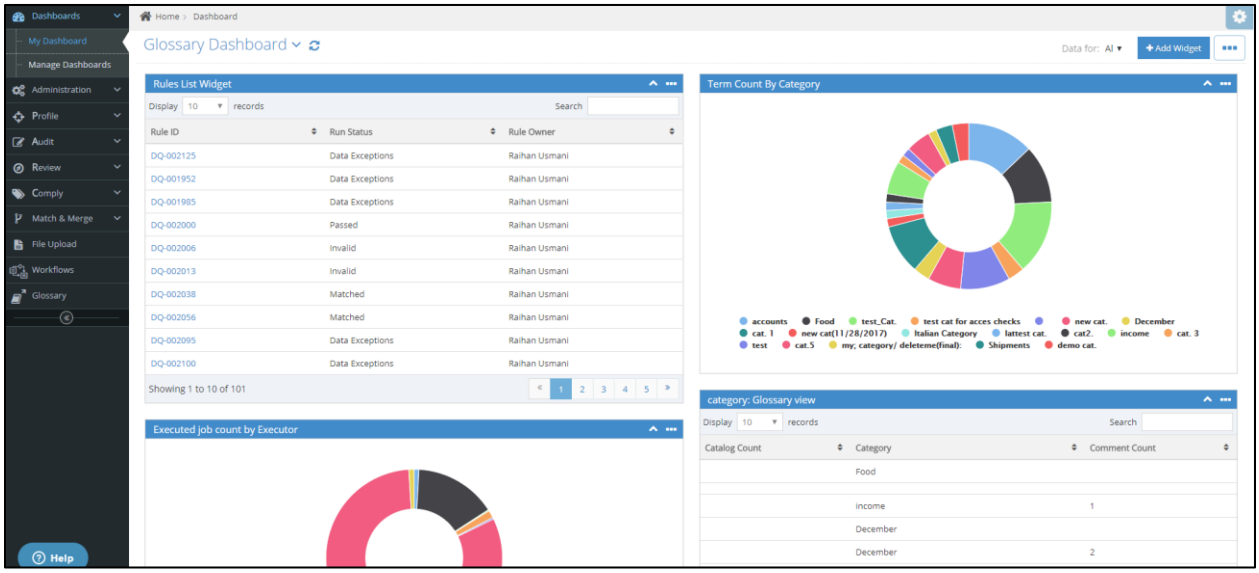
Go ahead and click on the link to verify your email address. It will open up a page where you can set your password.



Once your password has been set, you will be logged into the application automatically. You can login from your computer, tablet or laptop using any of the browsers. (Internet Explorer, Firefox, Chrome or Safari)

That's it. You are in! You should see the landing page for the application which is the dashboard.

(More details on dashboards is in the last section: **Demo Data**)



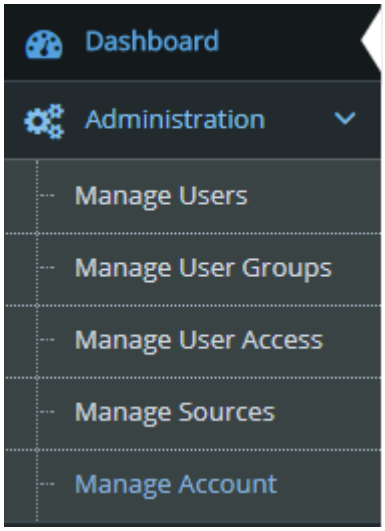
2 Owner/Administrator Quick Reference

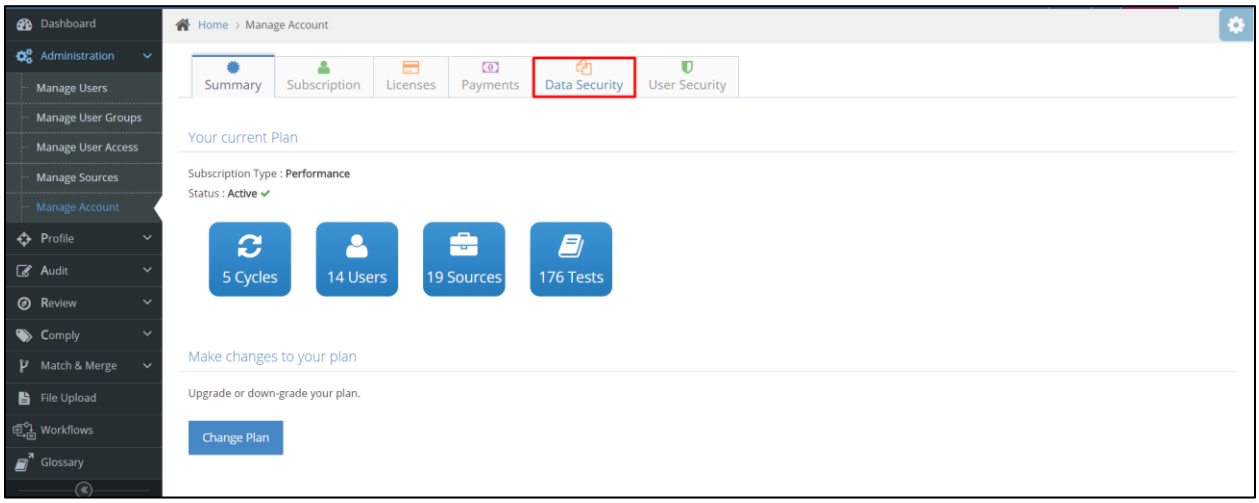
This section is a quick reference that all administrators or owners can use for administrator specific actions within DataPARC.

Applicability: Owner only

2 Configuring Security

Navigate to Data Security by clicking on Account on top right under your name and clicking on Data Security Table.



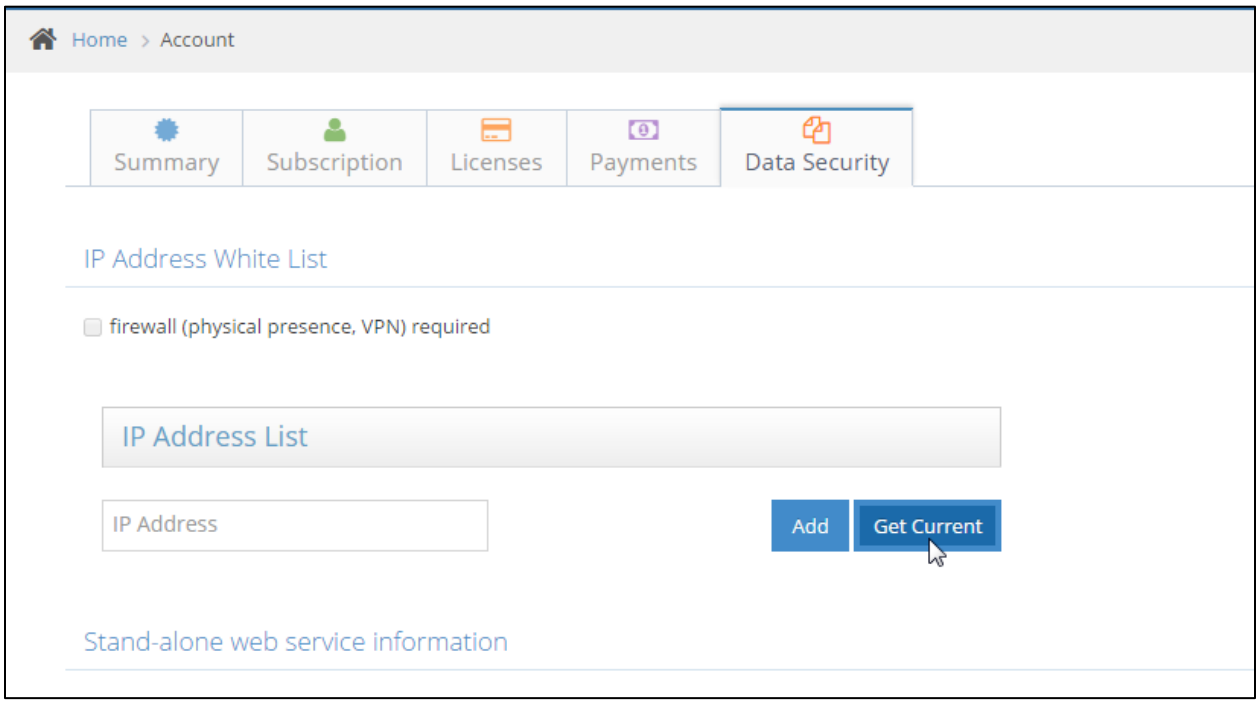


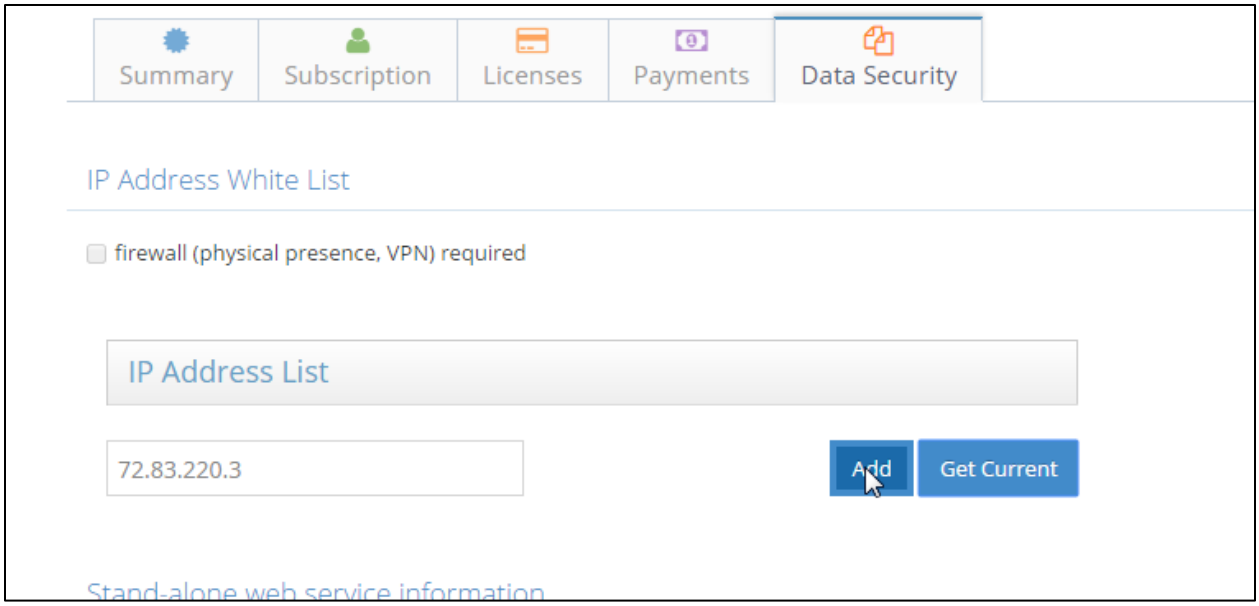
Scroll to the IP Address White List section. This is where you would indicate the domain within which users need to be present to execute audits and profiles. The most common scenario is your company's

external IP address which is used both when you are within your company premises and if you are connected to VPN.

CAUTION	If you are not within your company / client premises working remotely, then go ahead and first connect to VPN before executing the below step.

Click on “Get Current” button to automatically capture the external IP address of your domain. Then click “Add”.





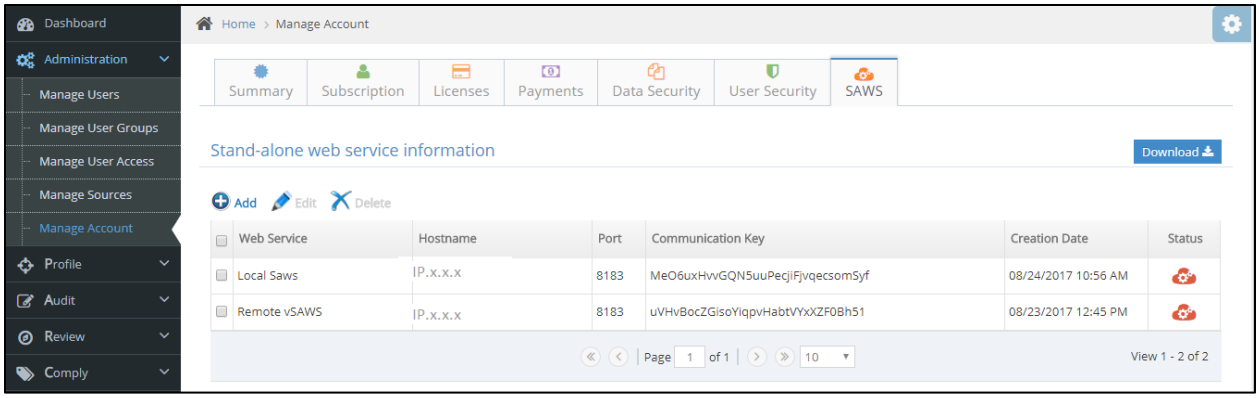
3 Web-service Configuration (SAWS as a Service)

This option deploys a web service that acts as the execution engine and can be used within the company premises or over VPN. It gives the flexibility of a single execution engine which all users can connect to. This option should be configured by someone who has been deemed as DvSum administrator. For a small team, this could be the owner role or a specified administrator user.

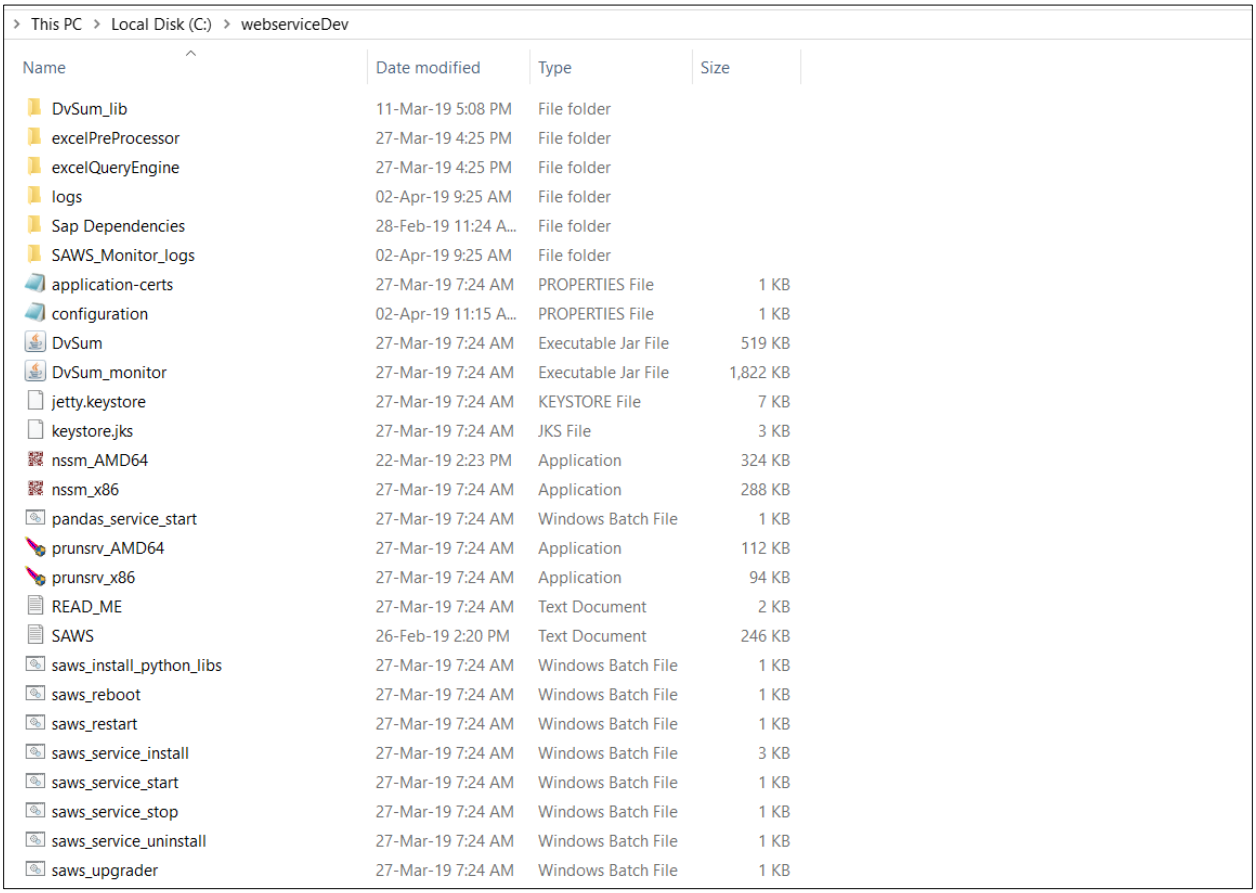
In a single-user or test mode, you can run web service on the PC. In a distributed environment, it is recommended that web service runs on a server.

Operating System	Windows XP, Windows 7, Windows 8.1, Windows Server 2008 and above.
Middleware	Java Runtime (JRE) 1.7_0_79 (for excel) or latest version
Disk Space	100 MB
Memory	1 GB

Step 1: You need to download the web-service archive. To do that, login to your account (as owner) and navigate to Accounts page >> SAWS tab.

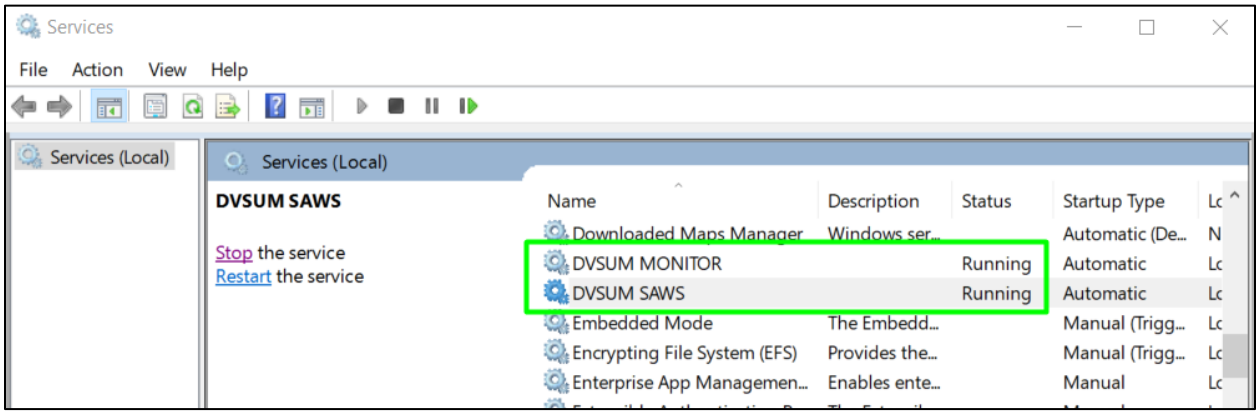


Download the SAWS zip file and copy it to the location (local or remote machine) you want your Web-service to run on. Unzip and extract the folder to a permanent location. (e.g. c:\dvsum or /opt/dvsum) . Below is how the extracted web-service folder should look like.



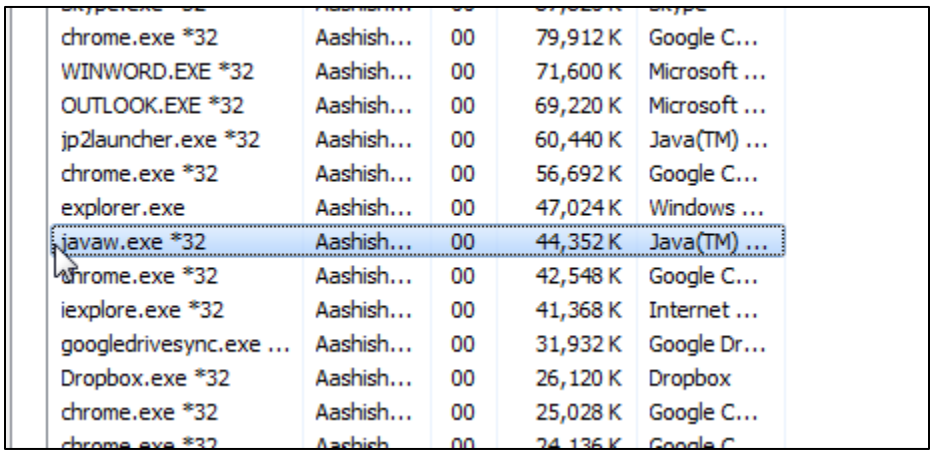
Step 2: You can start SAWS as a service by Right clicking on saws_service_install.bat and “Run as Administrator”. This will install and start webservice. To verify search windows for “Services”. This will show you services running on your local machine. Now search for “Dvsum”.

You will also see DVSUM MONITOR running. This monitors the status of SAWS. If by some rare reason SAWS crashes, the monitor will restart the web-service within 5 mins.

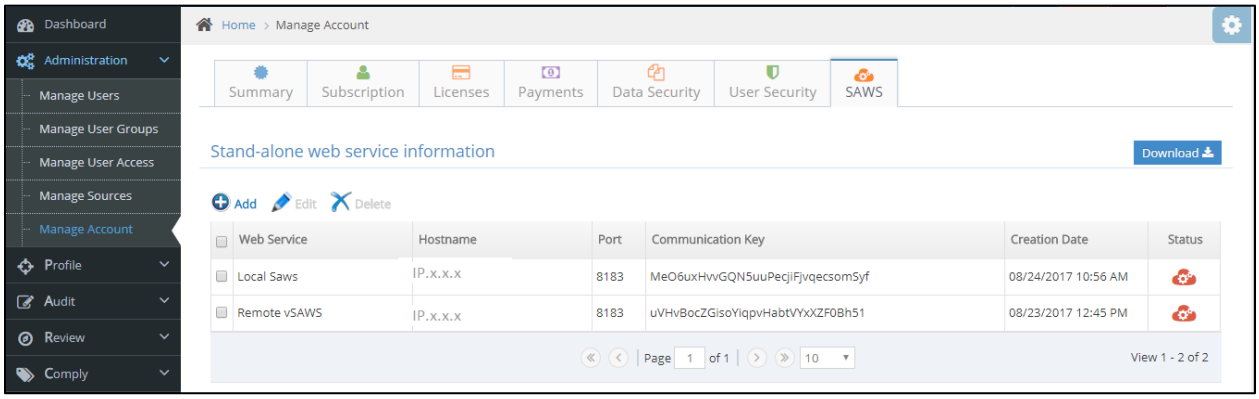


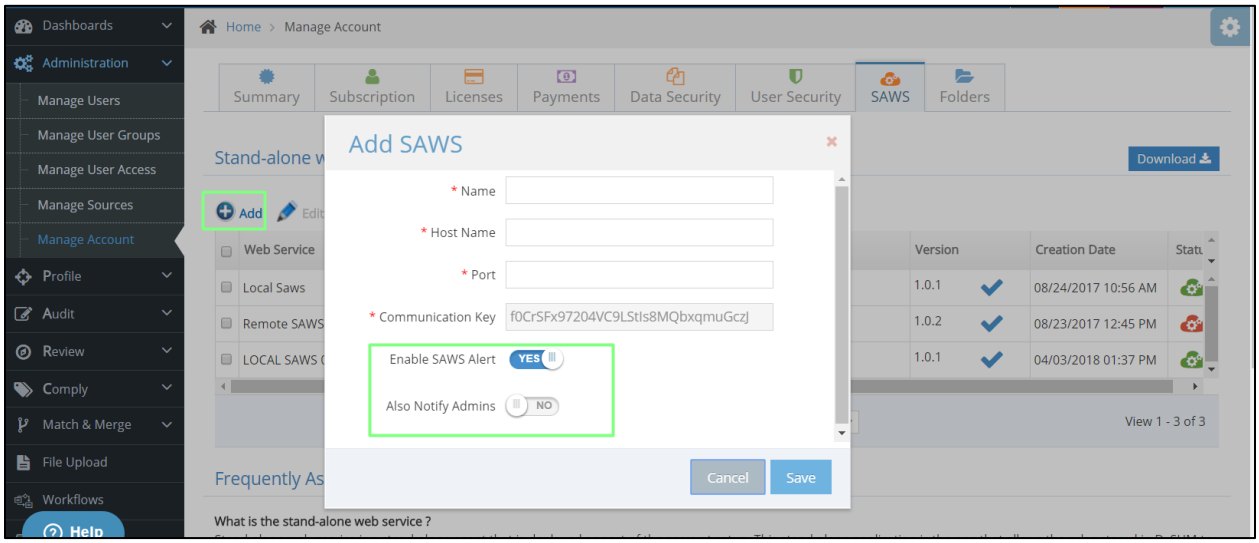
For detailed installation of SAWS including Python installation (for EXCEL source), please read our help article [SAWS 101: From Installation to Running as a Service]

Note: For manually starting the webservice, simple double-click DvSum.jar. It takes 5-10 seconds for it to get started. To verify open Task Manager and search for a process with name java.exe or javaw.exe



Step 3: Configure the SAWS you downloaded with the application by clicking on “Add” button.





Here you will enter the **Name, Host Name/IP** and **Port** for the configuration of web-service.

Note: You can also configure multiple SAWS with DvSum and you can map your Data Sources to the desired web-service.

SAWS Alert is the Capability to Send Email Notification to Account Owner and optionally Administrators when a specific SAWS is not able to communicate to DvSum web application and also when it resumes communication after failure.

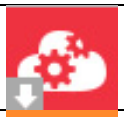
Enable SAWS Alert (Default is Yes)

Also Notify Admins (Default is No)

Note: If Enable SAWS Alert is set to NO, Application won't send alert Notification for this specific SAWS.

When the SAWS is successfully added and executing, the red cloud turns green. This cloud appears at the top right corner of DvSum application.

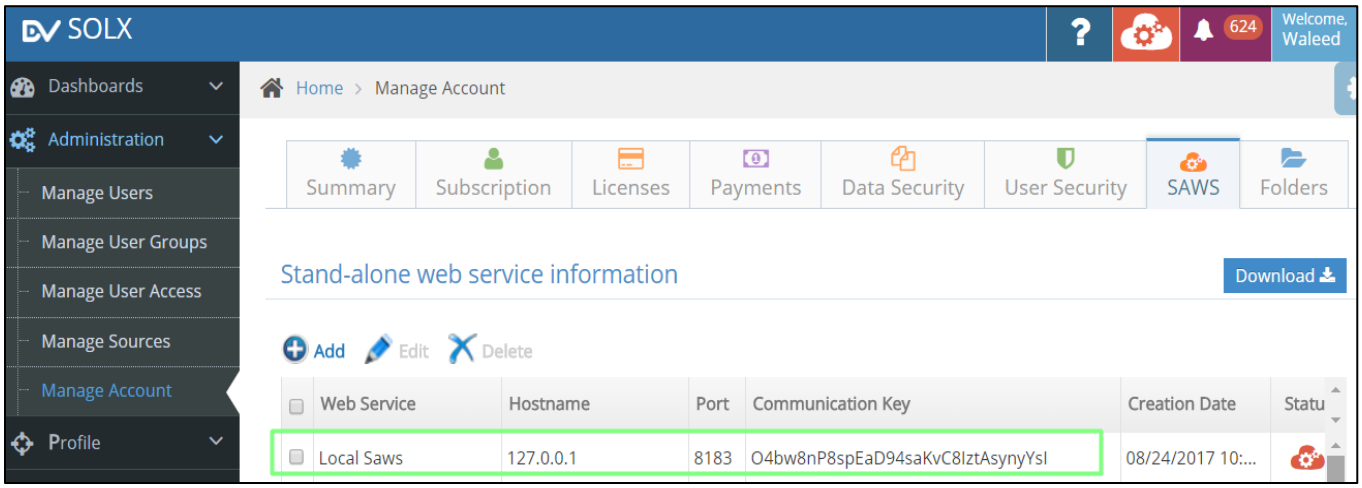
There are 3 states of this cloud;

RED: When all listed Web-services aren't running	
ORANGE: When some of the listed Web-services are running and some are not	
GREEN: When all of the listed Web-services are running	

At this point, SAWS cloud might be green but the process is not completed yet. We still need to define a communication key in the properties file of SAWS.

DvSum Communication key

Whenever you add SAWS, DvSum will generate a Communication Key which allows the communication of SAWS between DvSum Application and your company's Data Source.



Step 4: Copy the Communication Key. Go to Webservice folder and open **configuration.properties** file. Set this communication key as "api.token"

configuration.properties - Notepad

File Edit Format View Help

```
#Mon Jan 29 11:47:07 PKT 2018
logLevel=DEBUG
logFilePath=SAWS.log
scriptPath=C\:/temp/preprocess_csv.bat
port=8183
api.token=O4bw8nP8spEaD94saKvC8IztAsynyYsI
pool.size=10
maxDatabases=20
```

Read our article for [Advanced SAWS Settings (Configuration File)]

Step 5: Save and Close the file configurations.properties

This completes the installation and configuration.

Note – as a user you will still need to follow the remaining instructions for all-users in Chapter 2 to be able to use the service.

4 Identifying an account administrator

If you are going to use this application for individual needs, then you are the owner, administrator and the super-user. In which case, you can skip this section. However, most likely and eventually, you want more users to use this application. Determine if you are going to play the role of administrator or an existing system administrator needs to be assigned responsibility for DvSum.

The responsibilities of Administrator will involve – managing users into the system, managing data sources and executing batch jobs. Typically, this role would be best suited to existing administrators in your IT department for data intensive applications like ERP, Supply Chain etc. It is also absolutely ok for you as owner to play that role.

This completed the account owner specific configuration of the system. Please continue to the next section of **(Owner/Administrator Quick Reference)** to complete steps that are required for all new users logging into the system.

5 Manage Account

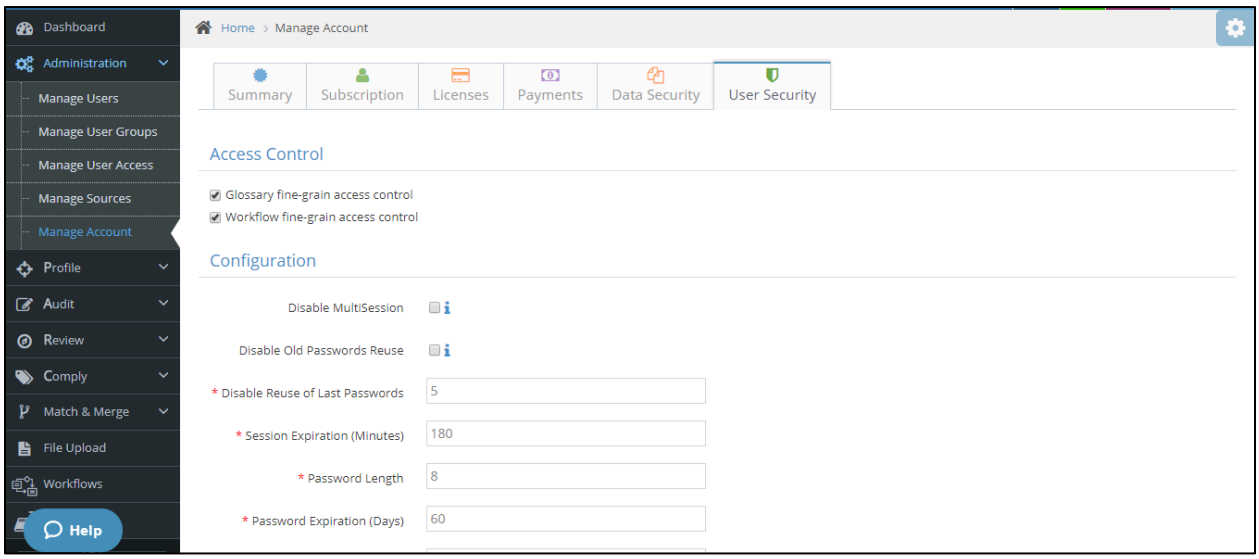
User Security tab is only visible to the Account owner. Here you can manage access control and account level security configurations.

5.1 Access Control

You can give Global or fine grain access control over the categories of Glossary and Workflow.

5.2 Configuration

Changing flags and setting values creates an impact on user login, user session maintenance, setting up password and user inactivity.

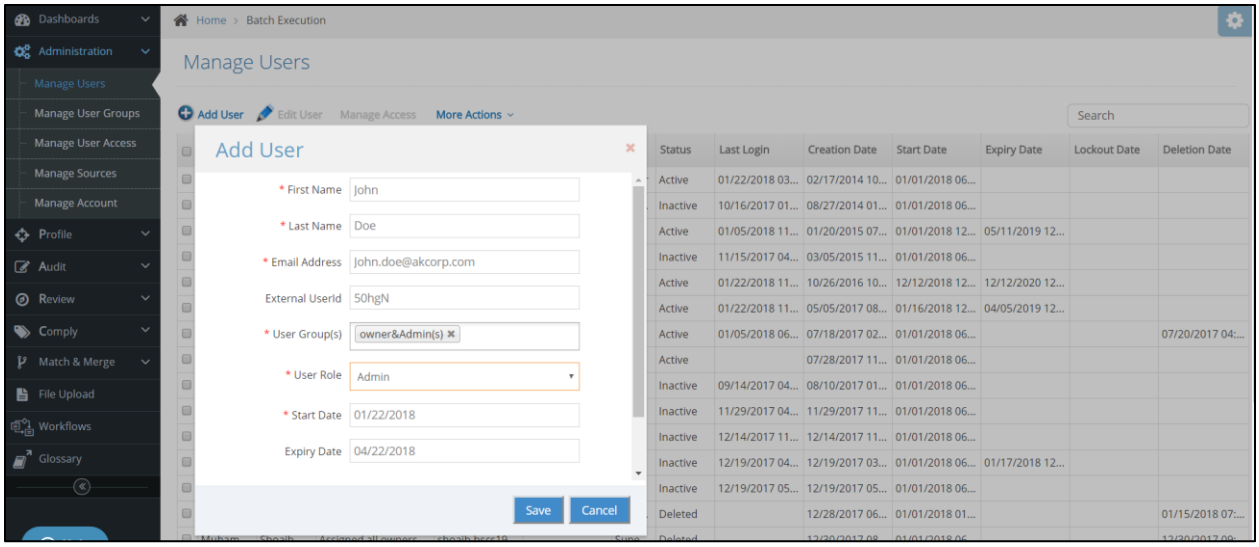


Applicability: Administrator and Owner roles

6 Manage Users

You add, delete and modify users using the Manage Users workflow. When adding a new user, you should select the role that is appropriate to the user.

Role	Description
Admin	Application Administrator. Responsible for managing the application, ensure its upkeep, manage sources, run batch workflows. In a small team, owner can double-up as Administrator or even super user.
Super User	Subject Matter Expert who is responsible for the data quality initiative or an experienced user who is a mentor to other users. Project Managers, Solution Architects or Business Line Managers should be given this role
User	A user who is responsible for maintaining certain business data within the enterprise. The user's role is focused on their own data



When you add a new user, you have the option to set the "Start" and "Expiry Date" for that user account on which a user has no control. This means, a user will not be authorized to sign into the DvSum application before the Start Date. Similarly, a user cannot sign in after he expiry date unless the Administrator resets it.

By Default, **Start Date** is the *current date* and **Expiry Date** is *current date + 90 days.* Both the fields are editable.

6.1 Manage User Account Status

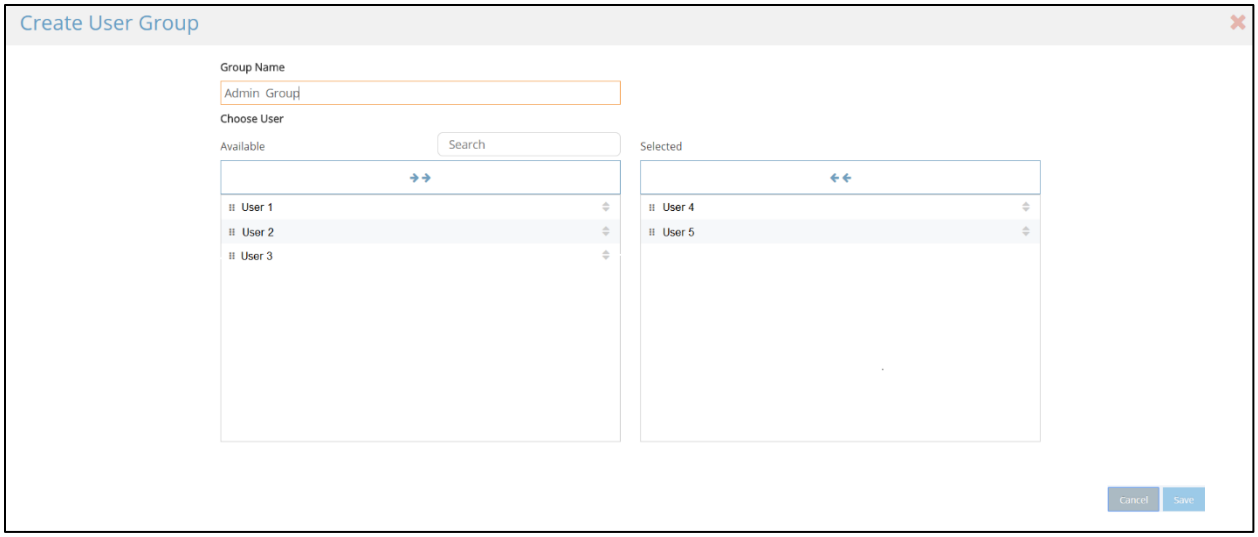
While users can only sign in the application within their Start and Expiry Date, you can also de-activate/activate them or delete their accounts by clicking on "More Actions". Keeping in mind that deletion does not permanently deletes a user. It will change their status to "Deleted" with their "Deletion date" populated in the user grid.

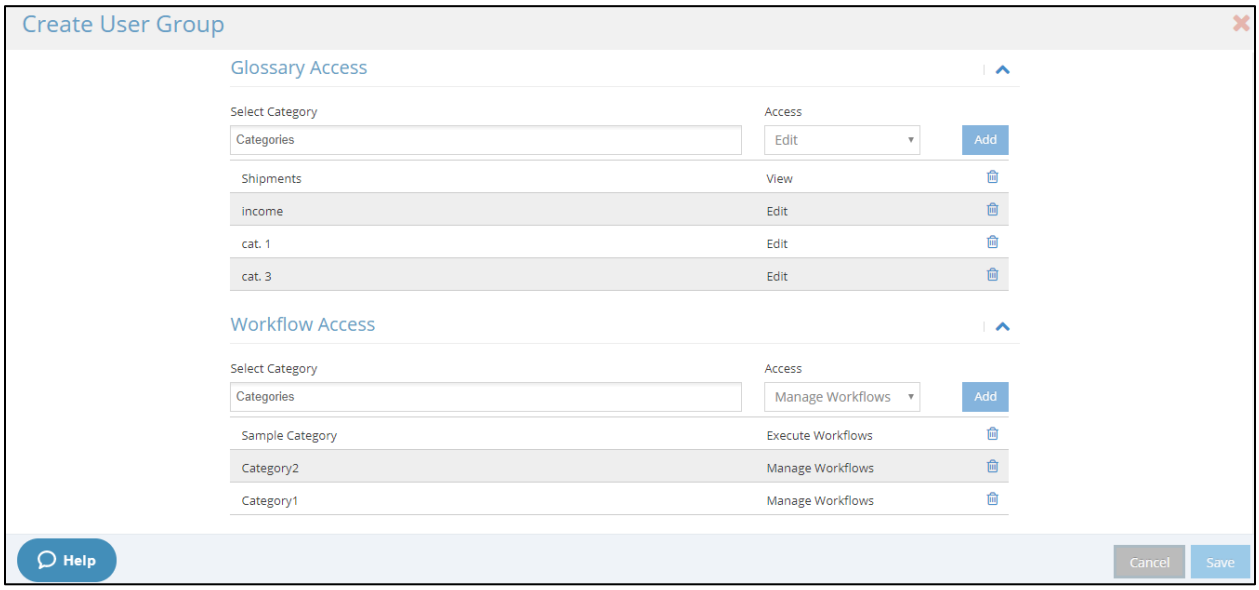
If you change the status of a deleted user to "Activate", it will treat the user as a newly added user.

7 Manage User Groups

Grouping multiple users together by specifying a group name helps the user to assign data quality or data management tasks to all the members in one go rather than selecting individually. Administrator can grant view and edit rights to

user groups on DvSum features like Glossary and Workflows. You can create a new user Group from the left navigation (Manage User Groups under the Administration tab)





After adding users, you can control access rights over multiple categories for Glossary and Workflows.

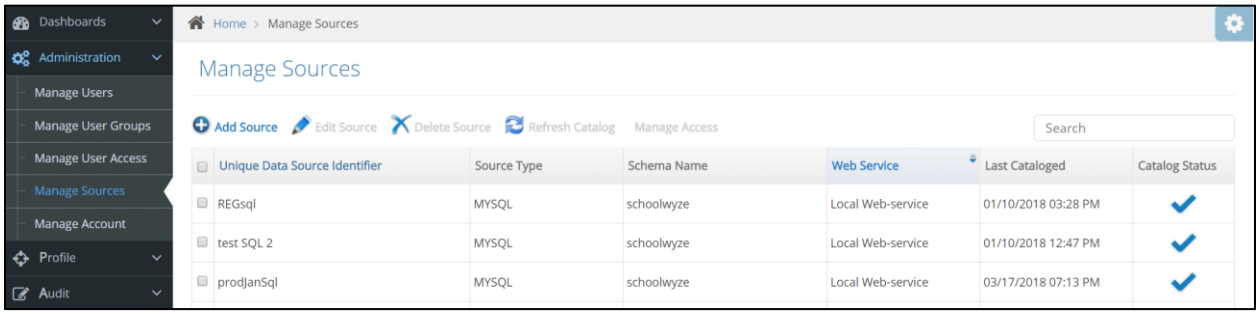
8 Manage Sources

This is a key workflow that defines the data source that will go through the data quality processes. Any new data source that needs to be setup has to be defined and cataloged first here. You can add the following types of data sources.

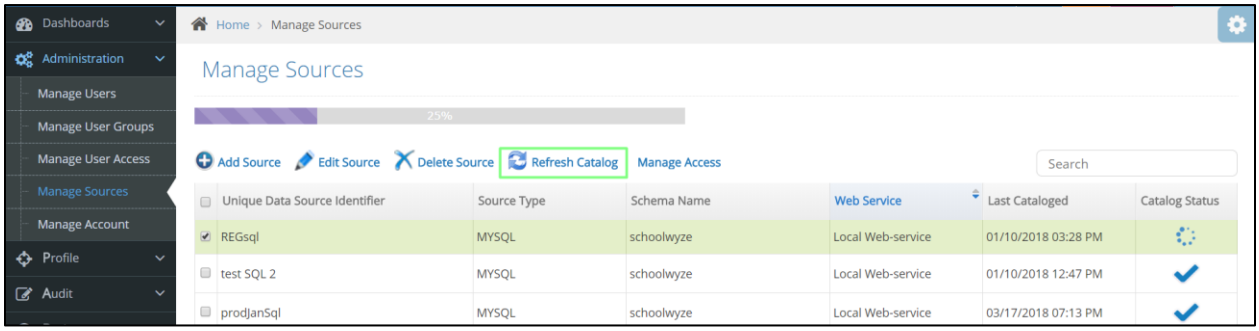
Data Source	Information Required
ORACLE, DB2, SQL SERVER, MYSQL	Host, Port, Instance, Schema, User, Password
Excel	Full path to the .xls or .xlsx file
SAP-ECC	Host, System ID, Client ID, User, Password, Source Application

8.1 MYSQL/Oracle Sources

While you are adding your Data sources, you need to map every source with a web-service as well.



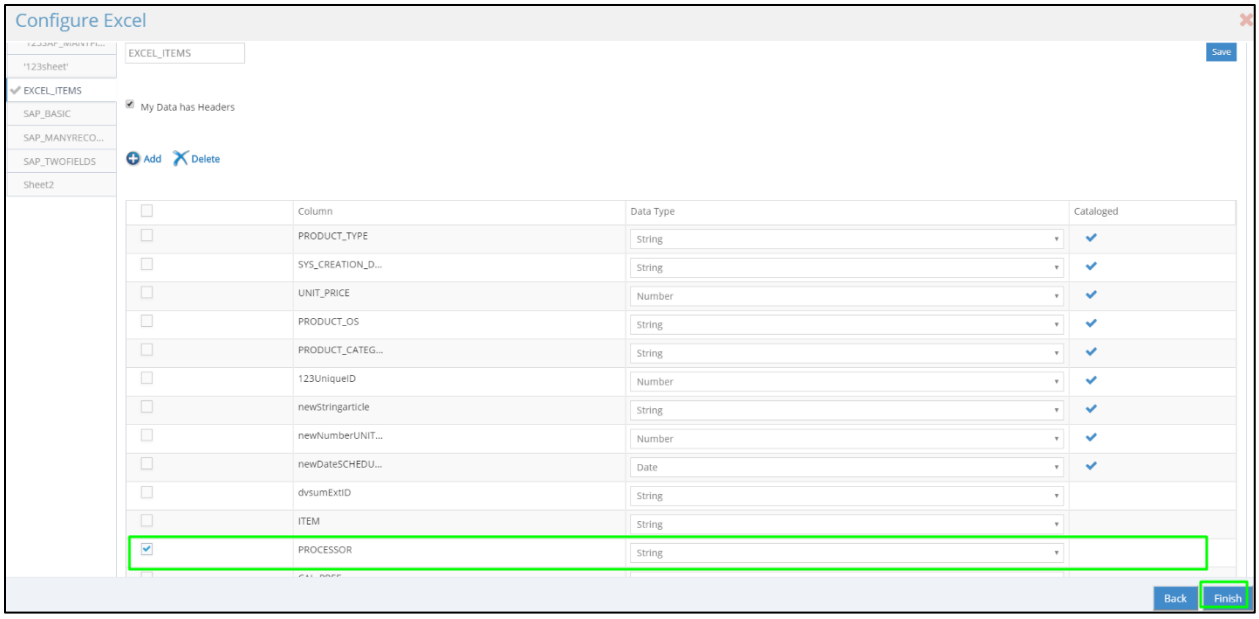
After adding any data source (except Excel), you would need to select the source and click on Refresh Catalog. Refresh Catalog is the process by which DataPARC goes and reads the meta data of the data source and saves it for use during profiling, data exploration, data auditing and in generating insights.



8.2 Excel/CSV Source

For Excel data source, you configure the data source by manually selecting the worksheets in the Excel workbook, the fields that they map to and field types. This results in creating or updating the catalog for that source within DataPARC.

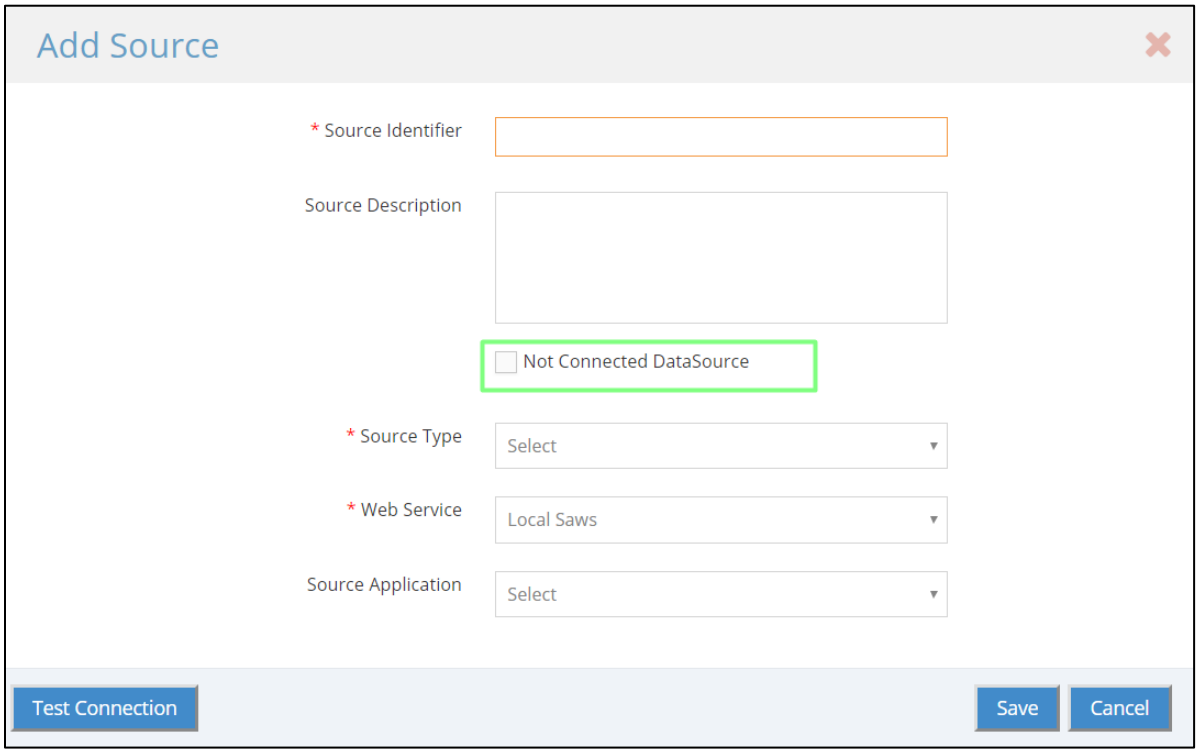
If an EXCEL or CSV source is already cataloged and profiled and a user wishes to add new columns, he can add, catalog and profile them individually and it will not have an impact on profiling of old columns or affect any rules created on source previously.



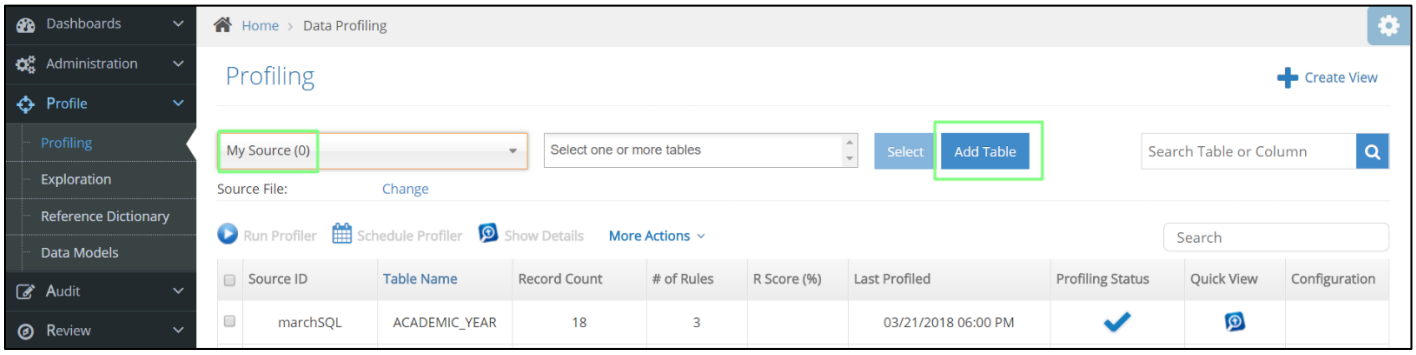
8.3 Not Connected Data Source

Not-connected sources are the ones which are not connected to the DvSum Web-service for the purposes of cataloging, profiling, executing rules or cleansing data.

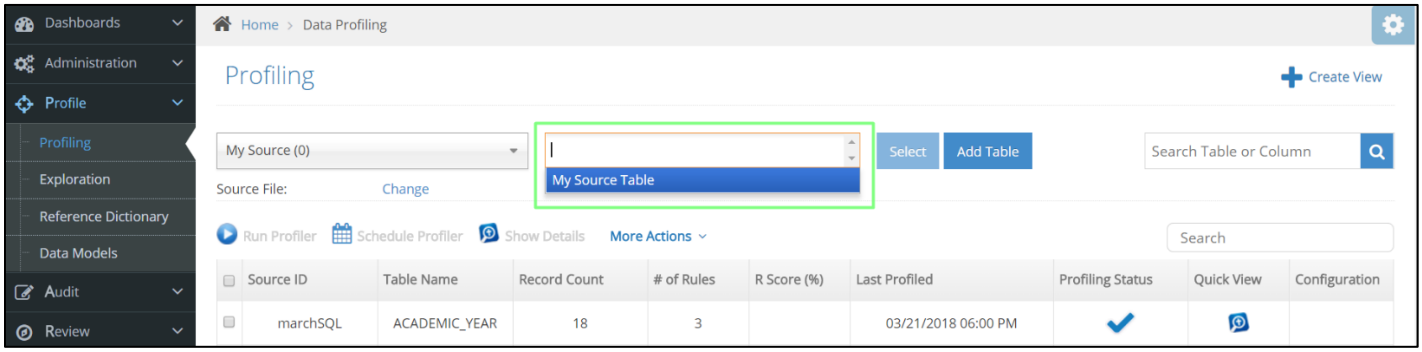
If you want to create a blueprint of a source or may require to create lineage, you can select this option. This means you would manually add the tables and the columns of your source.



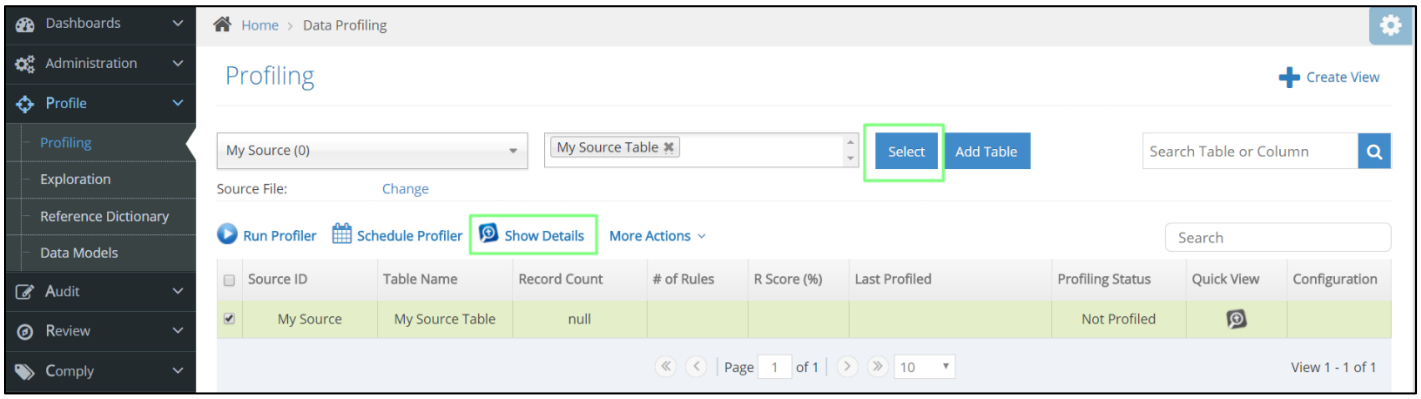
Once you have added the Not-Connected Source, you can go to Profile tab >> Profiling and select your source to define the tables manually. Only for a not-connected source you will see the "Add Table" button.



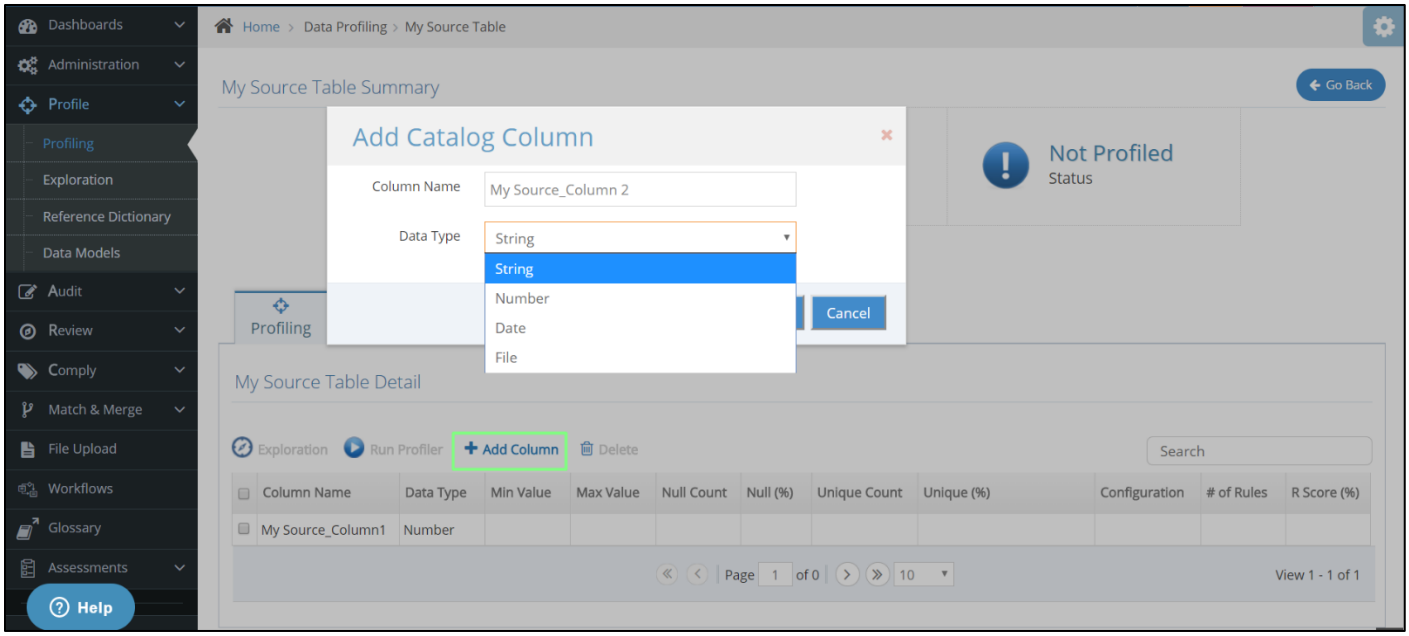
When you add the table, it starts appearing in the tables drop down.



Select this table and go to "show details" where you can define the columns manually.



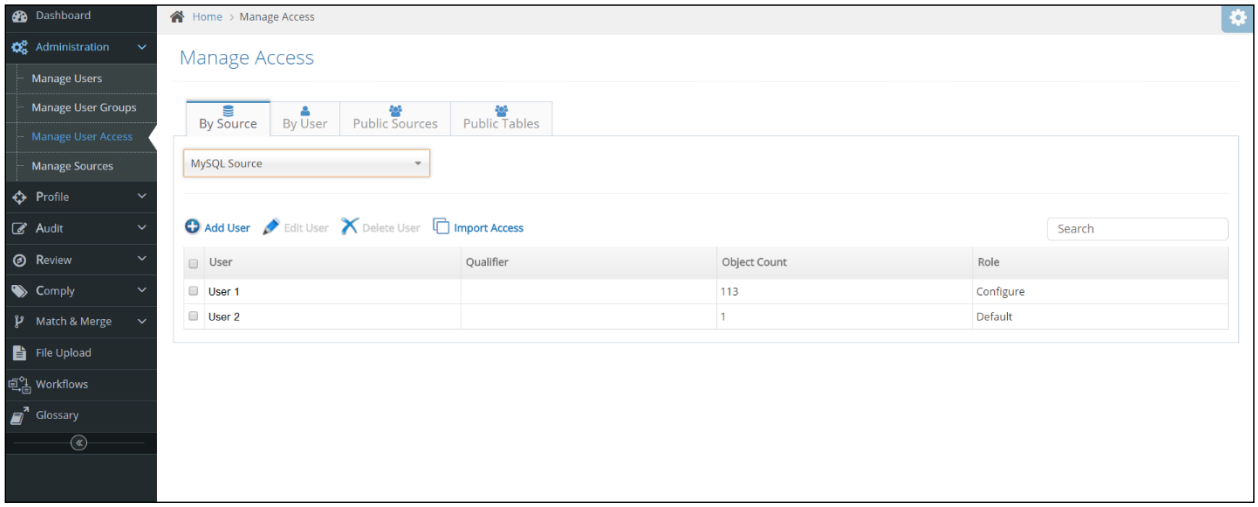
You can give name to your columns and select their data-type



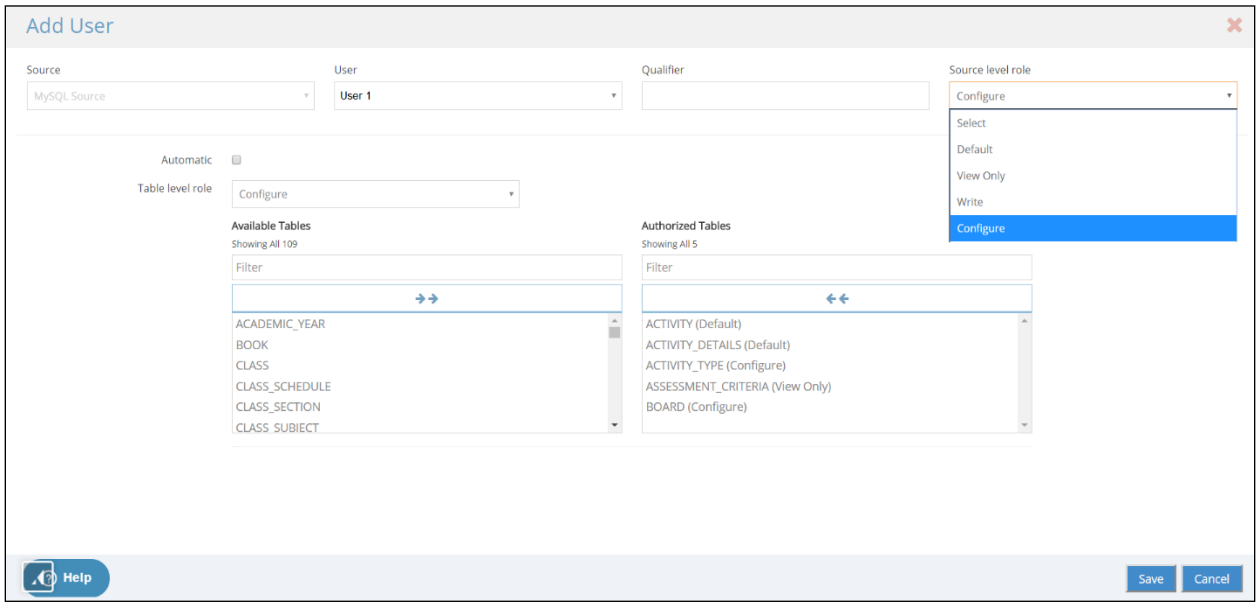
This completes the blueprint of your Database.

9 Manage User Access

An administrator can give access to user over a source or on particular tables of a source from this tab. Select a source from the dropdown menu and click on “Add User”.



Here you can select Source level role (Default, Configure, View only or Write) for the source and table level role for each table or for all at once.


Add User
Source
MySQL Source
User
User 1
Qualifier
Source level role
Configure
Select
Default
View Only
Write
Configure
Automatic
Table level role
Configure
Available Tables
Showing All 109
Filter
ACADEMIC_YEAR
BOOK
CLASS
CLASS_SCHEDULE
CLASS_SECTION
CLASS_SUBJECT
Authorized Tables
Showing All 5
Filter
ACTIVITY (Default)
ACTIVITY_DETAILS (Default)
ACTIVITY_TYPE (Configure)
ASSESSMENT_CRITERIA (View Only)
BOARD (Configure)
Help
Save
Cancel

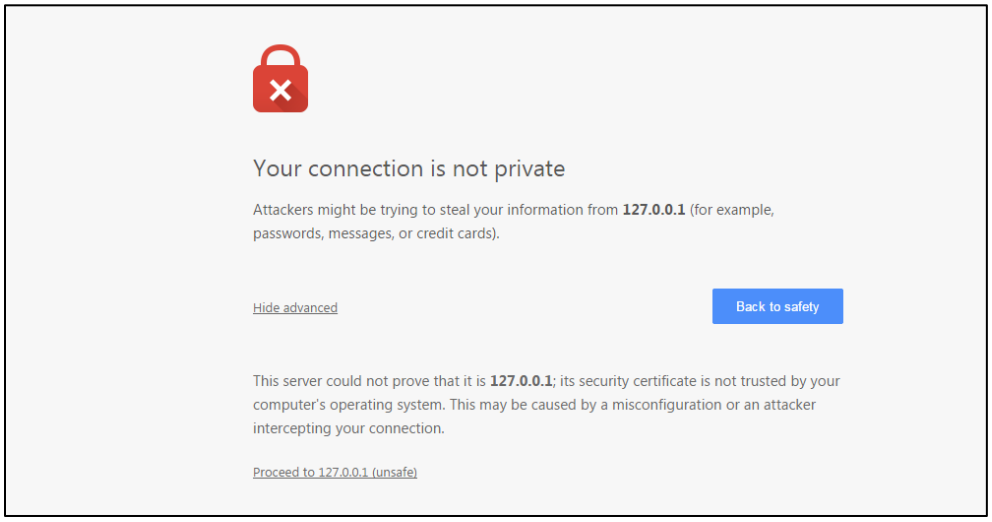
3 Configuration (Mandatory) for all users

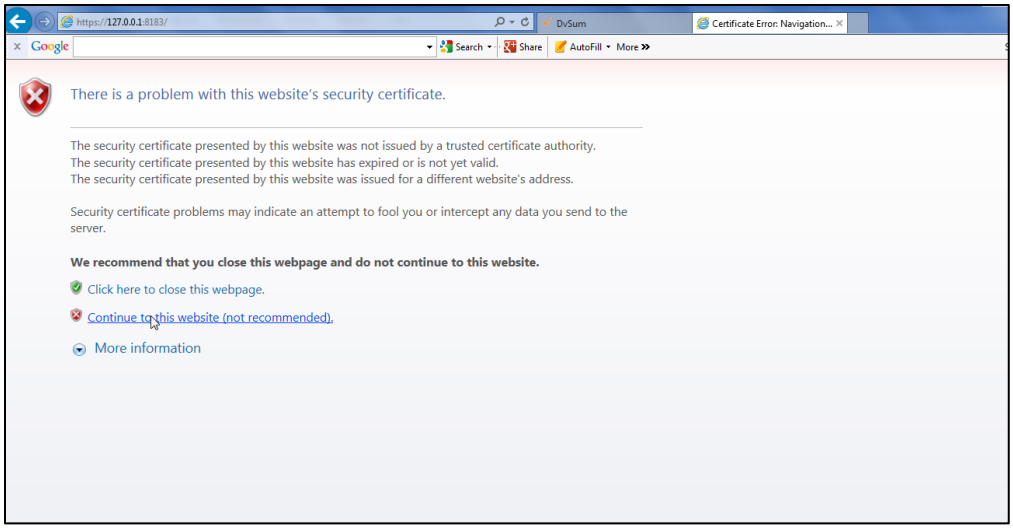
This section is required to be completed by all new users including Account Owner.

Applicability: All users

Login to the application. Look for the Cloud icon in “Red” or “Orange” color on the top right of the screen.

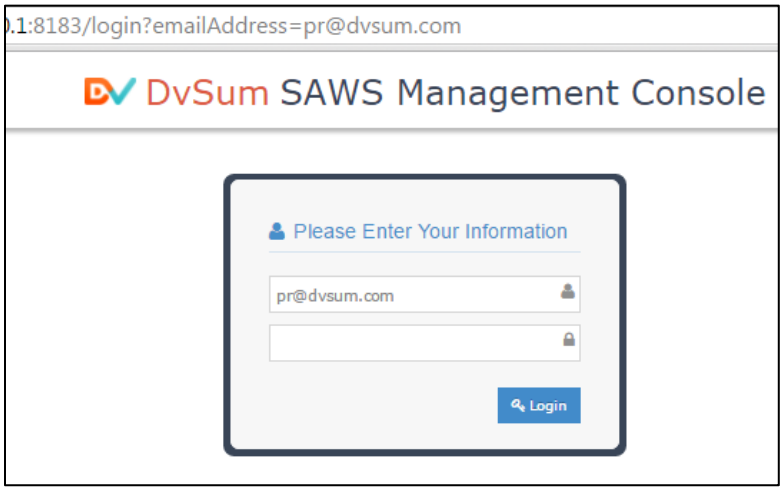
Click on the Cloud icon. It will open the list of web-services. Click on the web-service with the “Red” cloud. It will take you to the web-service page which will warn you of connection not being private. Click on “Continue” or “Proceed”. This is an internal company connection and it is fully secure.



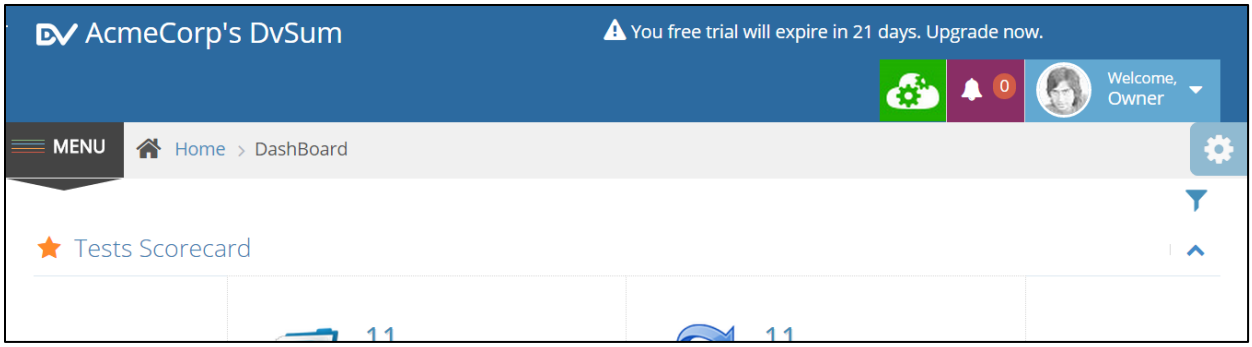


The reason this warning shows up because the SSL certificate generated by the stand-alone service is self-signed. This is not an issue because the communication that is happening is only between computers within the domain of your VPN connection or your company's subnet.

Once you click "Continue" or "Proceed", the next page should display the Management console as in picture below. This means that your browser can now connect to the web service over a secure connection. Go ahead and close this tab.



Now back on the application page, wait (less than 30 seconds) for the execution widget icon to turn green. Green color signifies successful connection to the web-service for execution purposes. Note that when all of the listed web-services are running, the cloud icon turns green. If any one of the listed isn't executing, you will see "Orange" cloud icon. But in the drop-down list, it will clearly show you the web-services that aren't running with corresponding red cloud.



Your configuration is complete!

If you run into issues, please contact DvSum support. DvSum Support information is available within the User-Interface and also in Welcome section of this document.

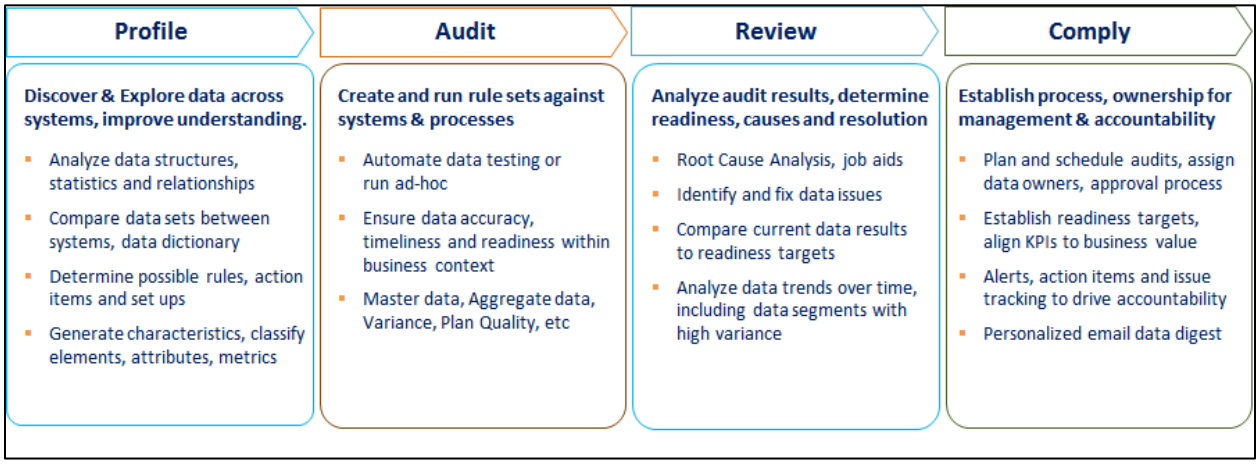
4 User Quick-Reference

This section is a quick reference that all users can use initially or on an on-going basis to understand what different workflow of DataPARC do and how to use them.

Applicability: All users

10 PARC– an integrated data quality improvement methodology

DataPARC stands for DvSum's data quality method of Data Profiling, Auditing, Review, and Compliance.



The workflows in the solution are aligned with the above principle.

11 Start with Data Exploration

Data Exploration is a workflow which allows you to understand what all data exists across your multiple data sources. It helps you visualize your data by showing it in a map format that you can click and find more information about that, identify missing relationships, profile the data, manage multiple audits for the same data element in the same place. The kind of use-cases that data exploration can assist with

- A business user has been asked to audit a particular data in the system for regulation requirements. They don't know which database, which schema that data exists.
- IT team is designing a new database. They want to find out where potential constraints and relationships (like Primary Keys, Foreign Keys) might be missing

- Data Governance – get information on areas where there is proliferation and duplication of data and integrity of data

Once your administrator has added data sources into the system, they are available to you for exploration. Alternatively, if you have just signed up, you can use the demo data sources that are automatically available within your account.

You can access Data Exploration from Left Navigation.

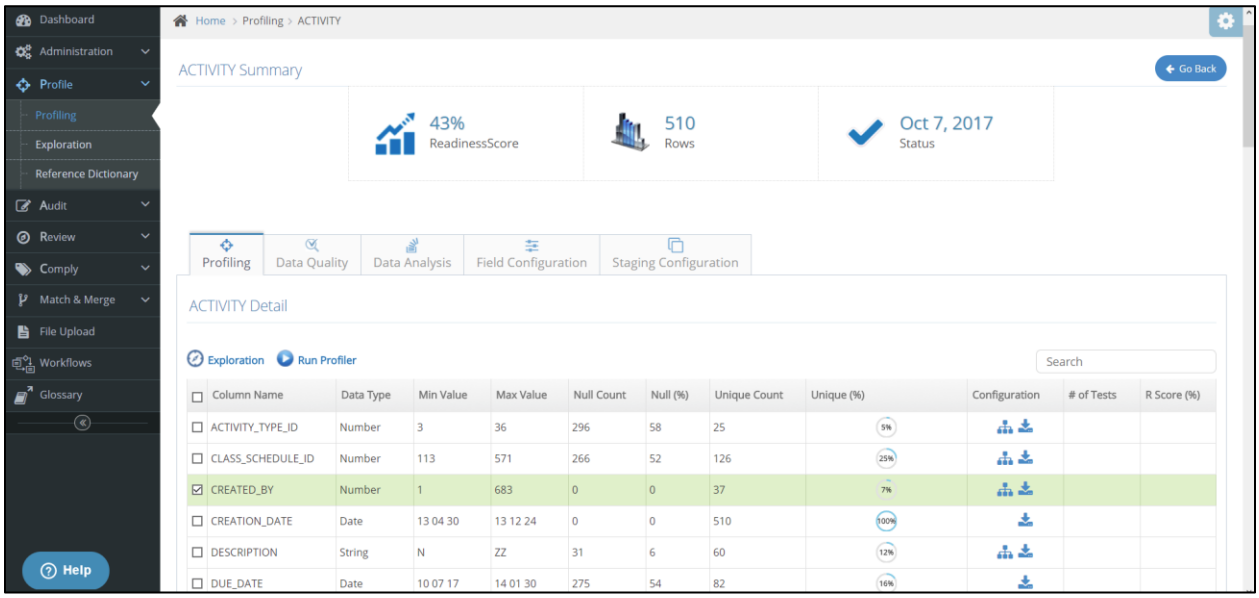
12 Data Profiling

Data Profiling is the workflow that allows you to better understand the characteristics of the data. It helps you answer the following questions

- Table level
  - How many records in the data table?
  - When was the data table last updated
  - What is the primary key in the data table?
- Column level
  - Sparseness / Density
  - Unique Values
  - Missing / Blank values
  - Minimum and Maximum values
  - Pattern of data
  - Distribution of data

Data Profiling is the first step in the PARC process of Profile. Audit. Review. Comply. You would typically do profiling and then set audits of the data based on evaluating the data profiles.

You can access Data Profiling from Left Navigation. Profiling information is also available when reviewing nodes in the Data Exploration section.



12.1 Catalog View

Data profiling provides you the basic information summary about a particular table of source which might not be enough for effective profiling and rule definition. Whereas, Catalog view generates customized information from multiple tables of sources. You can personalize these views depending on how you want to evaluate the data. That data is shown in the form of a table. You can then run profiler on that table just as on any other table of a source.

You can “Create View” from Profiling page.

From Profile tab in the left Navigation, go to profiling and click on "Create View". There are 3 types of catalog view.

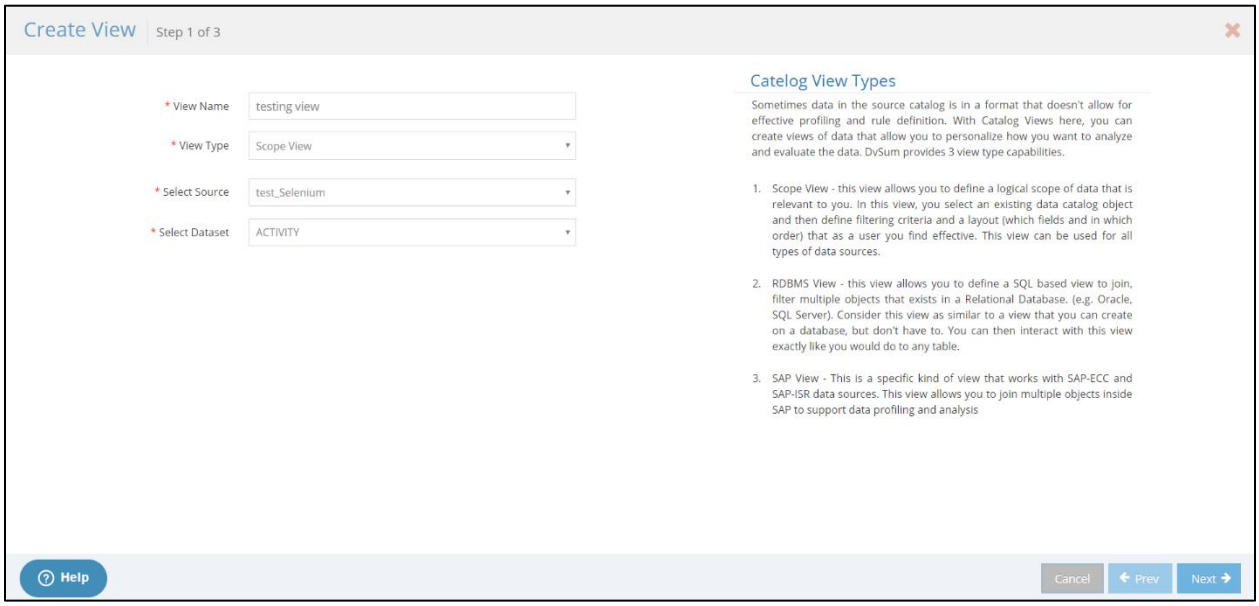
Scope View: This is the simplest view that you can create from selecting columns of a table from UI. This view is helpful when you want to see limited columns of a table or apply a particular filter condition on a column.

Blend View: Offers you to join two of more tables of a source with left/right or inner/outer joins. You can join tables from UI or provide join query.

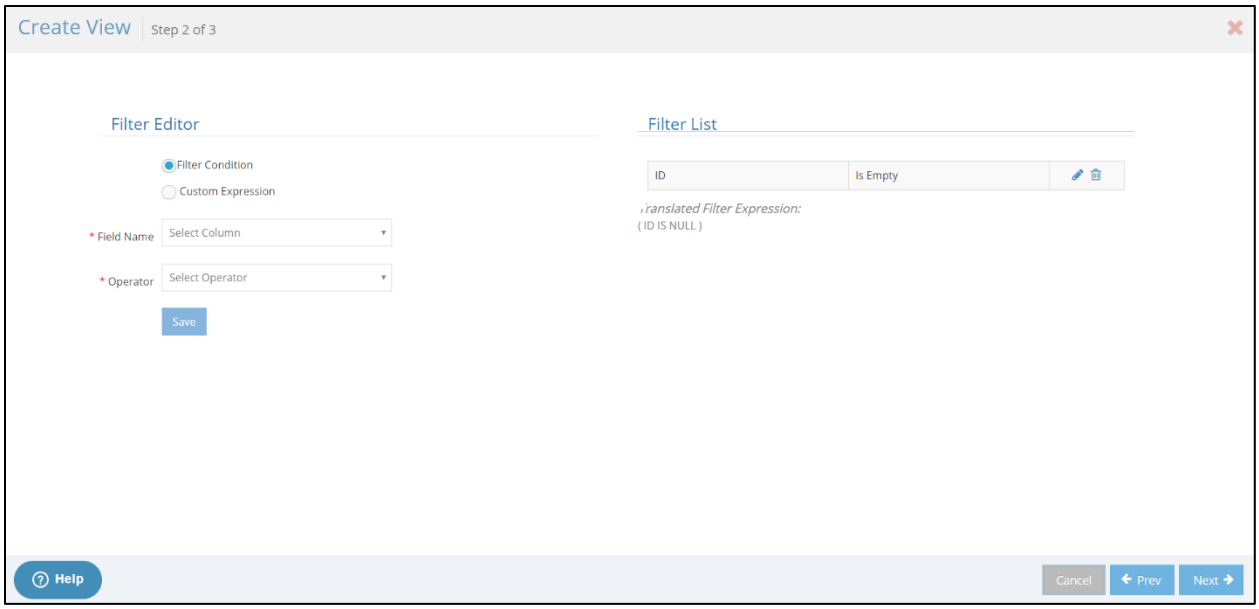
Database View: User needs to provide SQL query (and condition) to create db view.

To create Scope View follow along the steps below

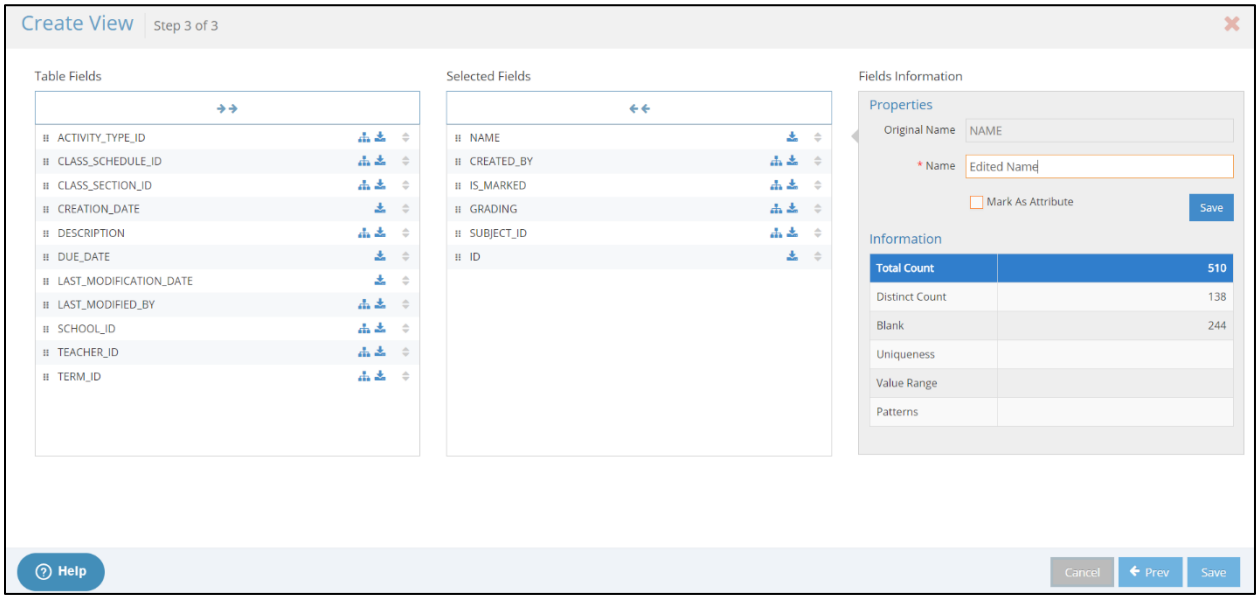
Step 1:
Enter View Name, type, source and table name.



Step 2:
Now you can apply Conditions to filter the data in the required form.



Step 3:
Now you can select the table fields of your source. Only these selected fields will appear as columns of catalog view table.



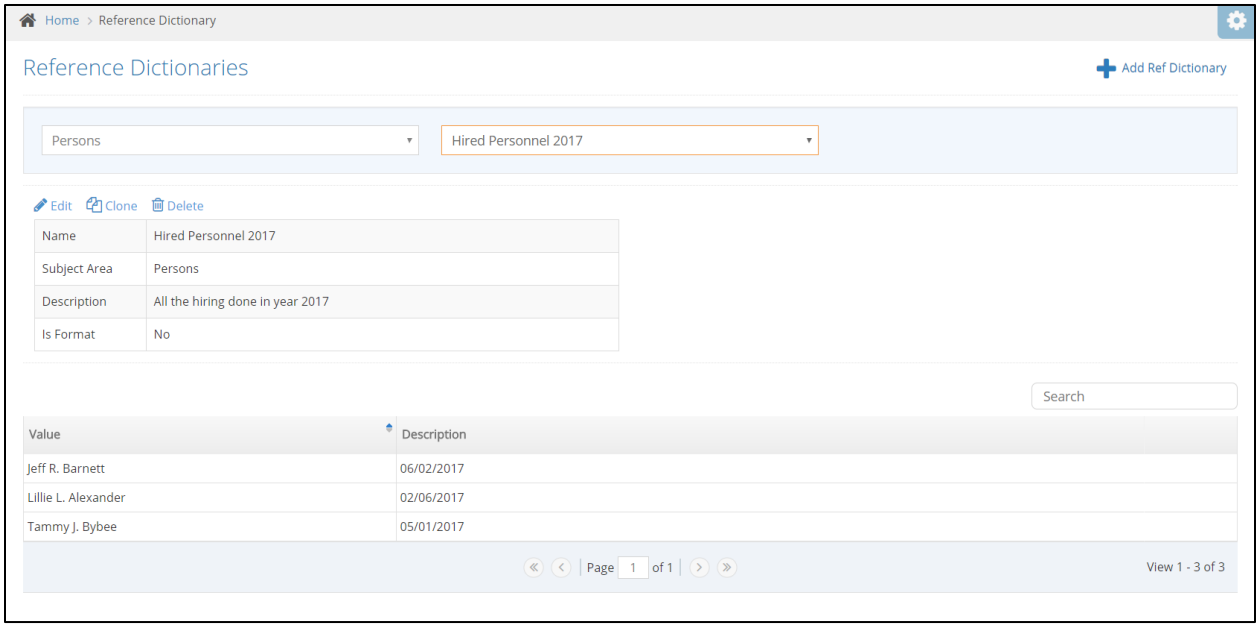
You can change the name of the columns as well from the Fields Information section. Now you can "Save" the view.
This will take you to the View detail page where you can profile or create rules on this view.

You can create blend view and database view in a similar way or you can read our support article on [how to create Blend View with join Condition.]

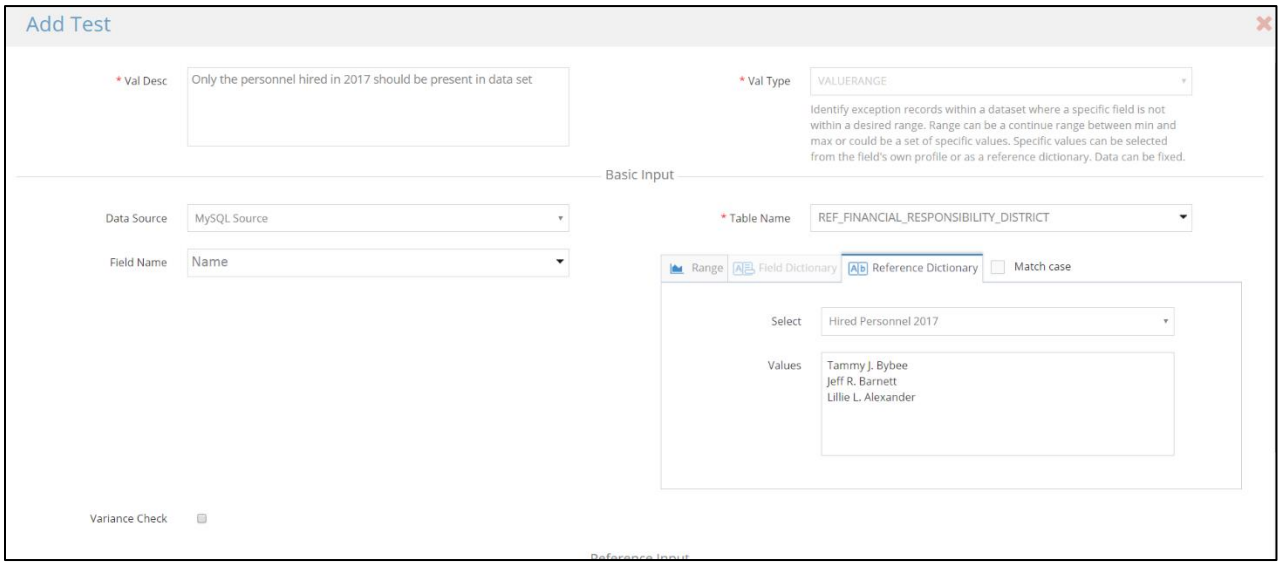
12.2 Reference Dictionary

Reference Dictionary is like a reference book containing values under predefined Subject Area. You can create your own dictionary by associating useful, organized information with it (i.e. set of values and their description). It is then helpful when DvSum validation Audits are created on the data sources. Instead of mentioning multiple data values, you can select a reference dictionary containing the list of data values.

Let's say, a user creates a reference dictionary "Hired Personnel 2017" which contains names and hiring date (as value description) of employees that joined their organization in the year 2017.



Now a "Value Range" audit is being created on a data source where you want to make sure only hired personnel exist in a particular column. Instead of mentioning the range, you can select the reference dictionary that you created.



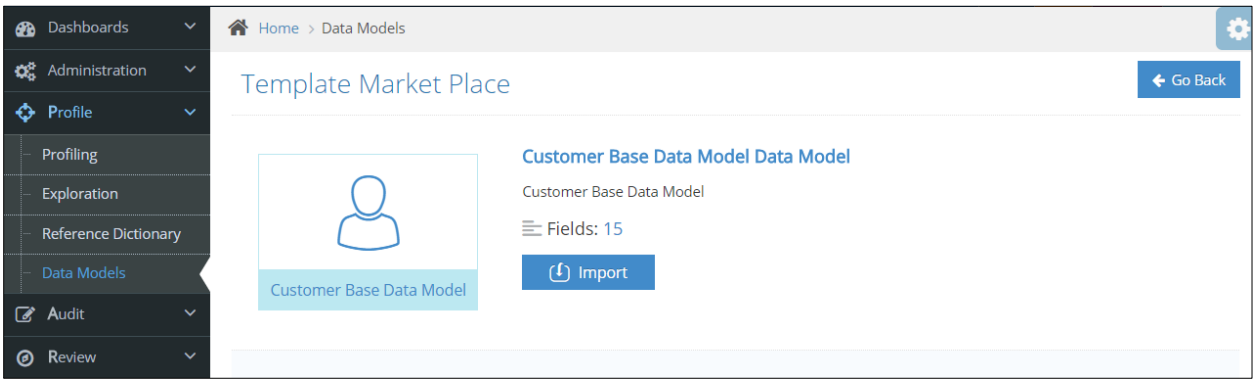
Now when you run the audit, it will validate your data source against these predefined values in this reference dictionary.

12.3 Data Models

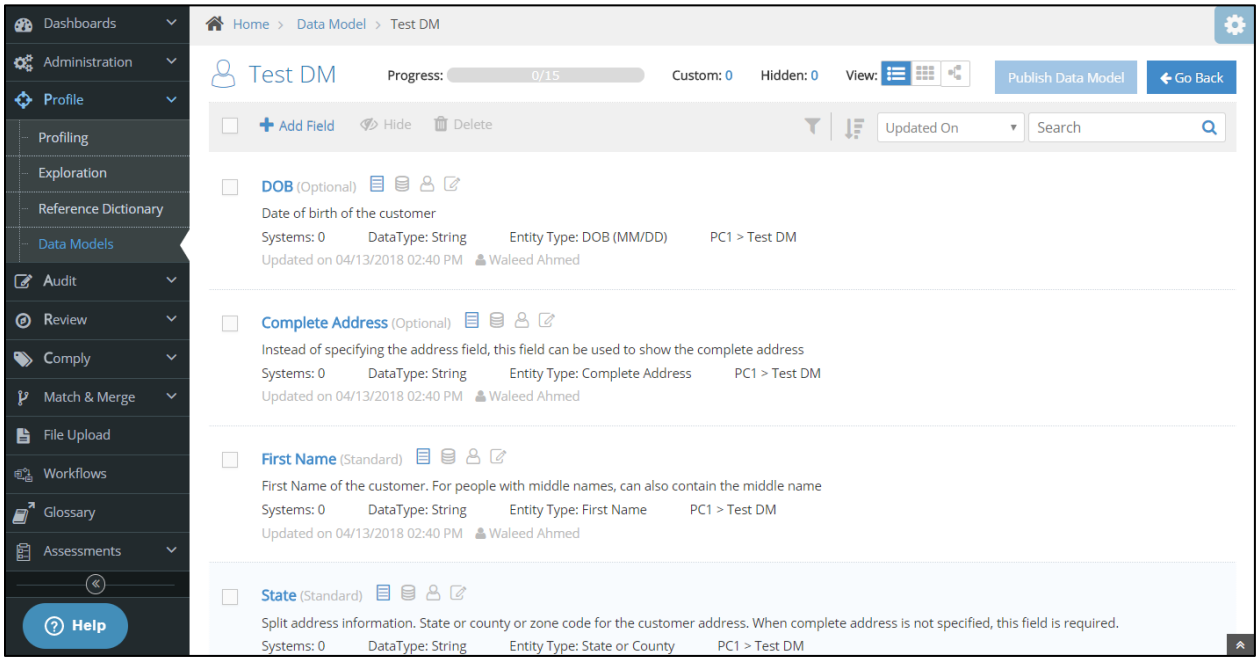
Data modeling is often the first step in database design to first create a conceptual model of how data items are related to each other. A data model template consists of fields defined as Standard,

Required and Optional. You can edit the fields as required to define Assets, User Roles, Data Policies or associate Workflows with the fields. You can also create Lineage on assets.

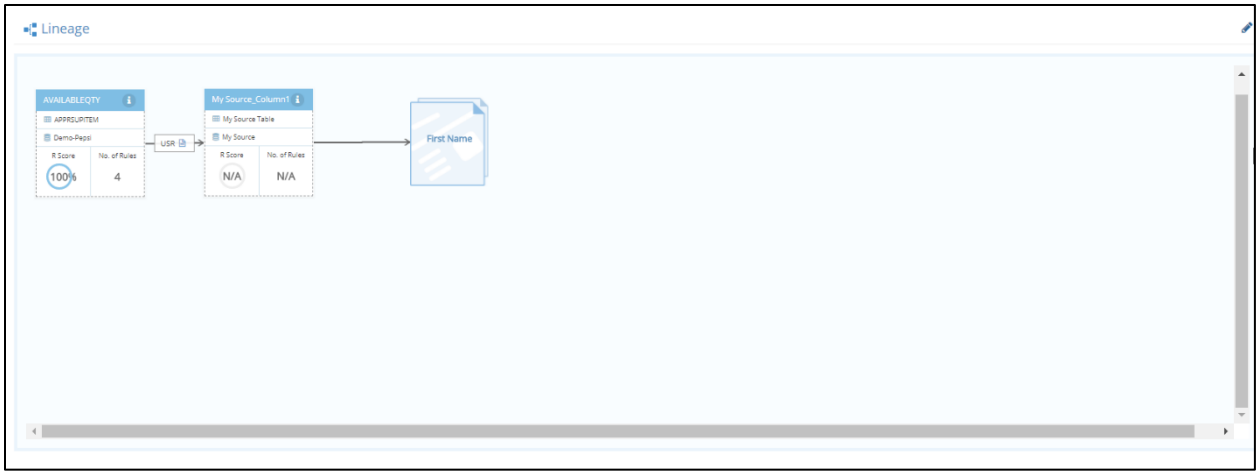
From the Template Market place, import a template and give it a title and description.



After importing the Data model, you can start working with the imported fields. Also, you can add Custom fields or associate Terms from Business Glossary to a Data Model.

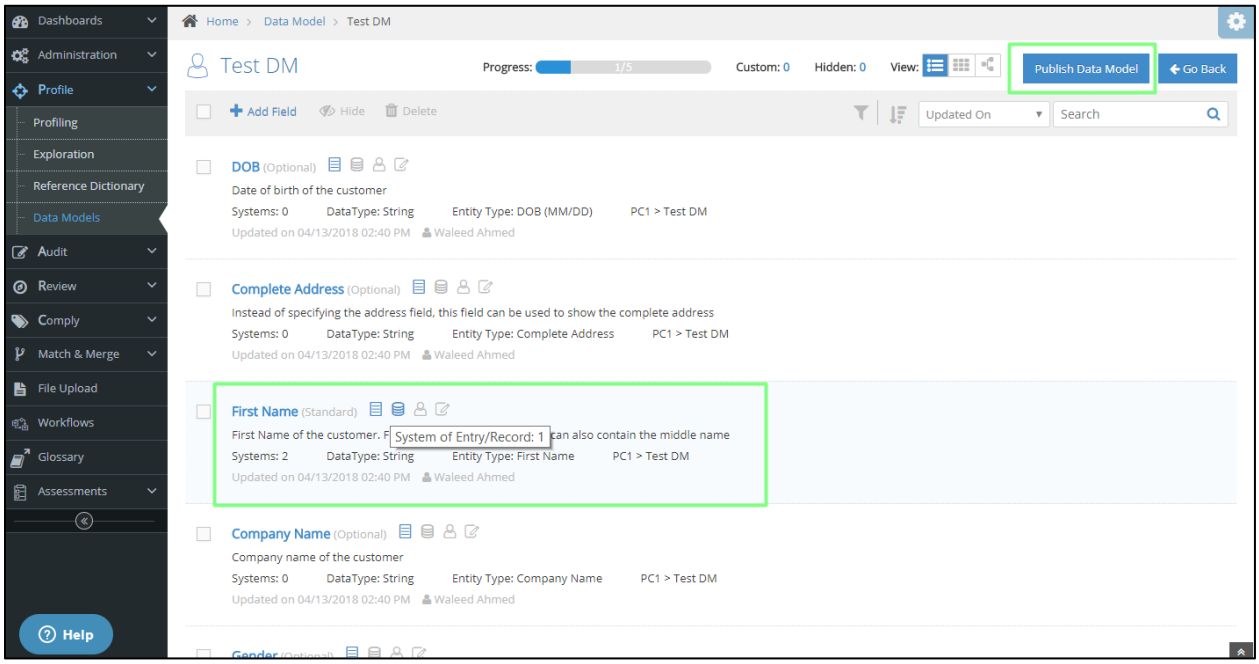


You can go to field detail page by clicking on the field name, where you can edit all sections Under the "Assigned Assets" section, you can define System of References and System of Entry which add the asset icon in lineage designer window. You can connect these assets to create lineage (like done from Glossary)



12.3.1 Publish Data Model

Standard/Required fields need to be completed before the Publish Data Model button is enabled. For a field to be complete it needs to have 1 System of Entry/Record with Api to delete.



The status of the fields is shown with 4 icons in front which show functional description, Systems, Field Owner and Data policies of field. Blue color appears if an entity is defined, otherwise its gray. The Progress bar on top will be showing you the completed fields out of the total number of fields.

Once you have published the Data model, you can evaluate it from the Assessments Section.

13 Audit

Auditing is the workflow that helps you validate the quality of data. With DvSum, you define a data quality audit rule once and the system can run the same one more multiple times, determine whether the data is passing the quality check or failing and if it is failing, how bad is it and what are the exceptions that require resolution.

The different categories of audits available are:

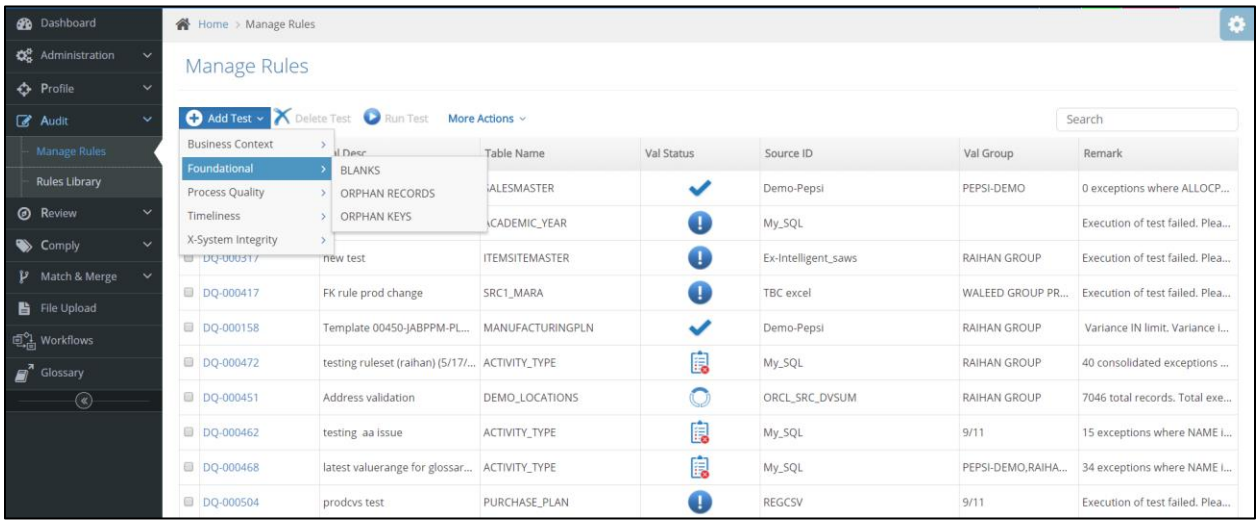
- Business Context
- Foundational
- Process Quality
- Volume Check
- X-System Integrity

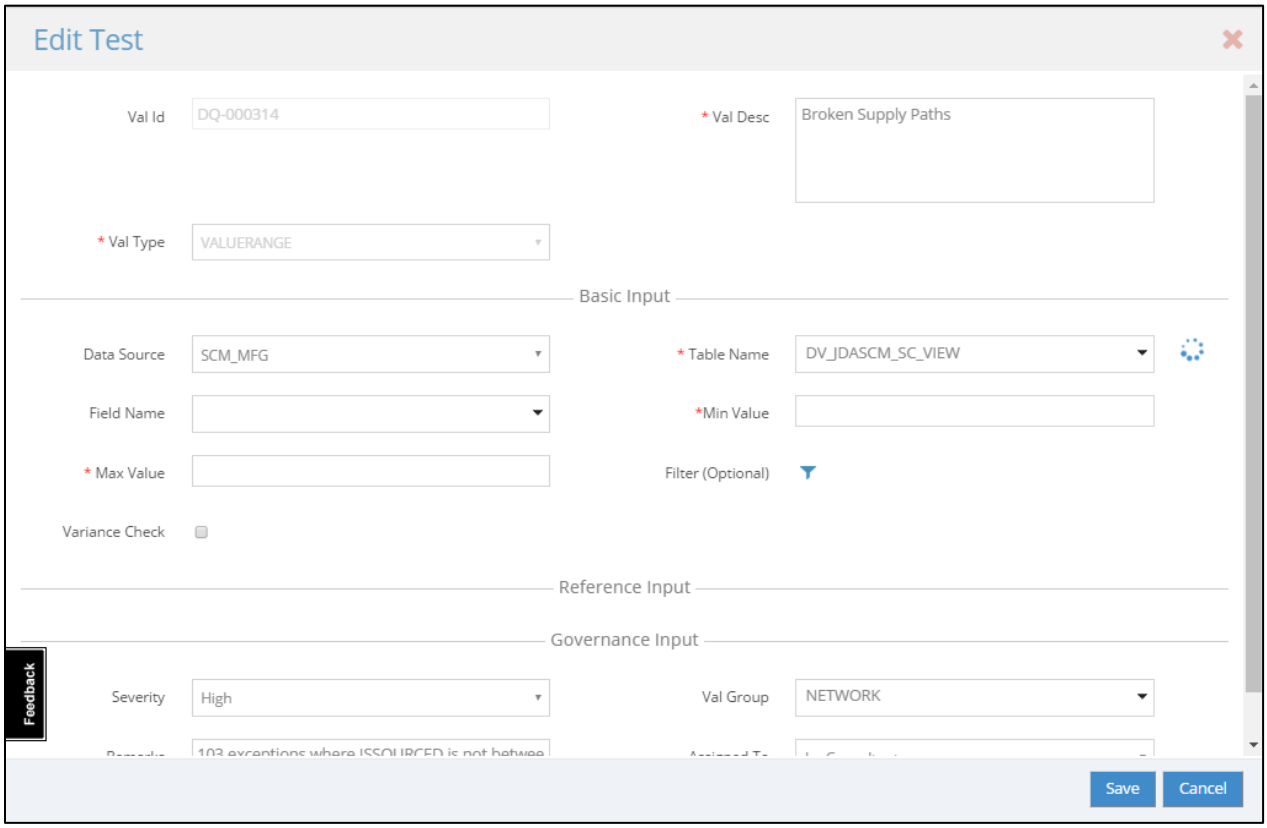
The different audit rules that you can configure with DataPARC are:

Type of Audit Rule	Definition	Example
BLANKS	Availability and completeness of data. Are their missing values?	All the records where address field is missing in customer ship-to information
VALUE RANGE	Are the values in data within the expected range?	E.g. manufacturing yield should always be between 0.01 and 1 Order_type in sales orders should only be one of ZOST, ZOCO, ZOFR, ZOSA Item_category in shipments should only be one of values in the item_category reference
UNIQUENESS	Are the values for a data element unique in the dataset?	Duplicate customers in customer table
COUNT	Is the count of records within the expected range?	Count of sales orders in open order extract should be between 800k and 1M records
SUM	Is the aggregated quantity within the expected range?	Total volume of open order quantity should be between 75M and 100M units
VARIANCE	Is the run to run variance of count or sum within the tolerance limit?	Total shipment plan from run to run should not vary by more than 1% Purchase plan sent to each supplier every month should not vary by more than 3% or 10K units
COMPARE COUNT	Compare the count of records between 2 different data sources at same or different granularity	Compare total shipments in ERP at transactional level to shipment summary loaded in data warehouse at aggregated level
COMPARE SUM	Compare aggregated quantity between 2 different data sources	Compare the total shipment volume for last 3 months in ERP with shipment

		volume loaded in data warehouse at aggregated level – by product line
COMPARE TABLE	Holistic comparison of count of records and volume for metric fields across different attributes	Compare Open Sales Orders in SAP with Sales Order extract in JDA for count, total of orderquantity, qtyopen by order_type, item_category, plant, key customer accounts, product line
CUSTOM RULE	Create your own custom validation rule	
ADDRESS VALIDATION	Verifies Address completion and validation using Google APIs	Validate the address of Cust_order_location to make sure the goods are delivered at the right location.
DOC - MATCHING	Removes the inconsistencies in the data among 2 or 3 different sources.	Reconcile Sales Orders with Delivered Orders for Cust_Name, total of orderquantity, order_type, item_category, payment.
DATA FORMAT RULE	Identify exception records within a dataset where a specific column does not match with required data format. (String, integer, decimal, whole number)	Sales Order_ID can only have column width of max 10 characters. Quantity should be whole number. Weight can be deciman number with maximum precision of 2.

You can create rules in DataPARC using the Manage Tests workflow. Some audit rules can be directly enabled from the Data Exploration workflow.



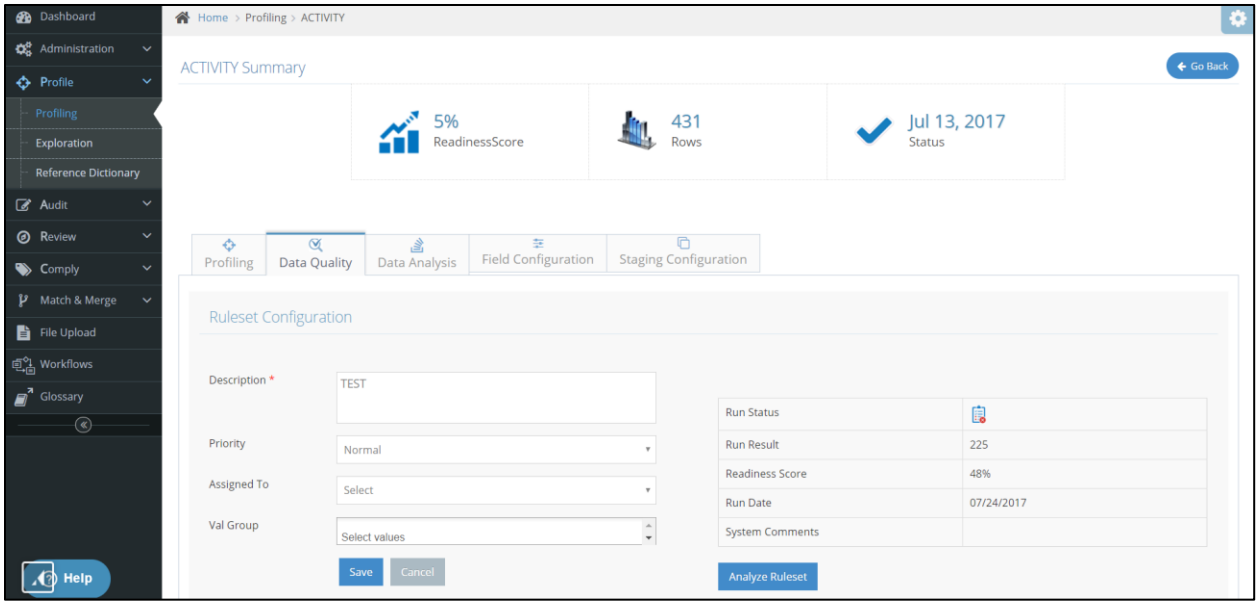


13.1 Executing audits

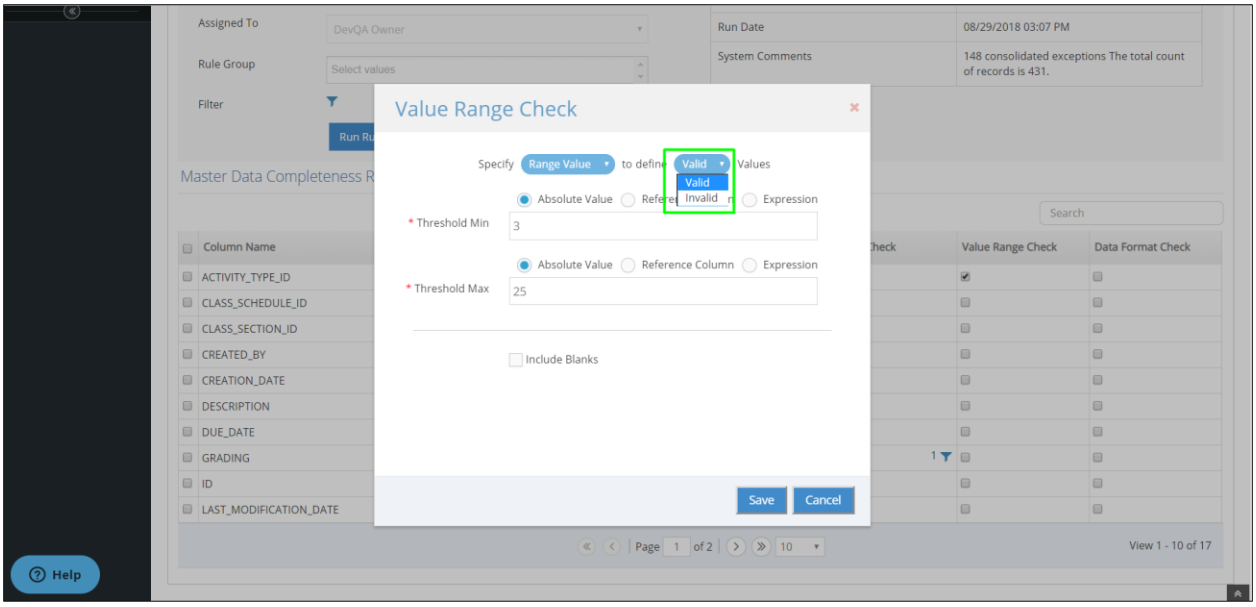
Once audits are created, they can be executed in multiple ways.

13.1.1 From Profiling Data Quality tab (MDC Ruleset)

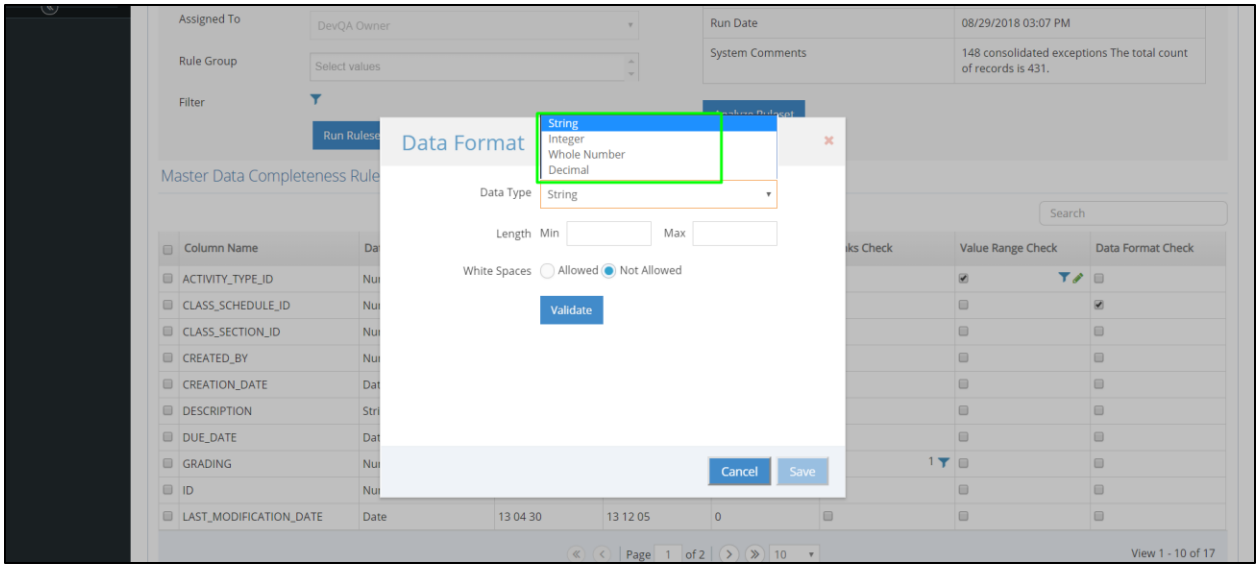
From Profiling tab, select a Source Table and click on “Show Details”. Click on Data Quality tab and enable ruleset for the selected Source Table. Enter the description for it and save it.

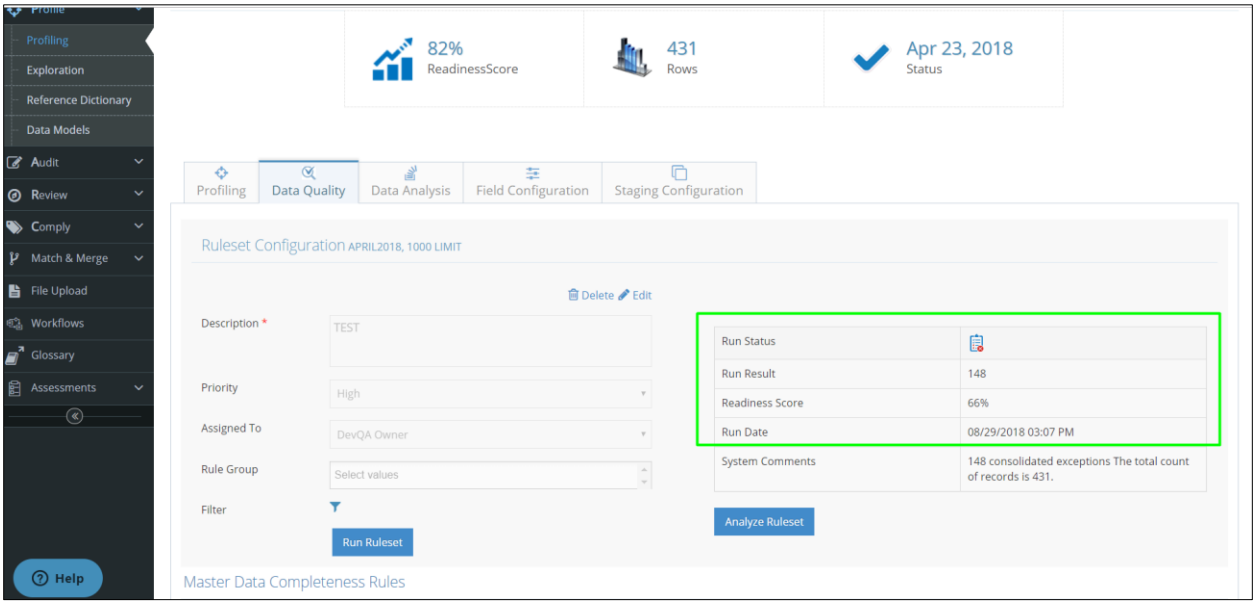


This creates the ruleset and now the columns of selected table are visible in the Master Data Completeness Rules Grid. Select the columns on which you want to apply Blank check, Value Range and/or Data Format check. For the Value Range check, you will have to mention minimum and maximum Threshold.



Similarly, for Data Format Rule, specify criteria which needs to be fulfilled. For String columns, you can specify column width and for integer, whole numbers or decimals you can create rule on single or multiple columns.

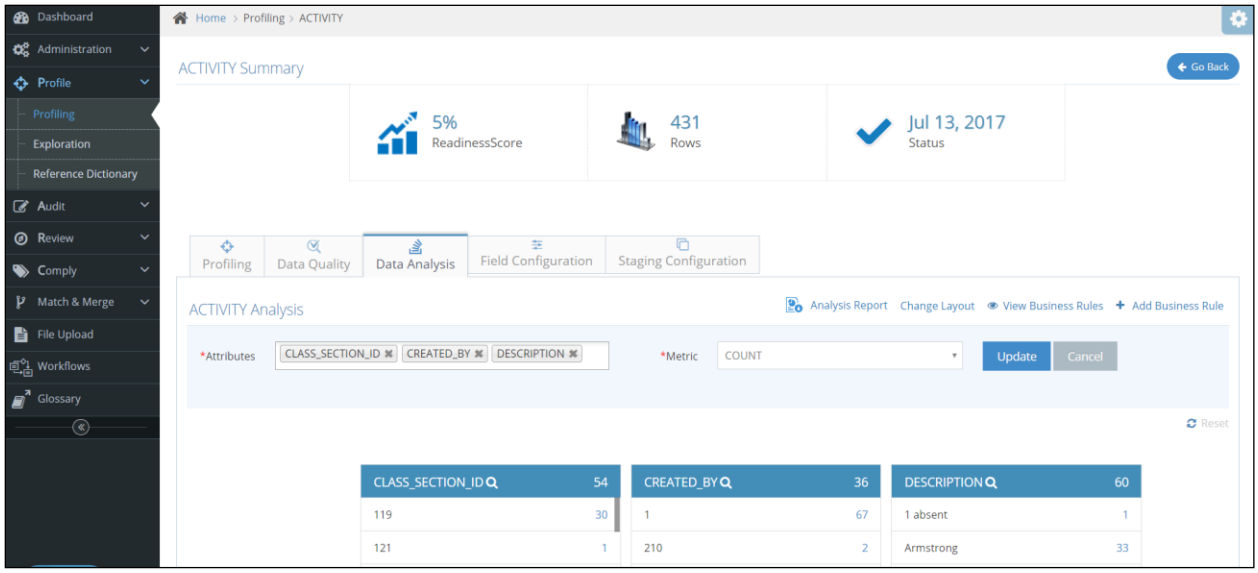




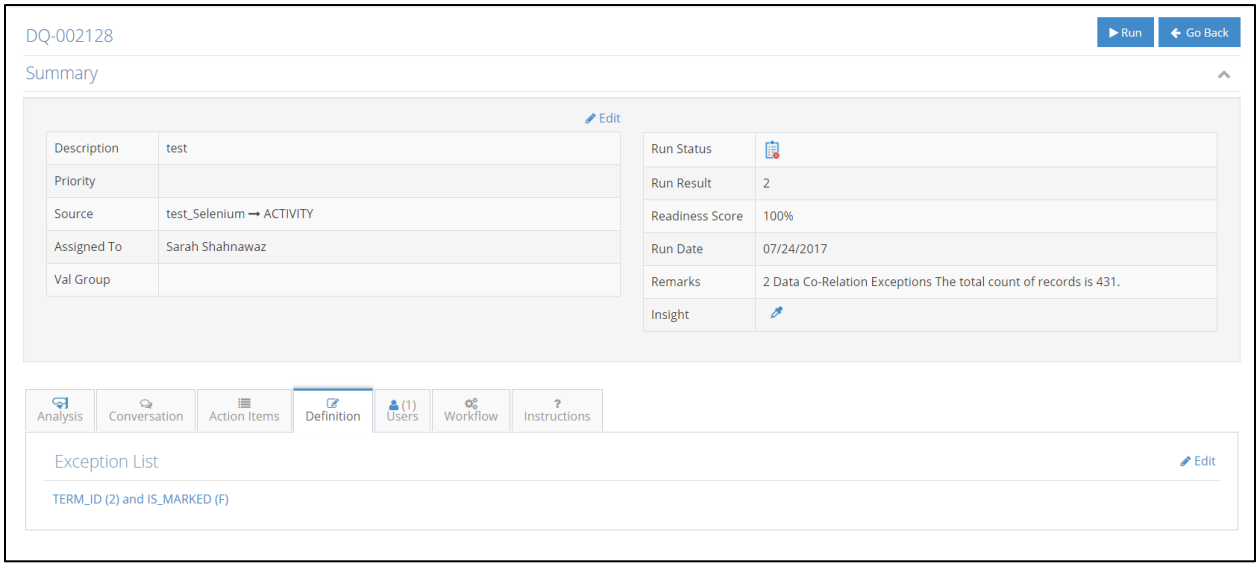
When you run the Ruleset, the Run Status and Run Result is shown accordingly. This rule is available as an independent rule in Manage Rules as well as part of the Rule Set.

13.1.2 From Profiling Data Analysis tab (DAE Rule)

From Profiling tab, select a Source Table and click on “Show Details”. Click on Data Analysis Tab. Select one or multiple Attributes (columns of a table) and the metric. Click on update to view the selected columns of a table.

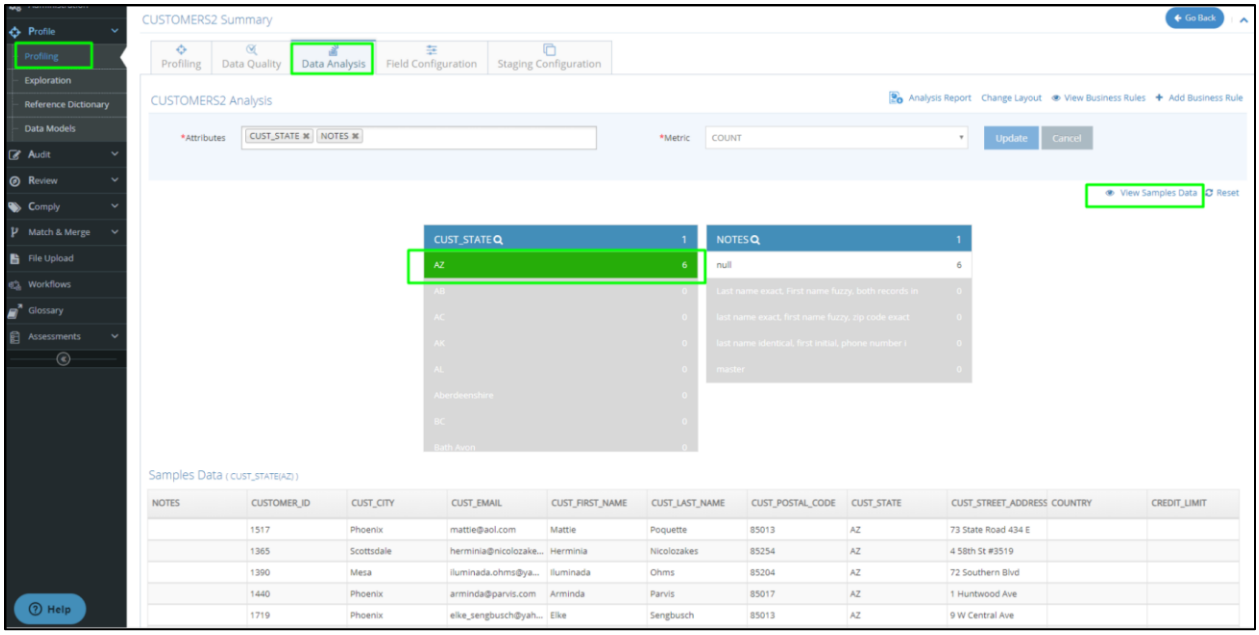


Now click on "Add Business Rule" and add description. Once the Data Analysis Business Rule is created, now you can add rules. Select the field and click on Add rule to the exception list. You can add multiple rules here. When this rule is run, the result is shown against the exception rules list.



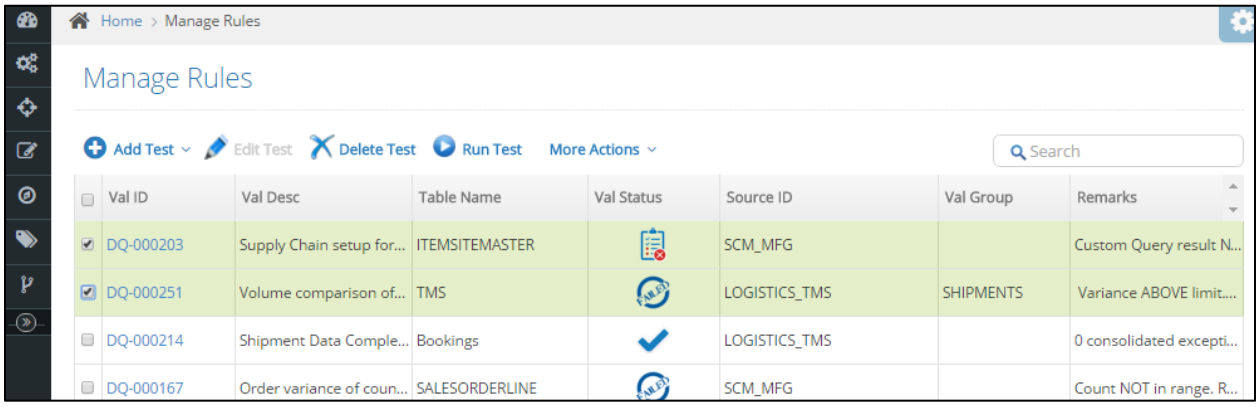
13.1.2.1 Sample Data Drill-down

If you wish to see drill down records based on the selected column value without the need to create DAE rule, you can select a specific value of an attribute and "View Sample Data". For e.g. selecting CUST_STATE as AZ and clicking on Viewing Sample Data will bring all records on the same page below. You can even select multiple attributes at a time.



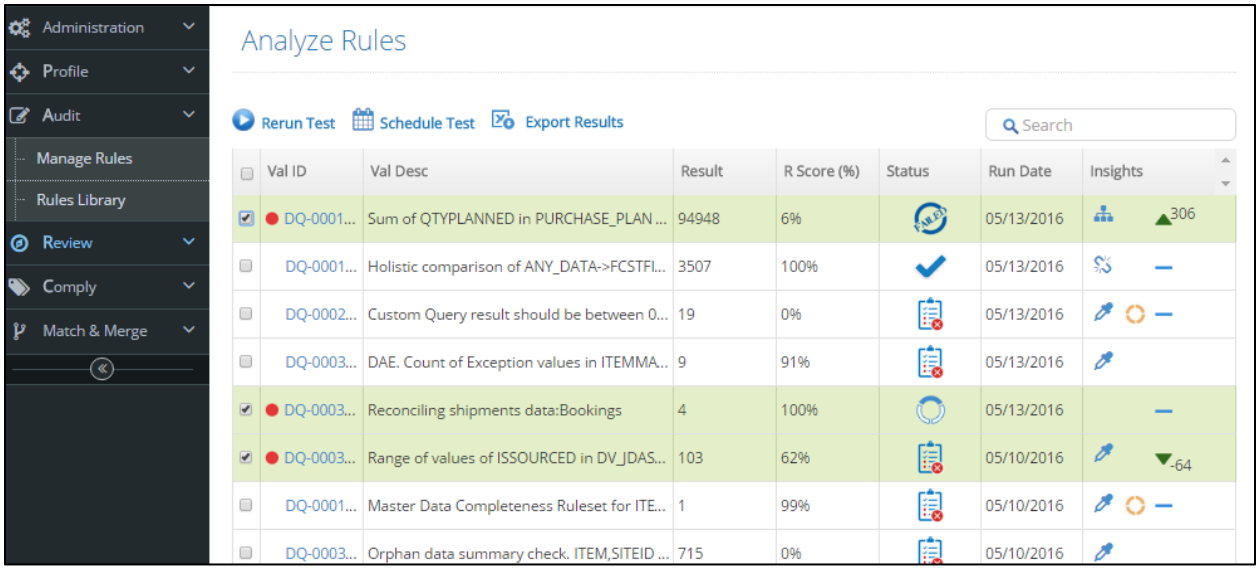
13.1.3 From Manage Tests workflow

When you create an audit using the Manage Tests workflow, upon saving the audit, it is automatically executed. You can execute on-demand by clicking on the Run Test button



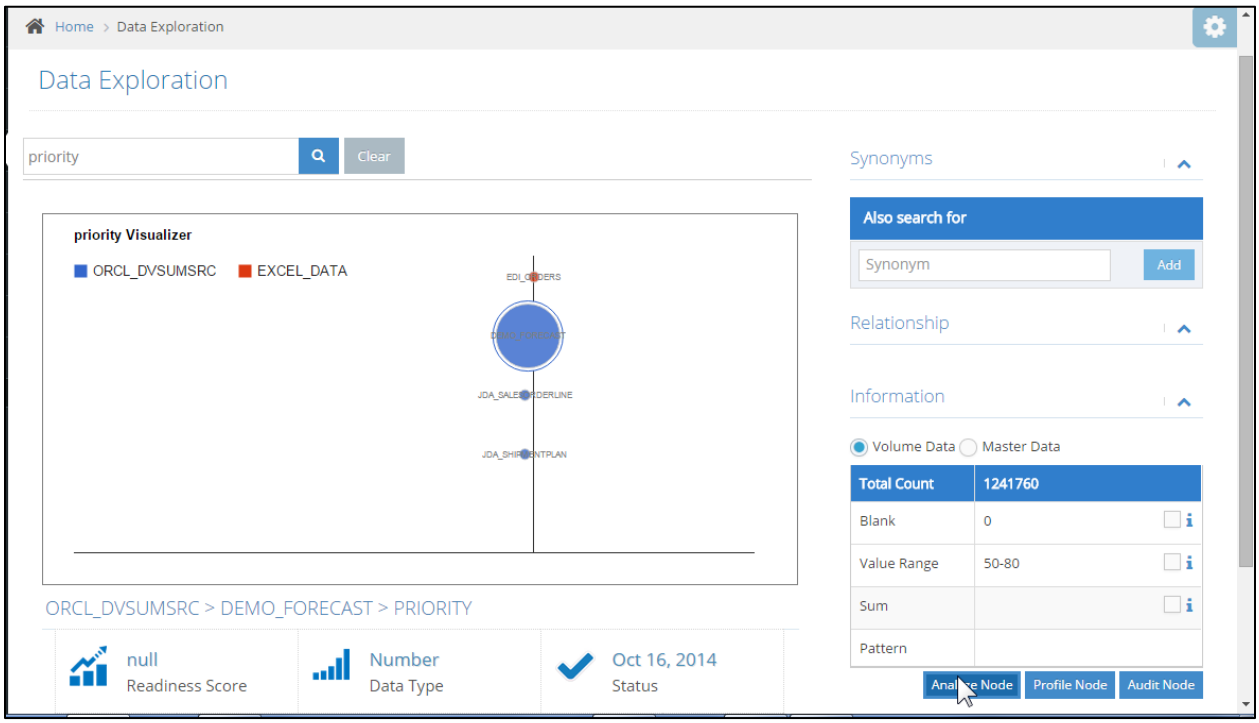
13.1.4 From Analyze Tests workflow

This is the view that is used primarily for analysis and is the most common approach to reviewing and re-running rules. You can select one or multiple tests and click on “Re-run Tests” to execute.



13.1.5 From Data Exploration workflow

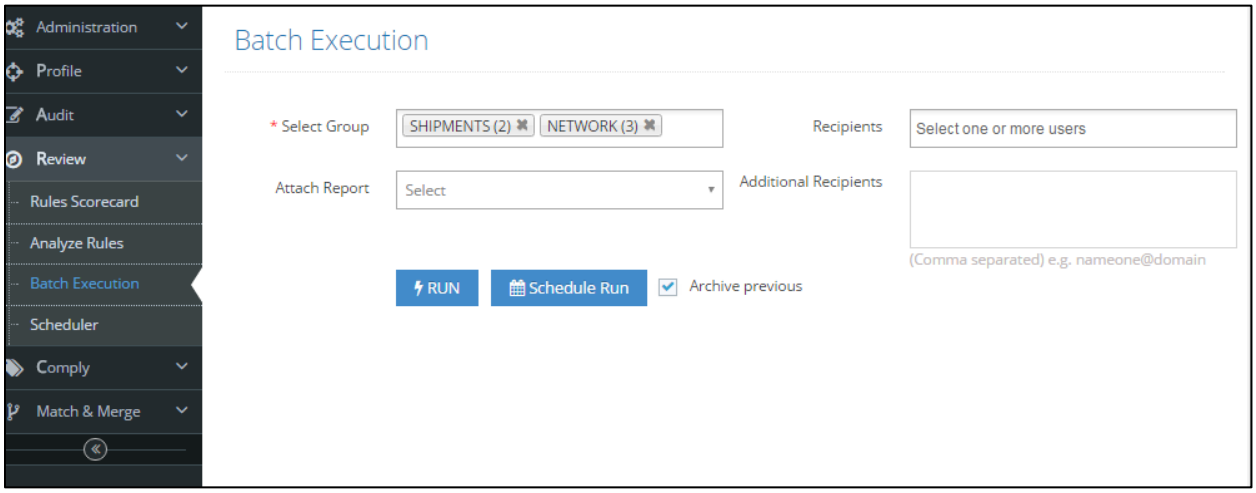
All audit rules specific to a particular node shown in data exploration can be execute by selecting the node and clicking on the Analyze Node button



13.1.6 Batch Execution

Batch Execution is a specialized workflow for batch execution of audit rules. It is only available for owner, administrator and super-user roles. In the Run all workflow, you can select audit rules by their group and

execute one or multiple groups. A history trend of batch execution statistics is also displayed for quick reference.



13.2 Custom Query (Advanced Rule)

The rule type like Blanks or Value range specifies what it does and the user only has to provide the data source and table name to apply these rules on. When there are exceptions, they can be “Exported” as an Excel file from Analysis tab.

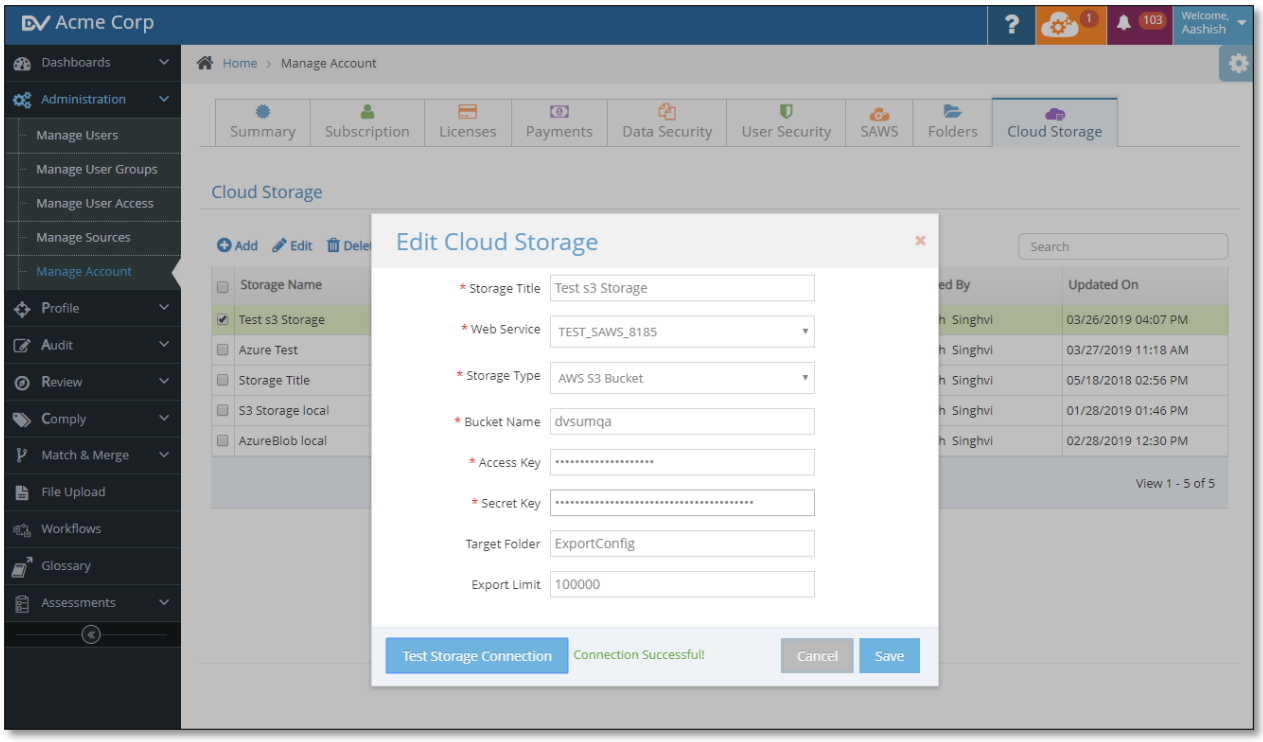
Sometimes a user may require to create a query if his own. For that you can select a Custom Query rule. This is an advanced rule and it offers a few options to users on how to manage data, i.e. where it should be uploaded to cloud storage like AWS S3 or Azure blob storage or it should be displayed on DvSum analysis tab. You can also select to display the run result outside the DvSum domain if the rule is executed using the Curl Command.

This feature is supported by all exception rules. On the Rule detail page, there is a tab "Script Configuration". You can set the export configuration to send the data to either AWS S3 Bucket/Azure Bucket/Folder or send the data to calling API.

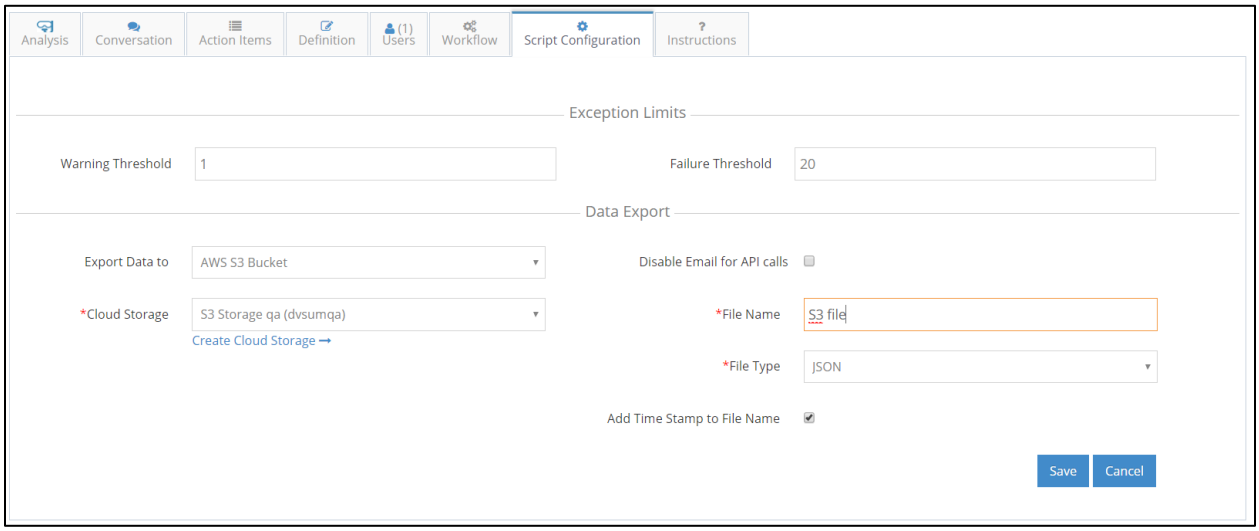
13.2.1 AWS S3 Storage/ Azure Storage

From Administration > Manage Account > Cloud Storage, click on Add and select the Storage type and the web-service. Depending on the Storage type, you will provide the Access key and fill out the mandatory fields.
“Export limit” allows user to set the limit for records to be extracted on Azure/S3. DvSum can extract 10 million records at max.

Once you save this, you can test your connection with Cloud Storage.



Now from rule detail page you can select this storage, give a file name and type, say the file name is "S3 file" and the type selected is CSV.



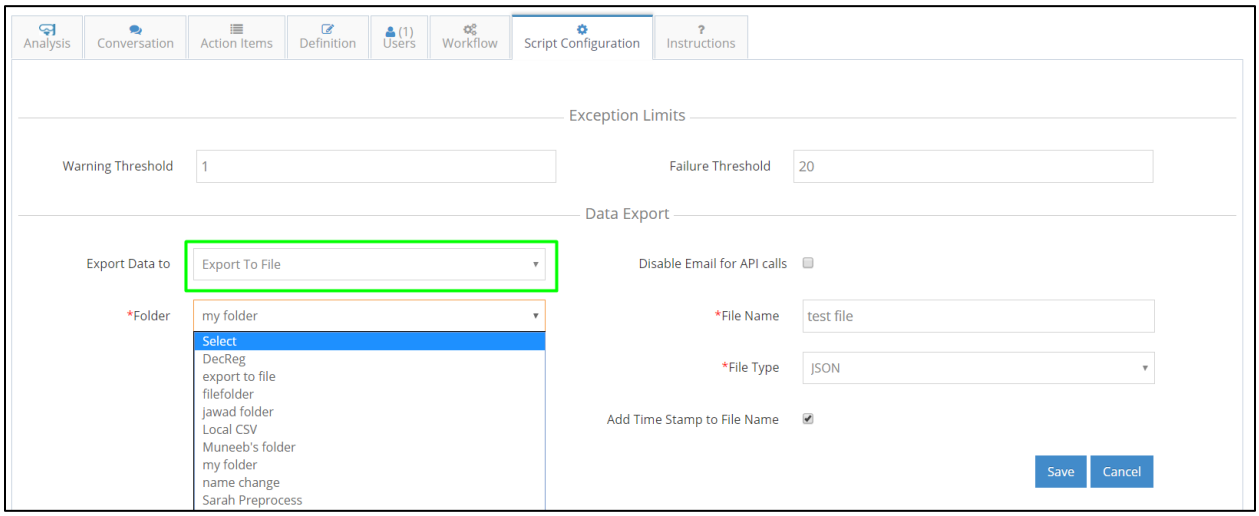
When this rule is executed, the results will be uploaded to AWS S3 bucket in a file named "S3 file".

Similarly, for Azure, you can set up Azure storage from Cloud Storage tab and select the same Azure Storage on rule detail page.

13.2.2 Export to File

Another option to manage rule report is to export it to a folder. This folder can be on your local machine or wherever the webservice is running. Folder is created from Administration tab, explained in File Upload section below (Heading 12)

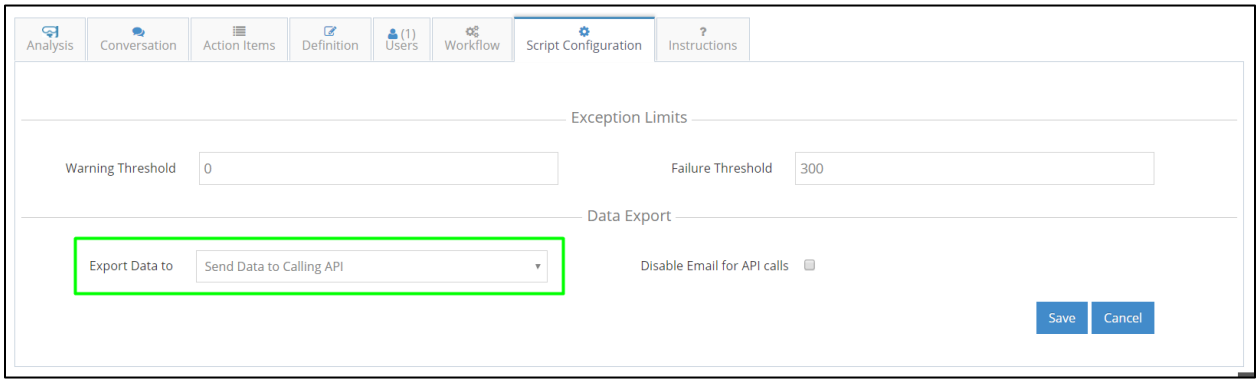
When the option “Export to File” is selected, you will be able to see the folders which are associated with the same SAWS as the source on which rule is created. i.e. The SAWS executing the rule and uploading the report to the folder should be same.



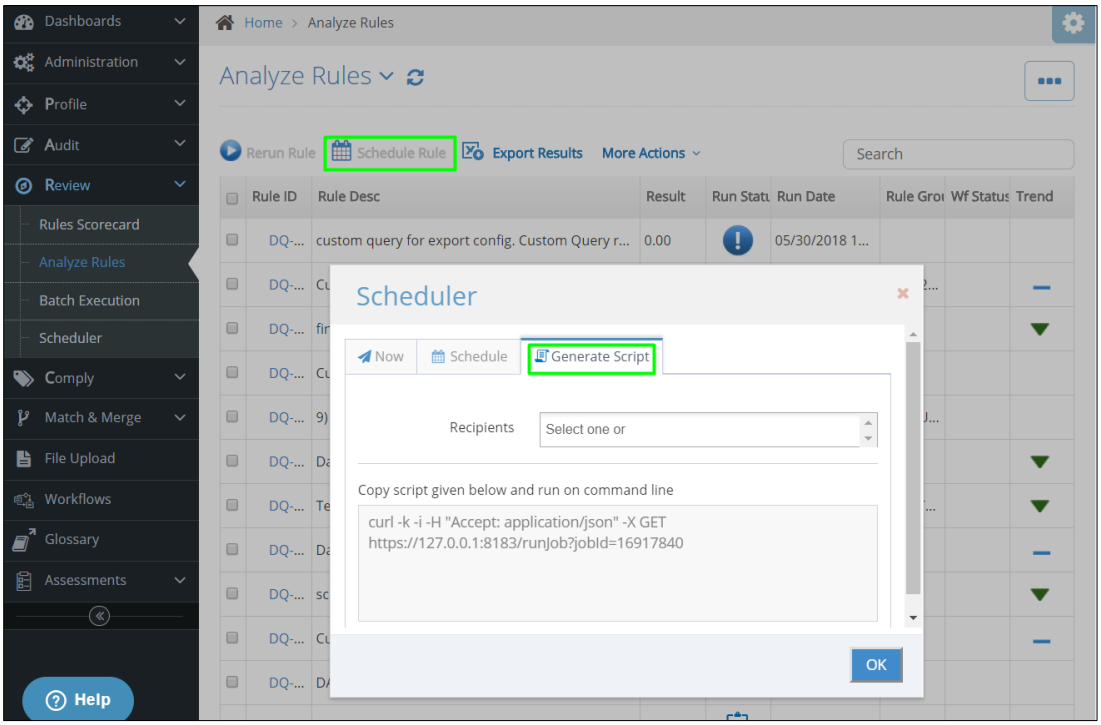
13.2.3 Send Data to Calling API

This option allows you to execute rule and fetch the results of that rule outside the DvSum domain. As the name suggests, the run result will be sent to the calling API.

From “Script Configuration” tab in rule detail page, choose the option “Send Data to Calling API” and then execute rule via API.

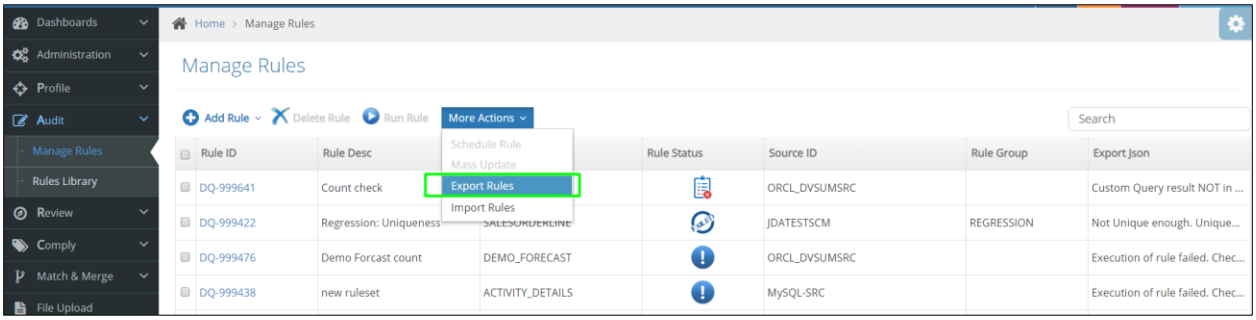


To generate script, schedule a rule and go to Generate Script tab. Copy the script and run from command line. The results for this rule will be returned in command line.



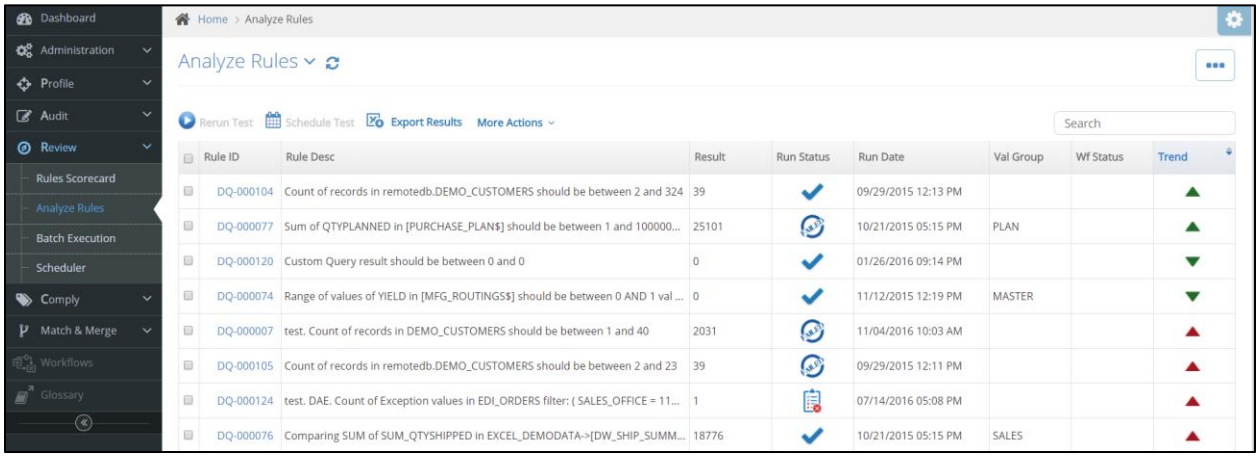
13.3 Export Rules

From Manage Rules page, you can download the excel report containing all rules and their related information.



14 Review (or Analysis)

You only configure the audit rules once. But you can execute them over and over again. Once you execute the audit rules, you need to check where rules are failing or they have exceptions and take corrective action. This is done using the Analysis workflow.
You can access the Analyze workflow by clicking on Analyze Tests on the left navigation or clicking on the Passed, Failed or Exception list on the dashboard.



14.1 Run Status

When a validation rule is executed, it can have either of the following statuses:

Status	Definition	Example
PASS	Audit rule is passing. If it is master data exception, there are 0 exceptions. If it is not of type master data, then the rule is meeting the criteria	Forecast Name is 100% unique in SALES_FORECAST extract
FAIL Or EXCEPTION	Audit rule is not of type master data and the rule is failing Or Audit rule is of type master data and there are 1 or more exceptions	Run over run variance of qtyplanned in PURCHASE_PLAN is 5% for Supplier X which is more than the limit of 3% There are 2 routing records where yield is not between 0 and 1

14.2 Readiness Score

Along with a status, a **readiness score** is also generated. Readiness score provides a second degree of information on the quality of data. If the data rule is failing, how bad is it. Readiness score creates a common unit for measuring the quality of data and allows result of audits calculating the overall data quality score across various data elements and types of audits.

Sample Audit Results	Audit Result	Readiness Score
Count of finish good items in item master is 6000 which is within the tolerance of 5500 and 7000		100%
Count of finish good items in item master is 5000 which is not within the tolerance of 5500 and 7000		91% (5000 is 9% short of 5500)
Count of finished good items in item master is 3000 which is not within the tolerance of 5500 and 7000		54% (3000 is 46% short of 5500). Same test as above is a fail but data quality is poorer than above
Forecast Name is 100% unique in SALES_FORECAST extract		100%
Run over run variance of qtyplanned in PURCHASE_PLAN is 5% for Supplier X which is more than the limit of 3%		33% (2% over limit of 3%)
There are 2 routing records where yield is not between 0 and 1. (total records are 100)		98% (98 out of 100 are good)

14.3 Insights

Once the audits are executed, what are the next steps? This is where the "Insights" functionality of Analyze Tests in DataPARC come into the play. Insights is additional information that DataPARC has generated to help with drilling-down to the specific data quality failures or exceptions and to identify the root-causes.

Following insights are available with your audit results

Insight Type	Definition
History	History for all audits
Exception List	List of Exceptions for master data audits
Drill-down Analytical	Drill-down with Variance for aggregation audits
Drill-down Integrity	Drill-down for comparison type audits

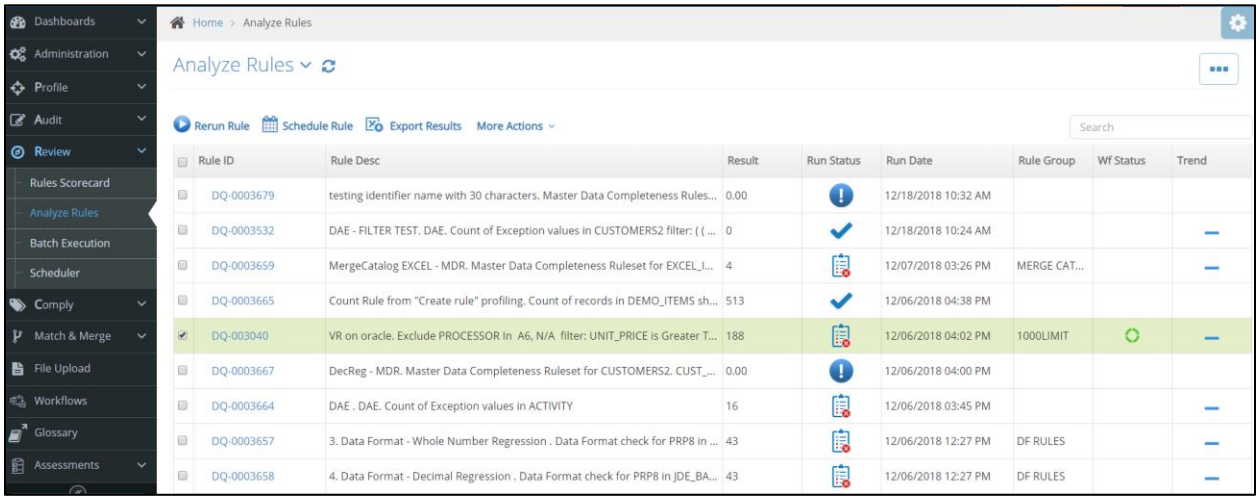
14.3.1 History Insights

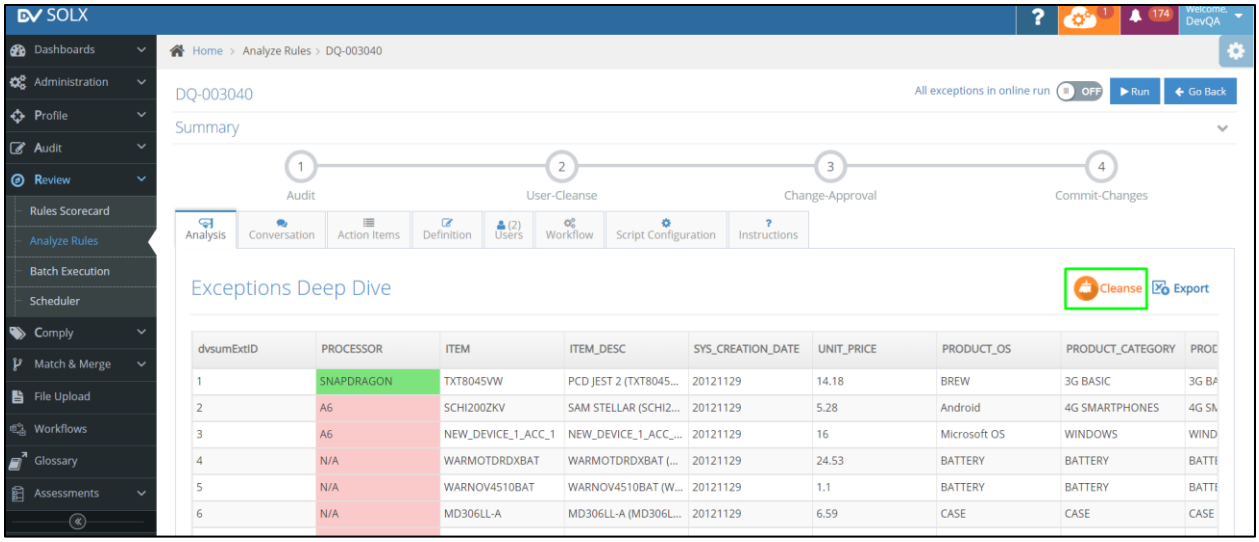
With history, you can get insights into the changes to the audit result from last run and also the trend of that audit over time. History insights can be useful to identify what audits to focus on.

History Trend	Definition	Focus Required
—	Flat. No change	Status quo
▲1075 ▼-105 or	**Positive Up** – Test was a fail before and now pass. Test has been failing because results too low but now trending up. **Positive Down** – Test has been failing because results were too high and now trending down.	Things are improving even if the tests are failing. May not need to focus.
▼-1075 ▲105 or	**Negative Down** – Test was passing before and is now failing. Or it has been passing but values are trending down and will cross the lower tolerance soon. Or it has been failing and things are getting worse **Negative Up** – Test was passing before and is not failing. Or it has been passing but values are trending up to cross the upper tolerance soon	Things are getting worse for failed tests – requires top priority focus. Things are trending in the wrong direction for passing tests – heads-up for future issues

14.4 Exception List

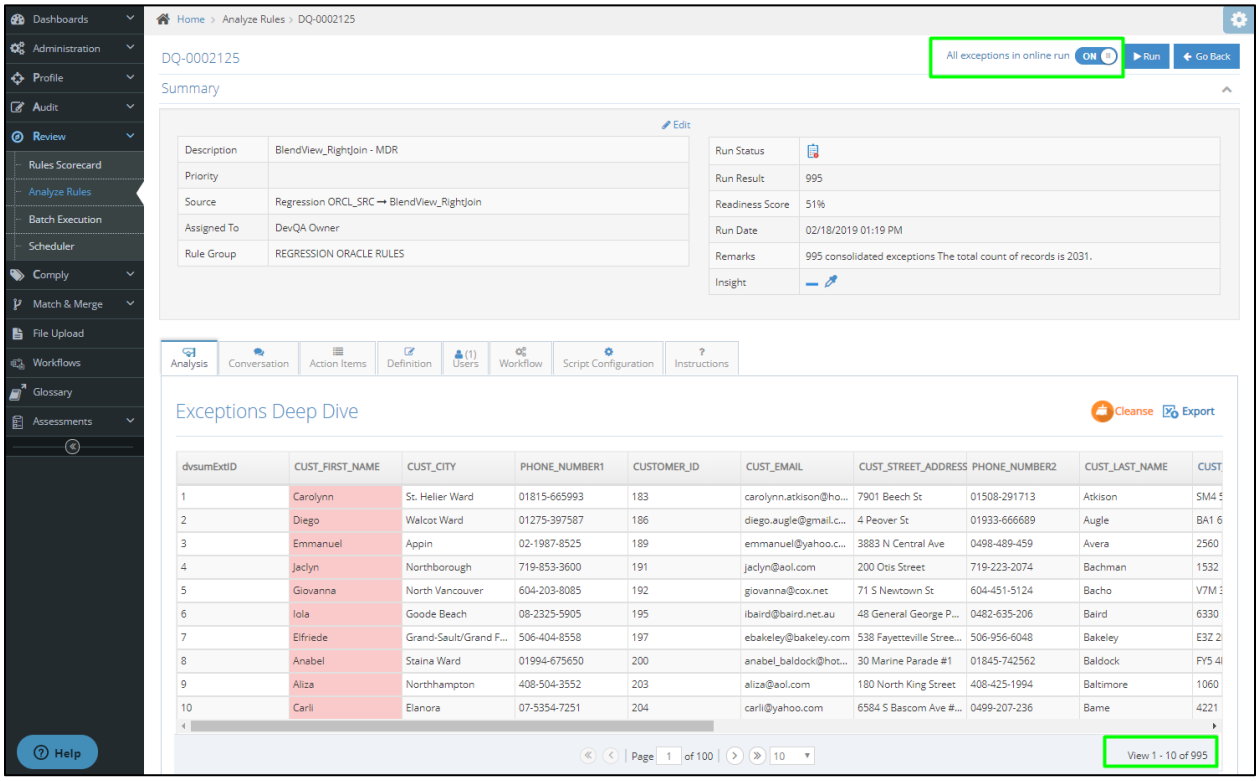
For master data exceptions, when exceptions exist, you can click on the exception rule and go to rule detail page.





Click on Analysis tab. Here you will be able to see the exception records brought by the rule you created.

By default, the rule execution brings only 300 records whereas your data source may have more. If you want to view all records on UI of Analysis tab, you must turn on "**All exceptions in online run**" button on top right. Now rule execution will bring maximum of 10,000 records.



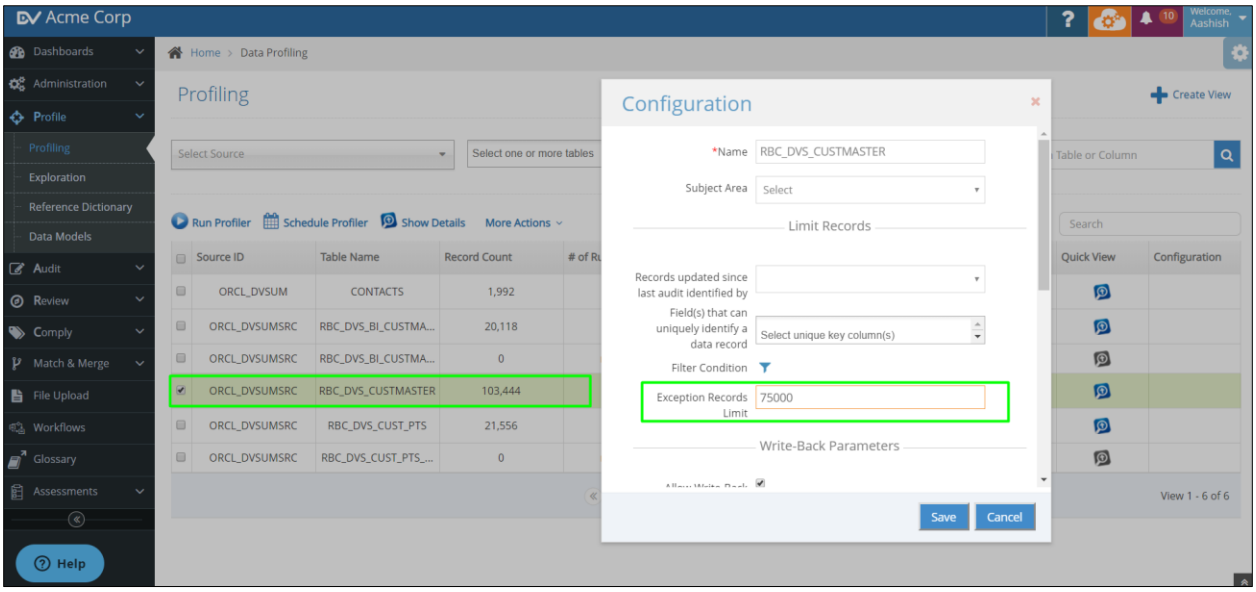
From Analysis tab, identify which records in the source system to update to fix the data. Once you fix the data, you can run the audit again to confirm that the exceptions have been resolved.

If you want to save the exception report, you can click on “Export” button next to Cleanse. It will download and excel file containing exceptions on this rule. If “All exceptions in online run” button is off, excel download will have 300 records and if it is on, it will contain up to 10,000 records.

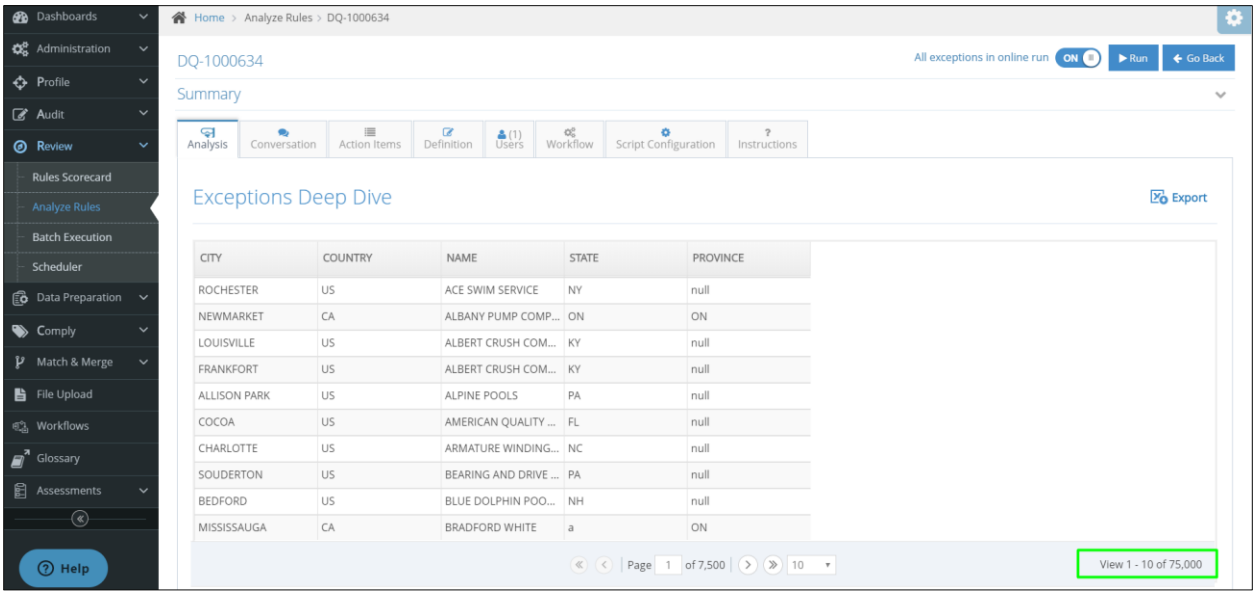
14.4.1 Increasing Limit of exception records

Please note that by default limit for “All exceptions in online run” is 10,000. However, you can increase it to a maximum of 75,000 records from table configurations on Profiling page.

From Profiling page select the source, table and from More actions, go to “Edit Configurations” and update the exception record limit to 75,000.

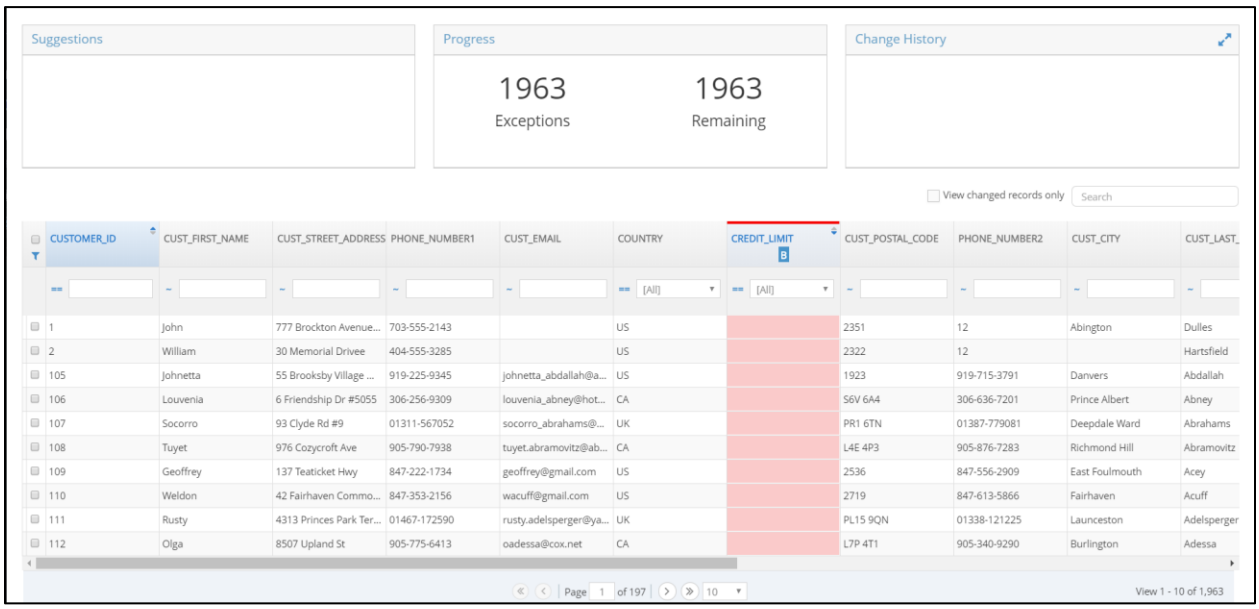


Any rule created on this table will now bring up to 75,000 exceptions on UI.



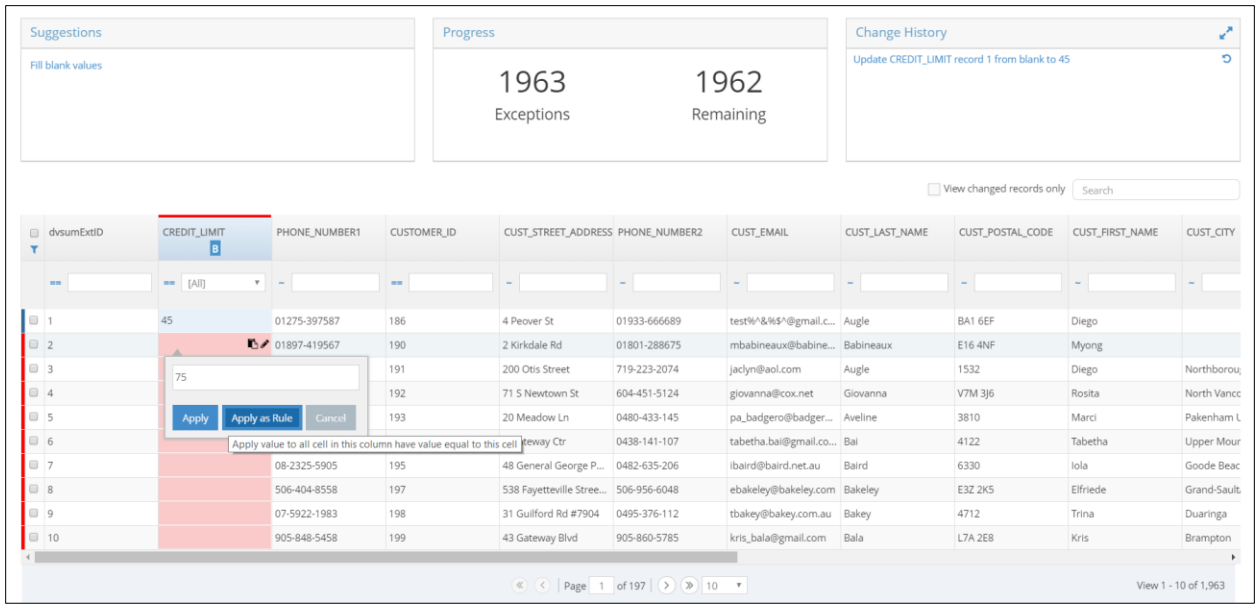
14.5 Cleansing Workbench

On Cleansing workbench, you can clean up the exception records brought up by the rule. From Analysis tab, go to Cleanse.

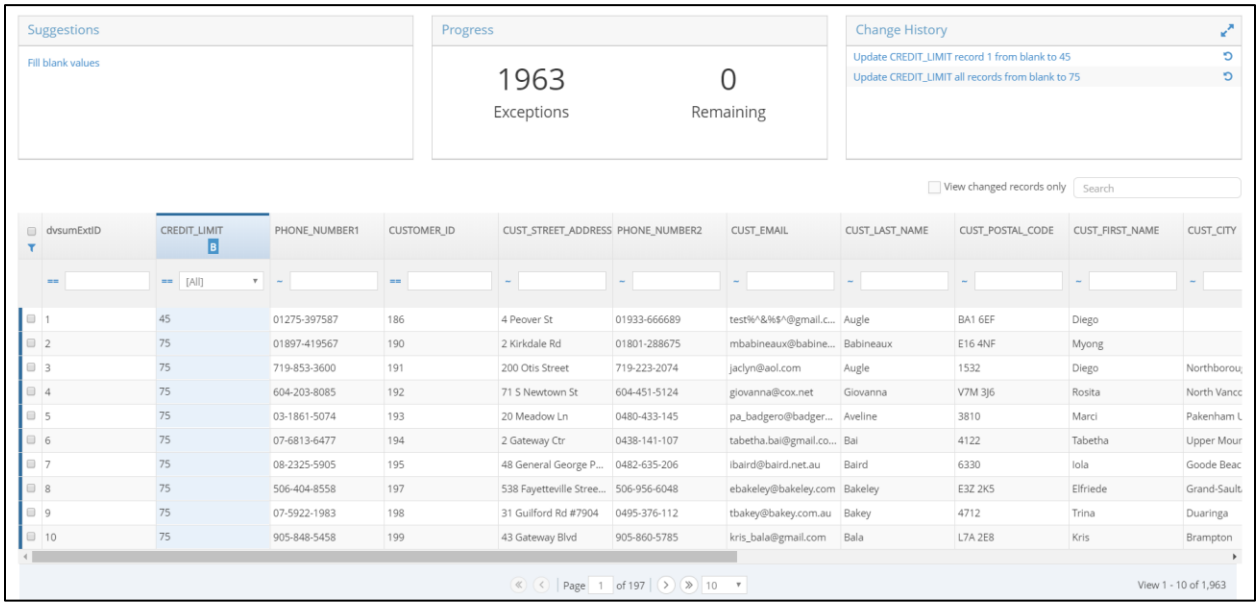


Exception records will be highlighted red and as you cleanse, they will appear as blue cells. There are multiple ways to update records.

1. Applying changes to a single cell; changes the value of one cell
2. Applying changes to selected cells; applies changes to all selected cells
3. Apply as a Rule; applies changes to all cells in this column which have the same value as this particular cell

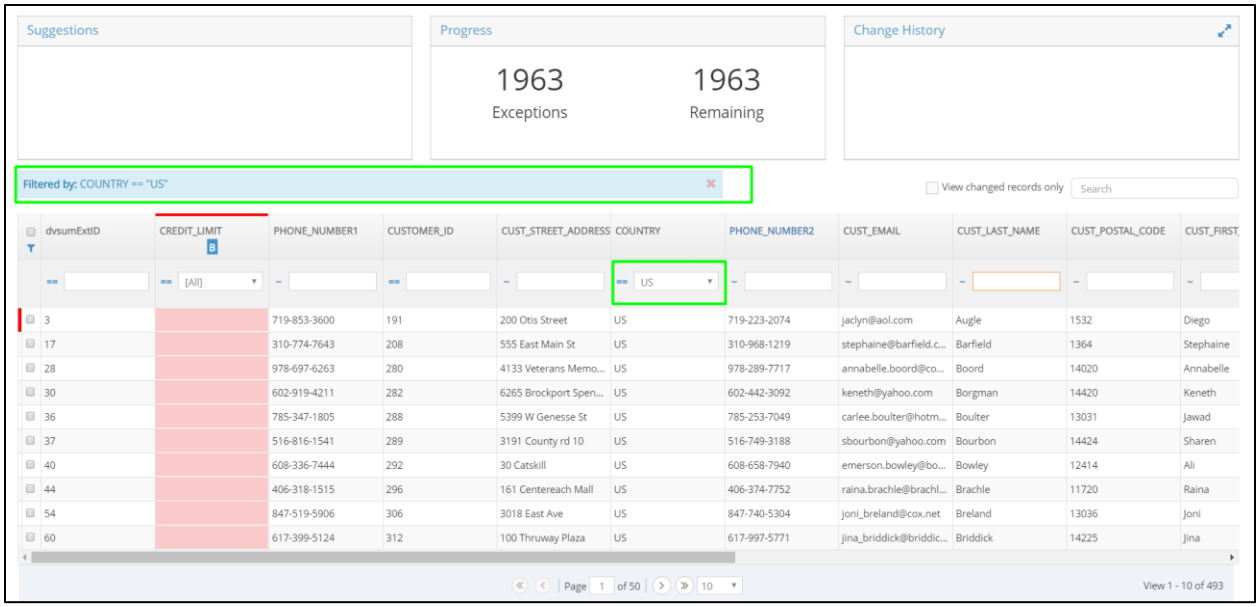


Applying as rule will replace all blanks of CREDIT_LIMIT with 75.



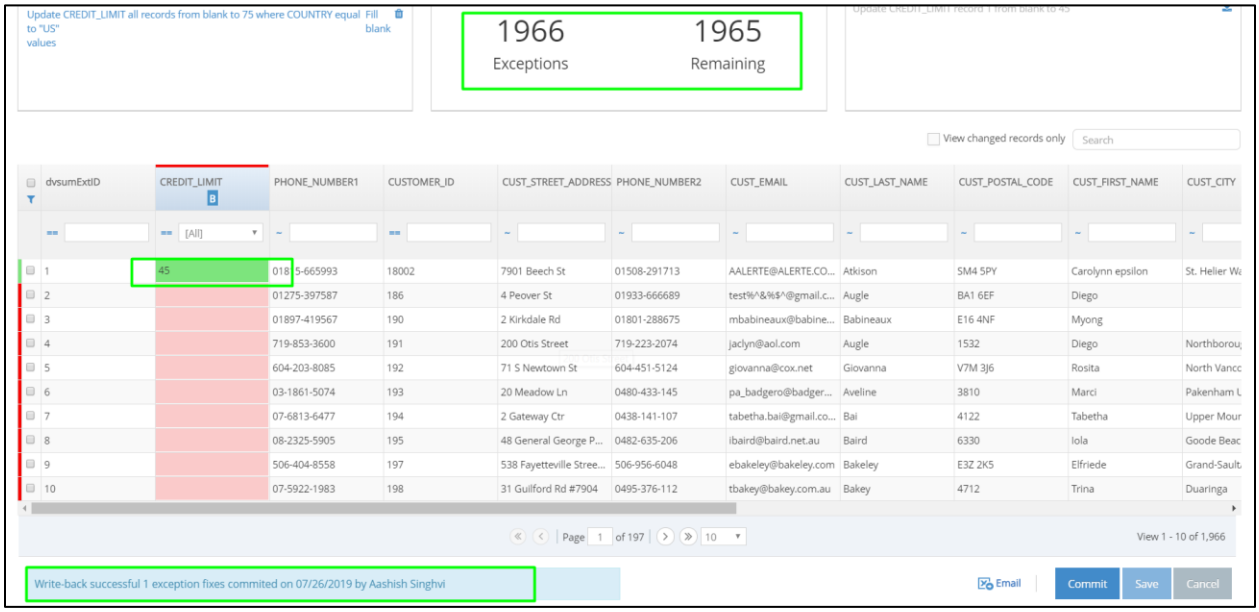
In Change History section, you can see all the changes you make. You can undo the changes from here as well.

A few more features on this workbench including cleansing records by applying filters on a column. For Example, you want to filter all blanks of CREDIT_LIMIT where COUNTRY = US



Once the changes on the workbench are completed, you can save them and commit back to source. Commit operation will write-back the correct records in data source.

Successful commit will show committed records by highlighting them as green. You will also be able to see the total exception count and remaining exception count on workbench.

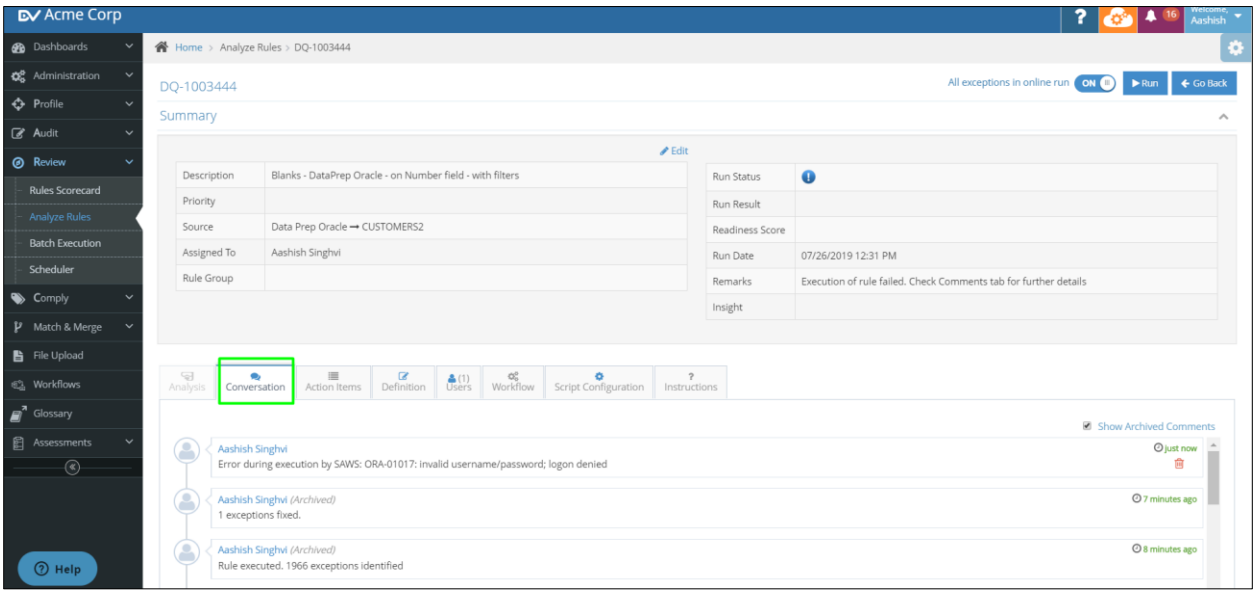


After successful commit, when you execute the rule, it will show exception count less than before.

14.6 Conversation

Conversation tab maintains the audit trail. It will show you when the rule was executed, how many exceptions were identified or how many were fixed. You can write your own comments here as well.

In case there is an error executing the rule, it will also show you the information what could be possibly wrong due to which the rule execution failed.



14.7 Script Configuration

Rules in DvSum can be executed from the UI (Run button), scheduled through the scheduler (heading 9.5), through batch execution (heading 8.1.6) and can also be called programmatically using an API. To run a rule or set of rules through a script, you can generate a script for those rules using the Schedule Rule Menu.

The script takes the form of https://<saws address>:<saws_port>/runJob?jobId=<jobid>

The script runs the rule exactly like running from the UI or the Scheduler, with the difference being what is returned back from the script. Below is a screenshot of what is returned from the script. The 2 elements are return_code and return_remark.

```
HTTP/1.1 200 OK
Date: Wed, 11 Jul 2018 12:32:16 GMT
Access-Control-Allow-Origin: *
Access-Control-Allow-Methods: POST,PUT, GET, OPTIONS, DELETE
Access-Control-Allow-Headers: Origin, cache-control,  X-Requested-With, Content-Type, Accept
Content-Type: application/json;charset=ISO-8859-1
Content-Length: 124
Server: Jetty(9.2.6.v20141205)

<return_code>-1</return_code>
<return_remark>DQ-999616 Warning. System Remarks: 256 exceptions.  Range:0-300</return_remark>
```

The return_code always contains 3 values

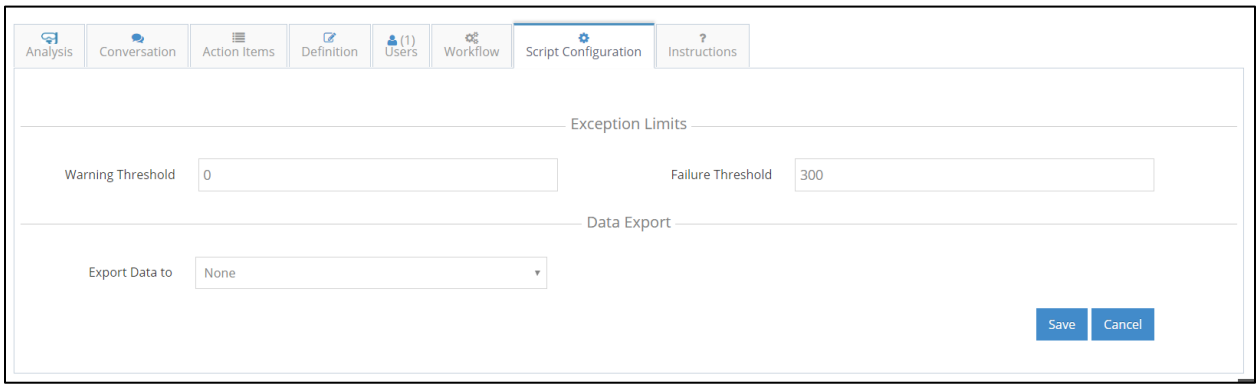
“0 - PASS” if rule has 0 exceptions or exceptions are less than lower bound of threshold

“-1 - WARN” if the rule has exceptions but exceptions are greater than lower bound but less than upper bound

“-2 - FAIL” if the rule has exceptions that are greater than upper bound limit.

The return_remark contains text that will provide the name of the rule additional information that you typically see in system remark field in the Rule Detail page.

You can set this threshold from “Script Configuration” tab in rule detail page.



Example

This rule has threshold set between 0 - 300

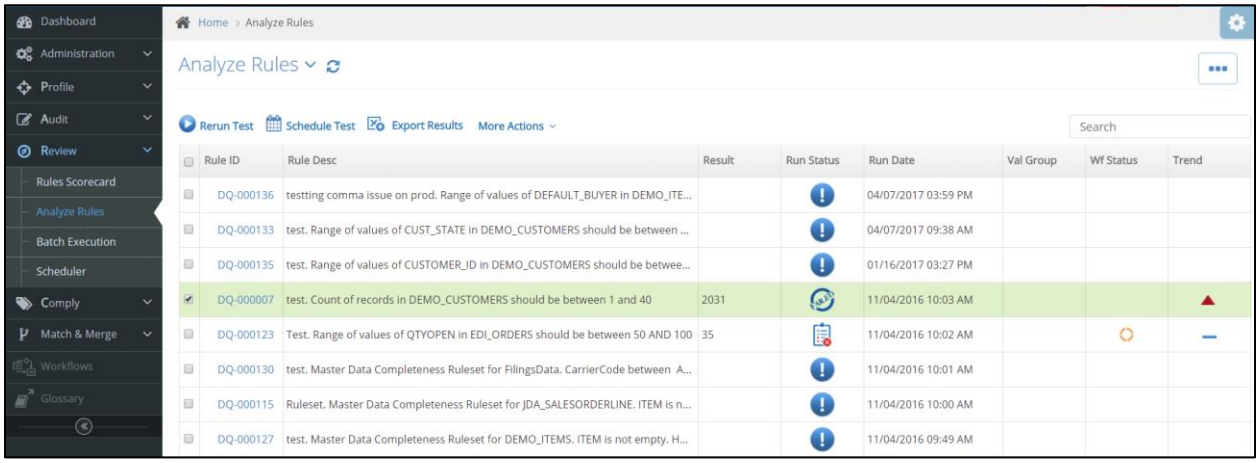
Run Status	
Run Result	256
Readiness Score	
Run Date	07/11/2018 05:32 AM
Remarks	256 exceptions. Range:0-300
Insight	

Since exception are 256 which is > 0 but less than <300, the return_code is -1 as a Warning.

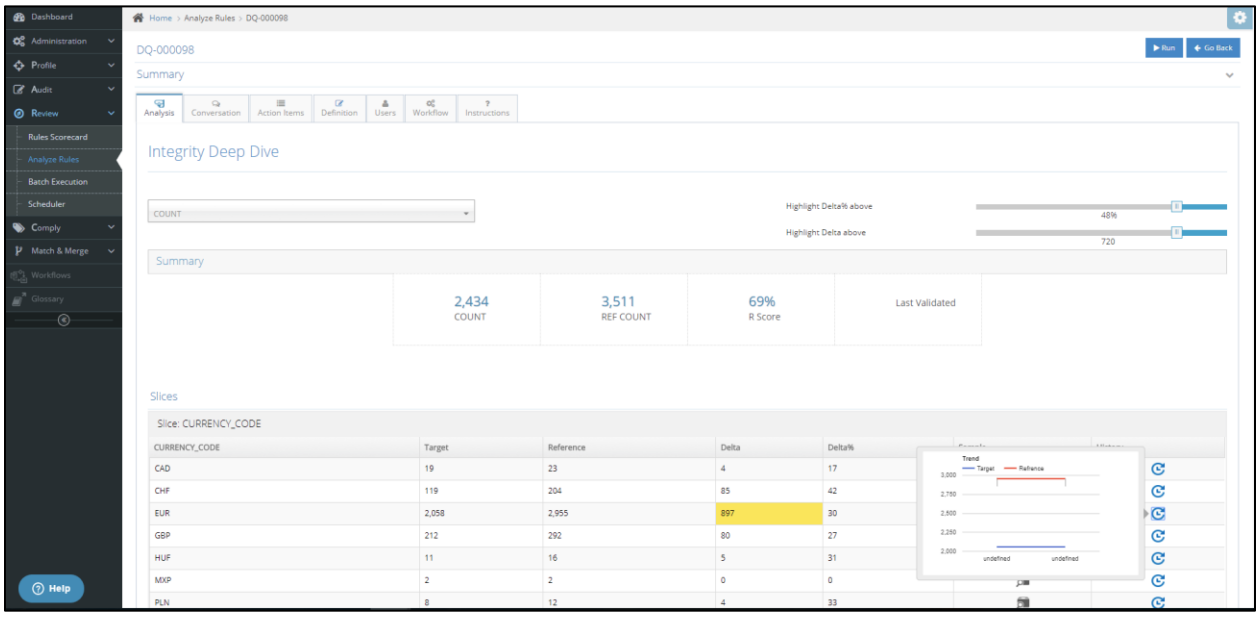
14.8 Drill-down Analytical

For aggregation audit rules like COUNT and SUM, when the totals or variance is not within limit, you can drill-down to determine what part of the data is driving the variance. You can specify the specific attribute

by which you want to drill-down in the audit rule. Or you can let the system automatically drill-down on all possible attributes. The system automatically highlights the top 20% variances using the Pareto methodology.

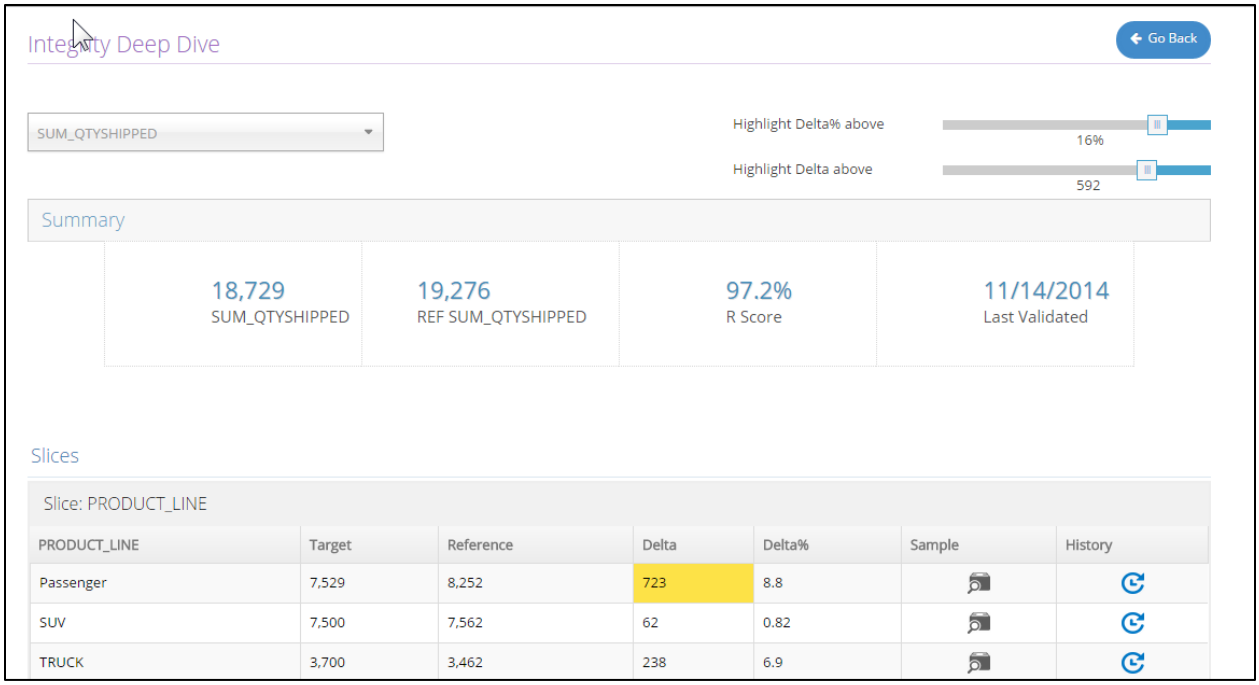


You can also see the history of the drill-down data which can help analyze what might have changed.

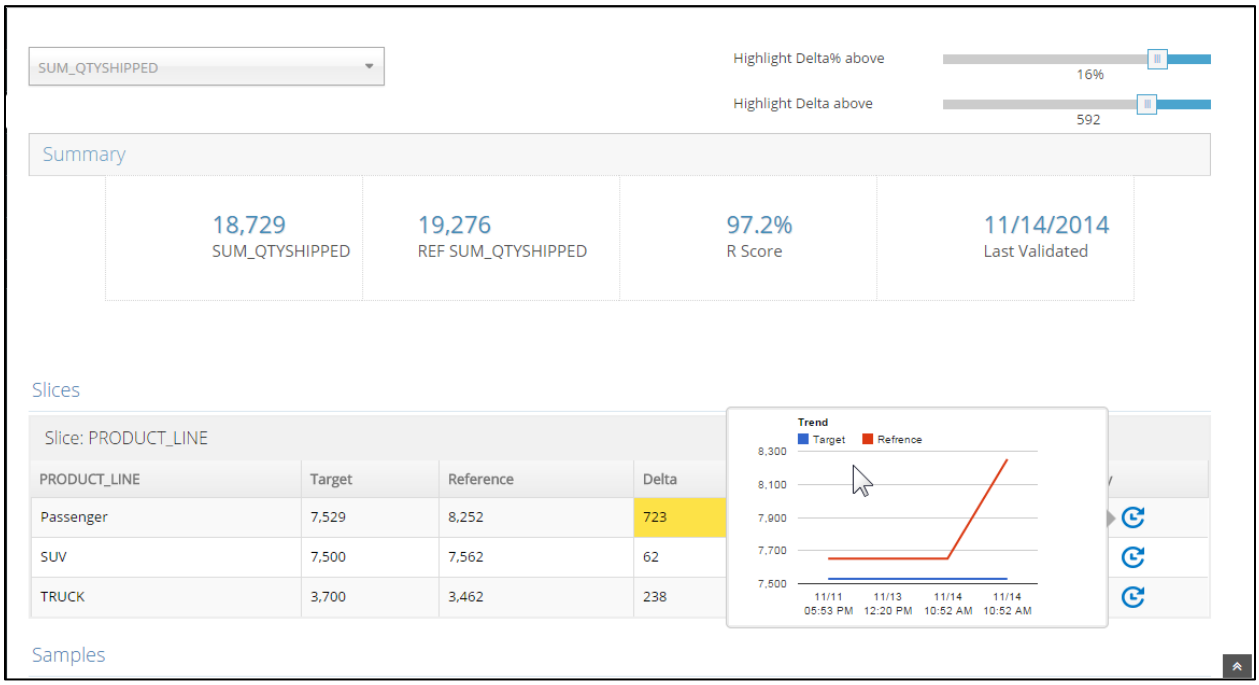


14.8.1 Drill-down Integrity

For comparison type tests like COMPARE COUNT, COMPARE SUM, COMPARE TABLE, or COMPARE CUSTOM QUERY, you can drill-down to determine what slice of data is driving the differences between the 2 data sources. You can specify the specific attribute by which you want to drill-down in the audit rule. Or you can let the system automatically drill-down on all possible attributes. The system automatically highlights the top 20% variances using the Pareto methodology.

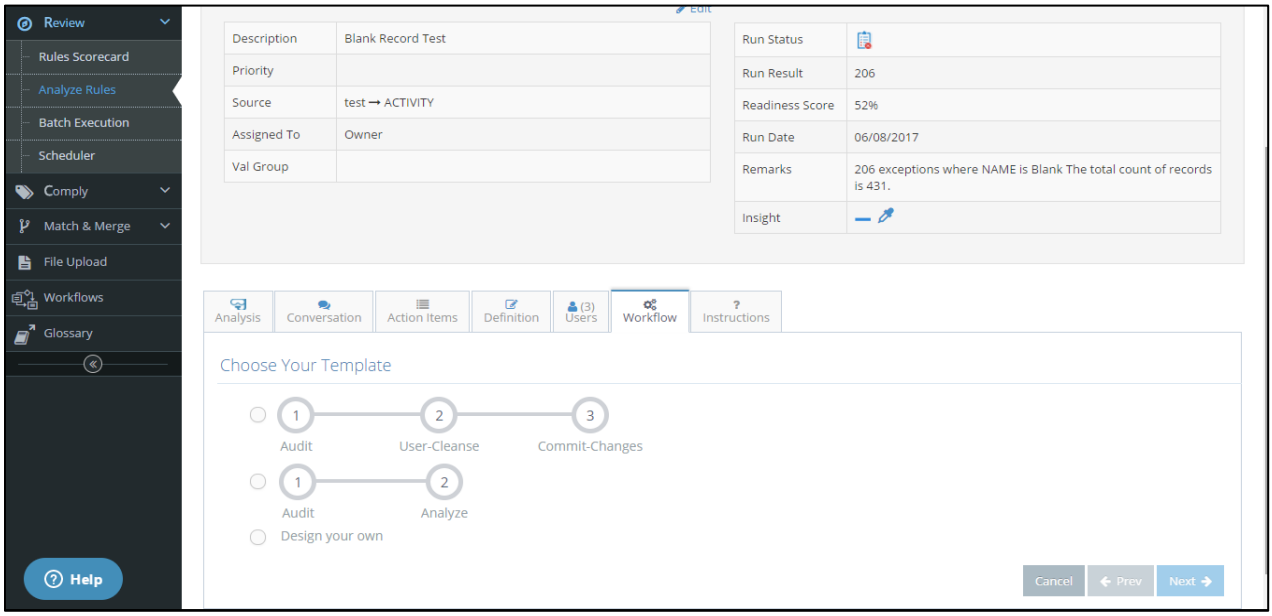


Here again, you can look at the history of the different slices which can help determine what has changed.

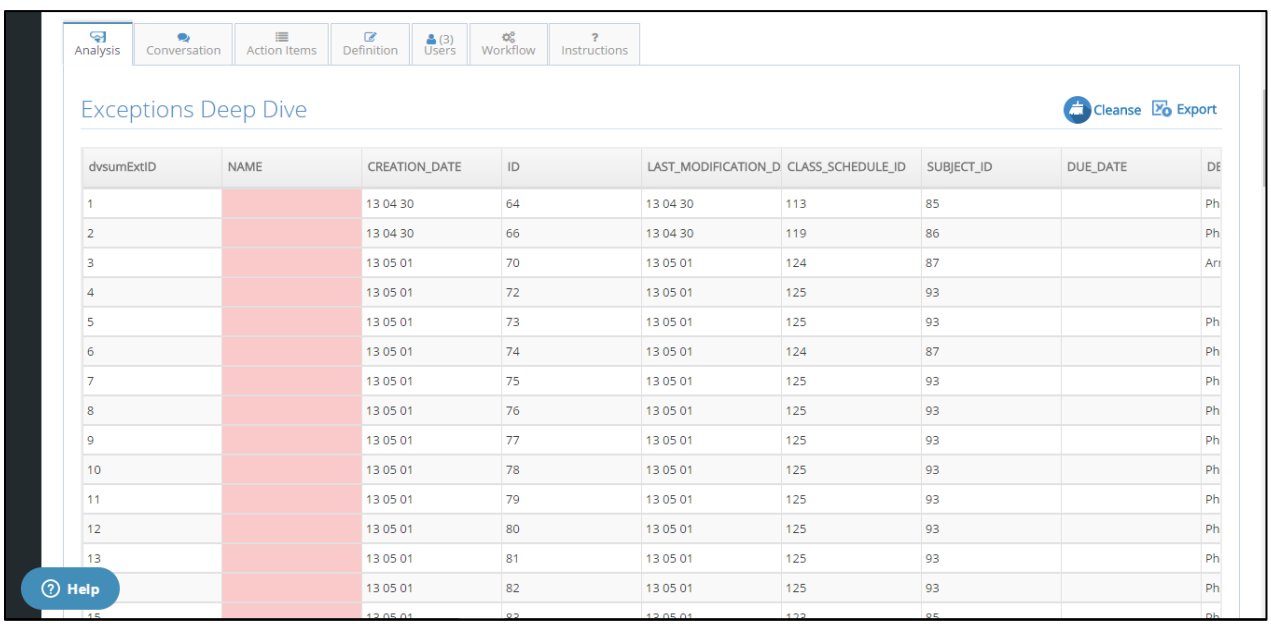


14.9 Data Quality Workflows

DQ Workflows provide users a methodological way of cleansing data by identifying exceptions, sending it to responsible user to cleanse the data and then committing to system after the changes have been approved by the authorized person. Another workflow could be identifying discrepancy, sending it for analysis and create action items. DvSum also empowers users to design their own workflow.

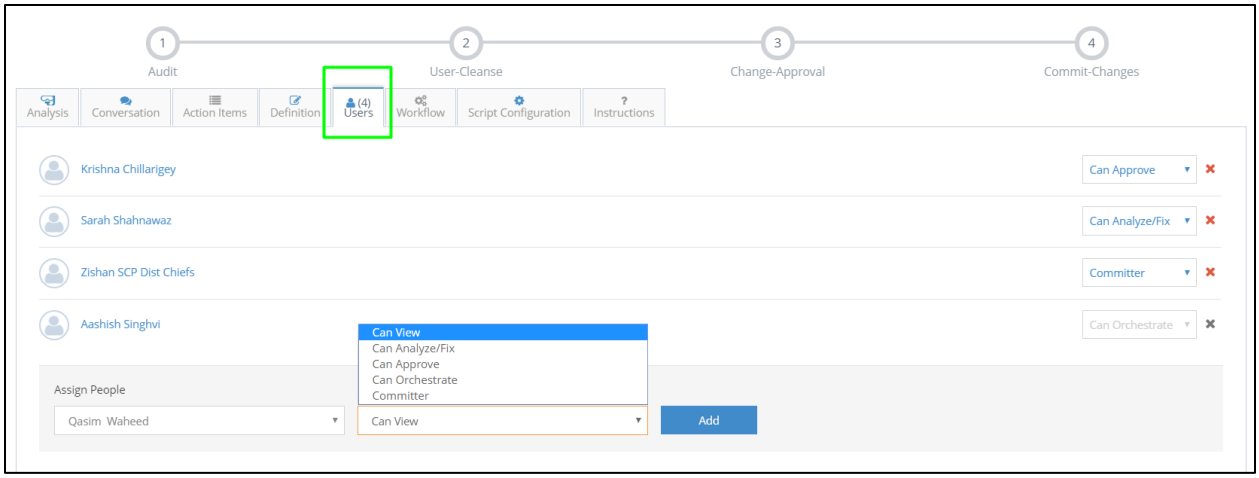


The audits are created on the column level of a source table and once the audit has been executed, the exceptions (if any) can be found highlighted pink in the Exceptions Deep Dive insight.

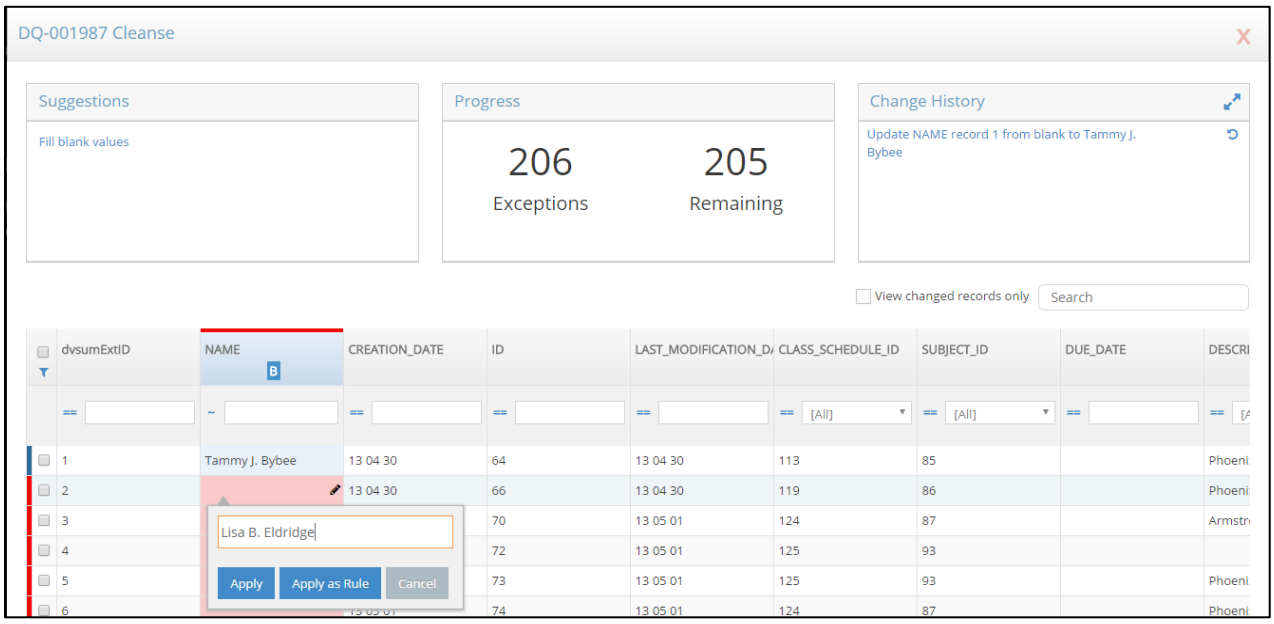


dvsumExtID	NAME	CREATION_DATE	ID	LAST_MODIFICATION_D	CLASS_SCHEDULE_ID	SUBJECT_ID	DUE_DATE	DE
1		13 04 30	64	13 04 30	113	85		Ph
2		13 04 30	66	13 04 30	119	86		Ph
3		13 05 01	70	13 05 01	124	87		Ar
4		13 05 01	72	13 05 01	125	93		
5		13 05 01	73	13 05 01	125	93		Ph
6		13 05 01	74	13 05 01	124	87		Ph
7		13 05 01	75	13 05 01	125	93		Ph
8		13 05 01	76	13 05 01	125	93		Ph
9		13 05 01	77	13 05 01	125	93		Ph
10		13 05 01	78	13 05 01	125	93		Ph
11		13 05 01	79	13 05 01	125	93		Ph
12		13 05 01	80	13 05 01	125	93		Ph
13		13 05 01	81	13 05 01	125	93		Ph
		13 05 01	82	13 05 01	125	93		Ph

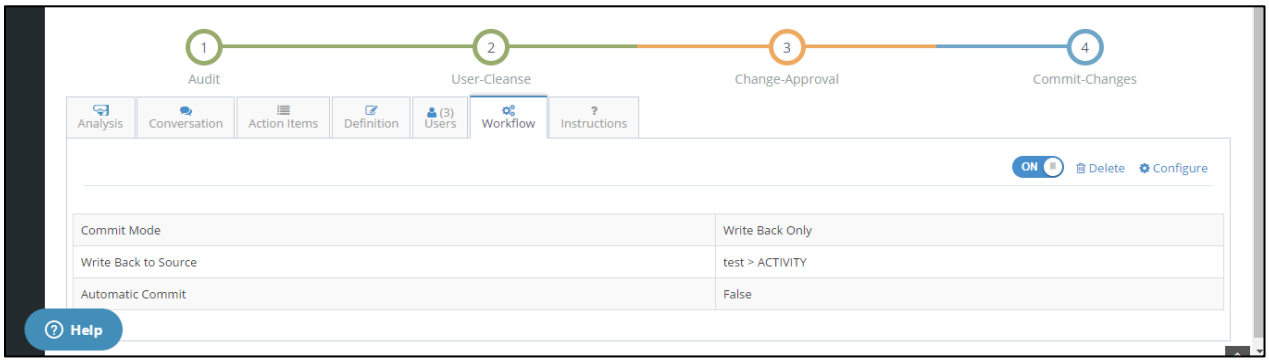
The orchestrator must assign the workflow tasks to users to make it functional. On the rule detail page, there is Users tab right next to Workflow. You can assign workflow tasks to users and define their roles in the user tab.



After the authorized person for fixing data cleanses the inaccurate or corrupt records, then the changes are submitted for approval before they can be written back on database.



The currently active step is shown as “orange”. The steps that are completed are shown as “green” and “blue” shows the succeeding steps.



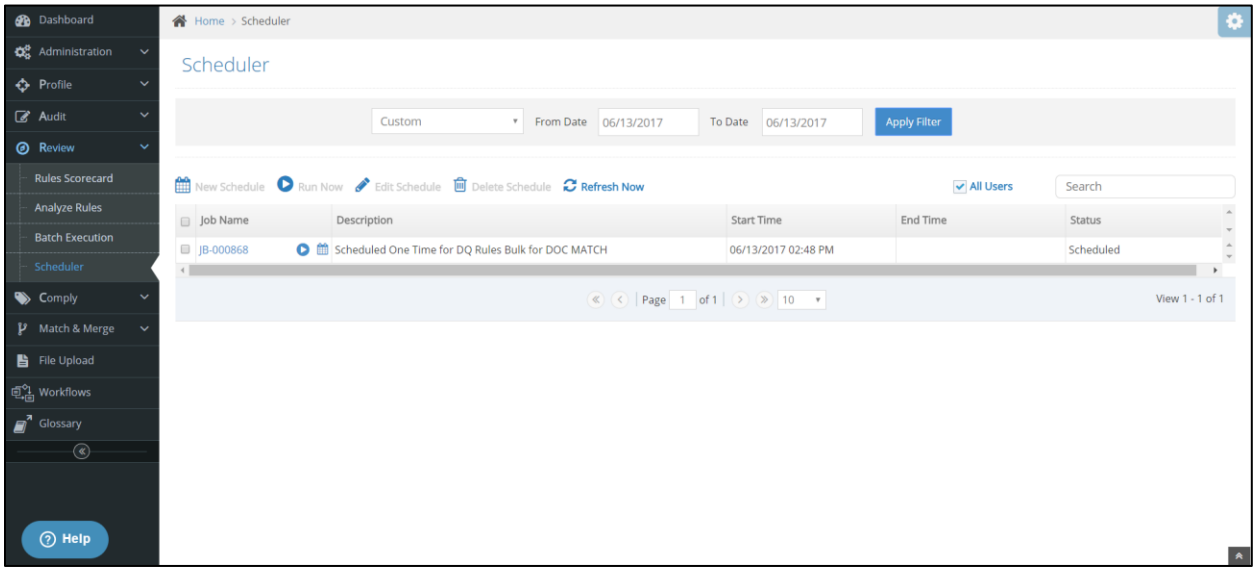
Among the Workflow Configuration Steps, there is a setting of commit parameter as "Off-line Write-back to source".

By default, when a DQ Workflow executes, DvSum writes back the fixes of exceptions to the source. With offline write-back, user can cleanse the exception records on cleanse workbench UI and save changes but it would be required to manually commit fixes to source outside DvSum.

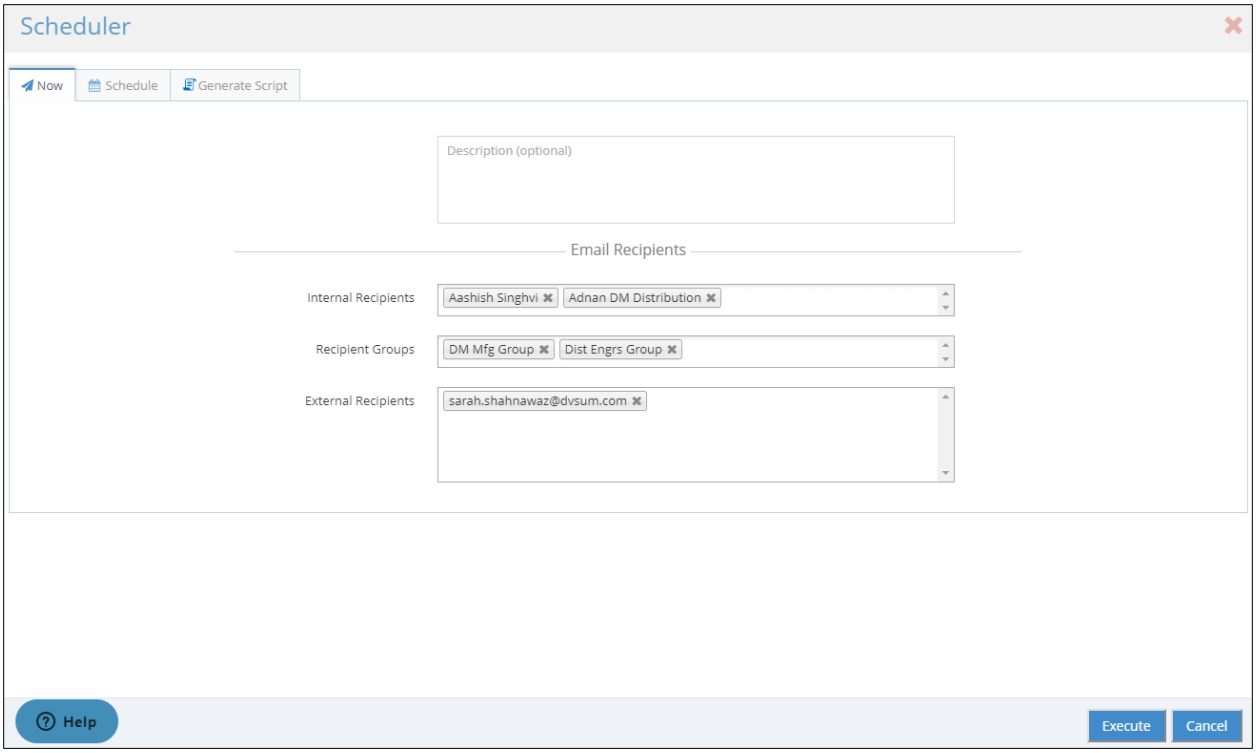
14.10 Scheduler

DvSum Scheduler allows you to schedule and automatically execute jobs – such as profiling data sources and running audits - on the cloud. You can run jobs right away, on a recurring schedule, or pin it down for future.

Scheduler can be used for profiling multiple sources and multiple audits respectively. It can also execute numerous scheduled jobs at the same time. These scheduled jobs can be edited, created, deleted and viewed from the scheduler tab by applying the required filter.



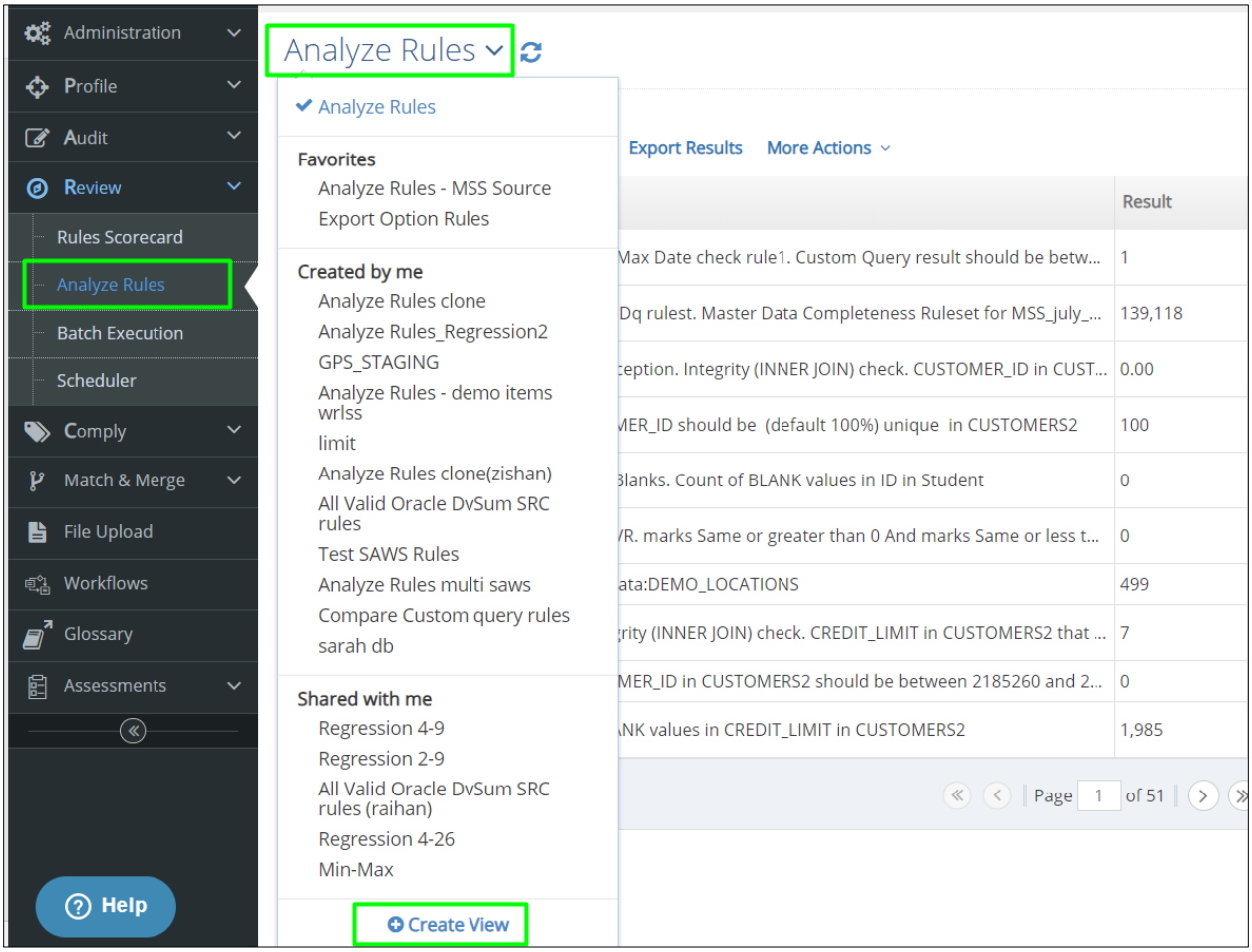
You can also send scheduler emails to other users which may be internal or external users.

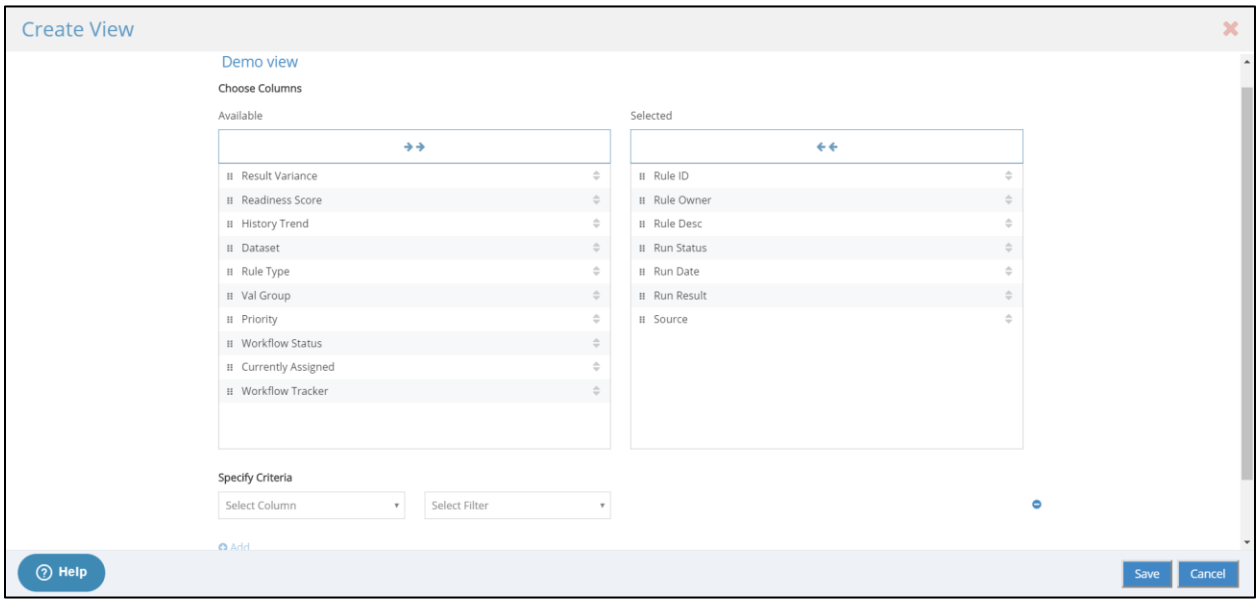


Read more details on [How to Add External Users in Scheduler]

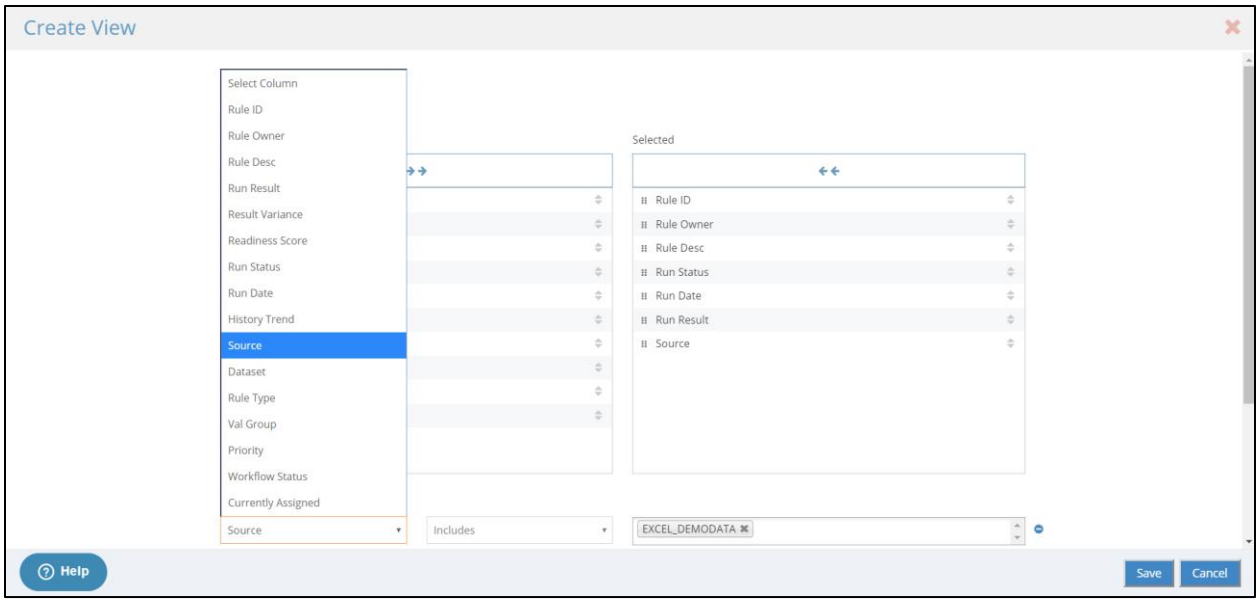
14.11 Governance Views

Governance views allow you to set up and save the rules view according to specific criteria. From Review > Analyze Rules > Analyze Rules drop down > Create View.

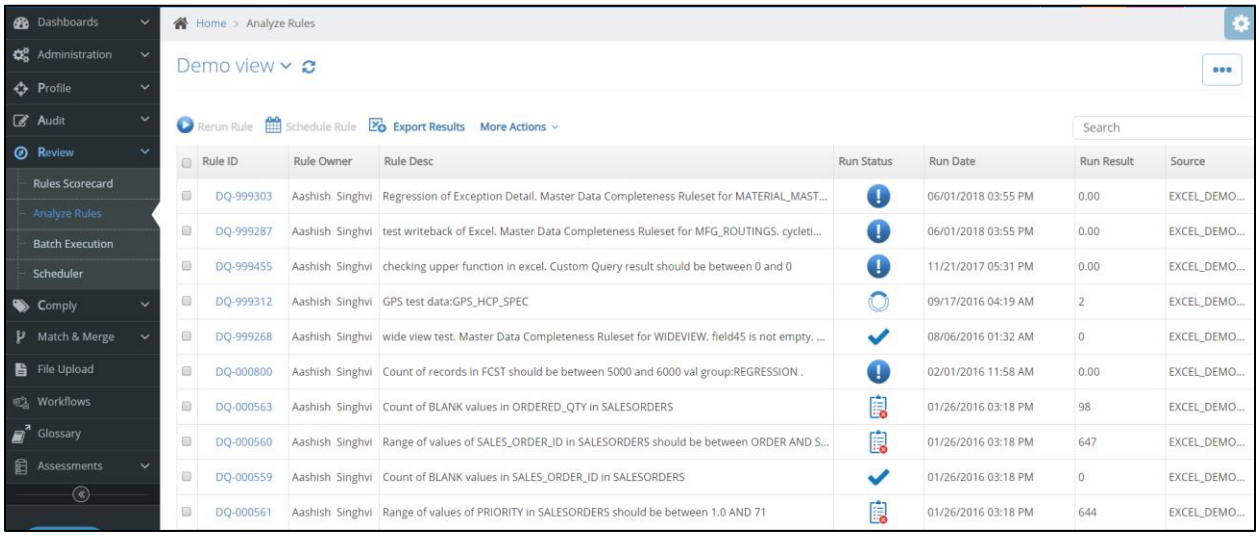




Here you can select the columns that should appear in your custom view and also specify criteria.

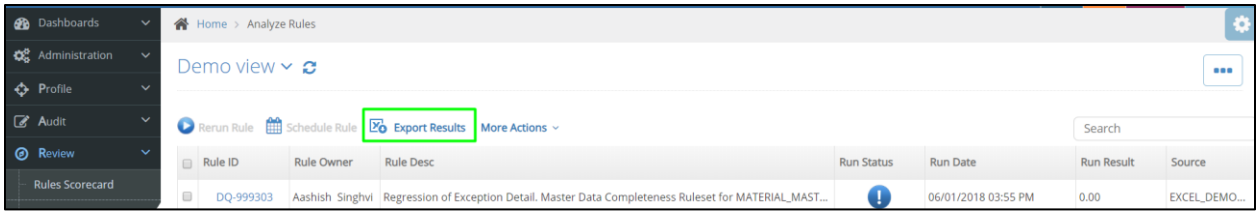


Saving this will create and display the view with the selected columns and specified criteria. You can also specify more than one criterion.



14.12 Export Results

Once a view is created, you can download the excel report based on this view by “Export Results”.



15 Compliance

The end goal of the data quality process is to drive improvements in data quality. The PARC methodology by DvSum enables that process. The compliance section of the solution allows you to monitor the overall quality of data, the performance of the data quality process and drive improvement actions.

The compliance workflow consists of following elements

- Set measurable data quality goals and milestones
- Monitor the trend and performance of data quality goals and identify areas of focus
- Drive accountability and action items to improve performance

15.1 Setting goals

The first step in improving the overall quality of data is to set goals and milestones. You may not know upfront what should be the goals. That is fine. You can define and run an initial data quality audit to baseline the data readiness score. Once you know the baseline, you can establish improvement milestones and align them with existing project milestones or calendar milestones in on-going operations.

An example of project-based milestones that you can configure in DvSum would be:

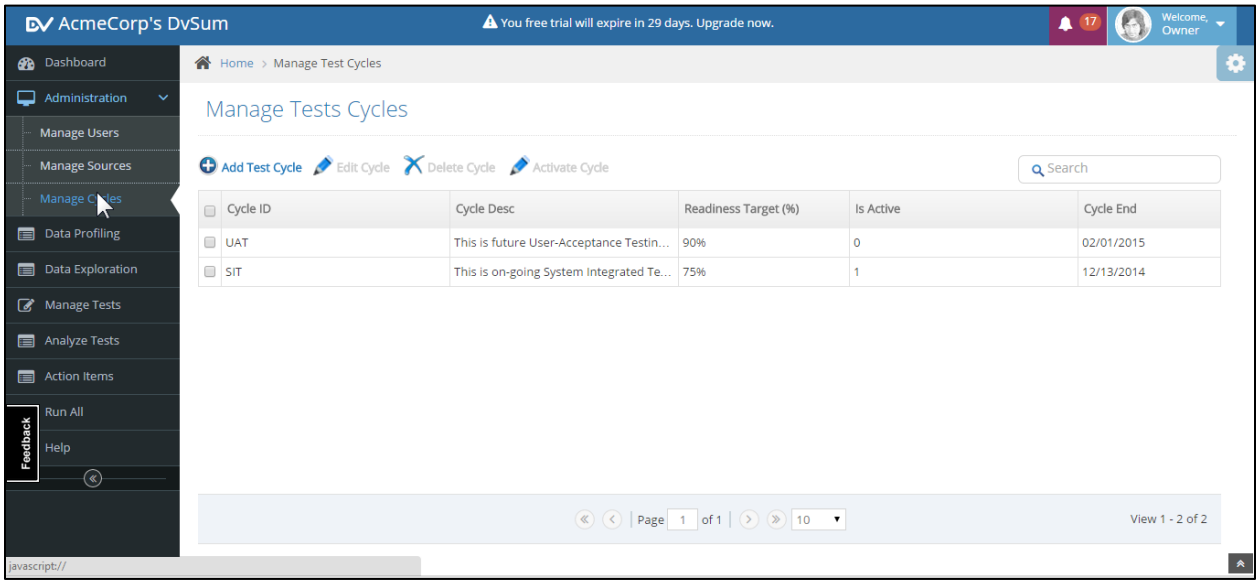
Project Milestone	Scheduled Start Date	Data Readiness Target
Design Phase	4/1/2014	Baseline
Build Phase	6/1/2014	50%
System Integration Testing Phase	8/1/2014	75%
User Acceptance Testing Phase	9/1/2014	90%
Go-live	9/10/2014	95%

An example of calendar-based milestones that you can configure in DvSum would be:

On-going Operations Milestone	Scheduled Start Date	Data Readiness Target
Data Quality Initiative Kick-off	10/1/2014	Baseline
1 Quarter	1/1/2014	65%
2 Quarters	4/1/2014	90%

3 Quarters	7/1/2014	97%
Maintain at		97%

You can set the Data Readiness goals using the Manage Cycles workflow. This is only accessible by owner and administrator roles.

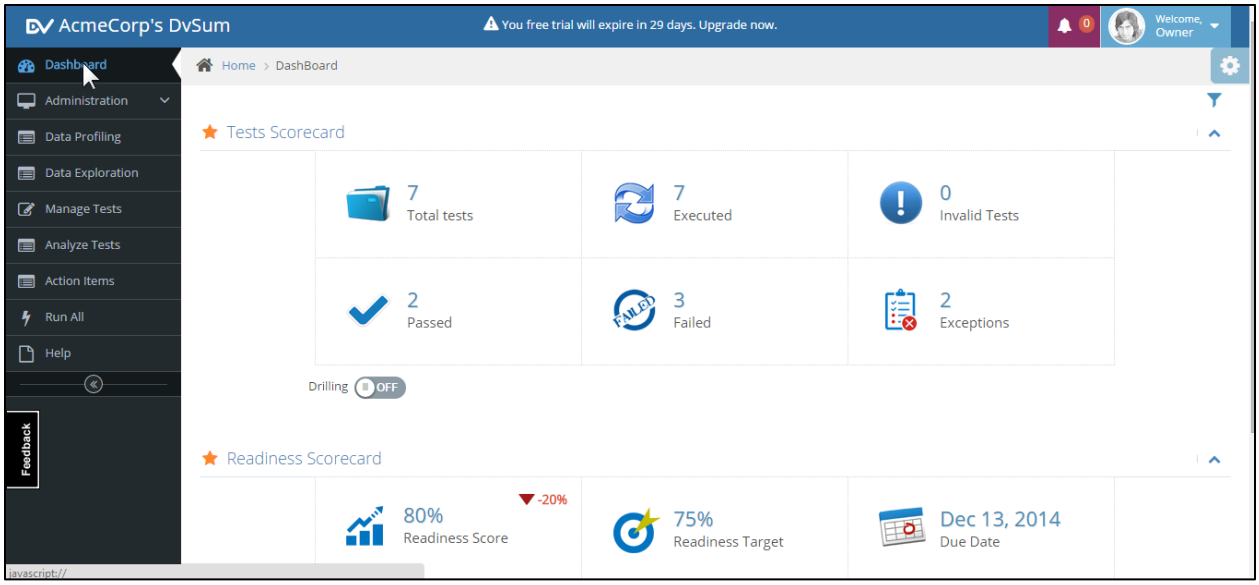


15.2 Monitor Data Quality

Once you have established goals, the next step is to monitor the progress of the quality of data. The DataPARC dashboard is the starting point for reviewing the current status of the data quality. Here, you can see the key metrics in the Tests scorecard and the Readiness Scorecard.

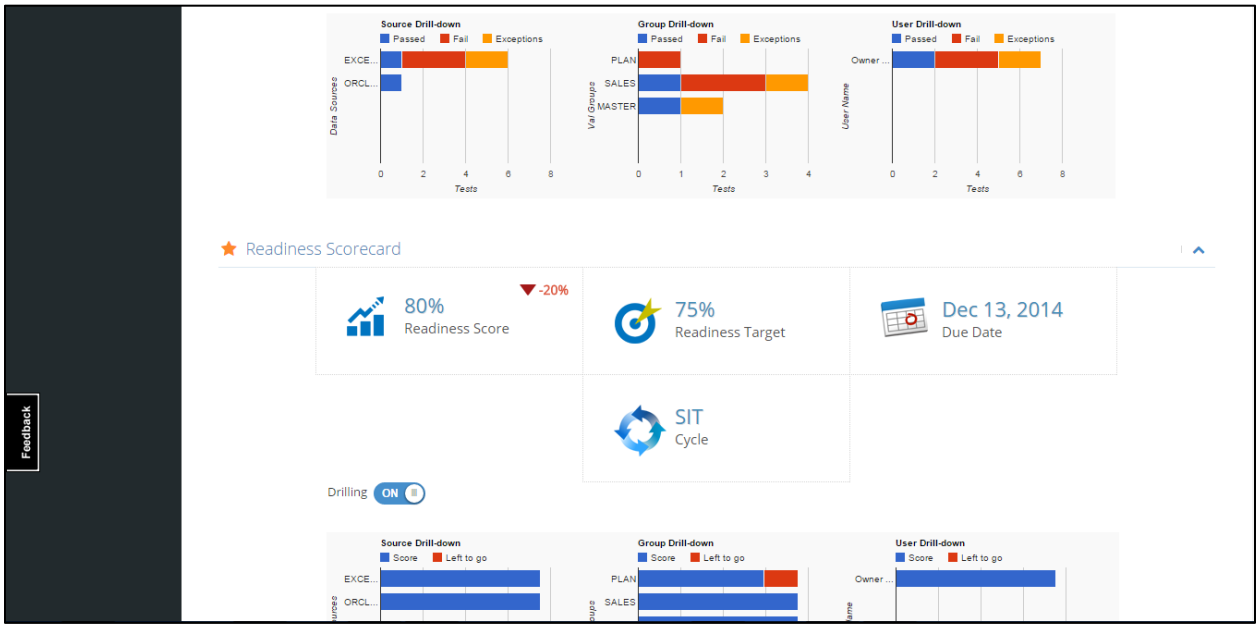
The tests scorecard gives a summary of the total number of audits in the system, how many are passing, how many are failing, how many exceptions.

The Readiness scorecard gives you a summary of the current readiness score, trend of the overall readiness and how it is faring against the currently active target and milestone date.

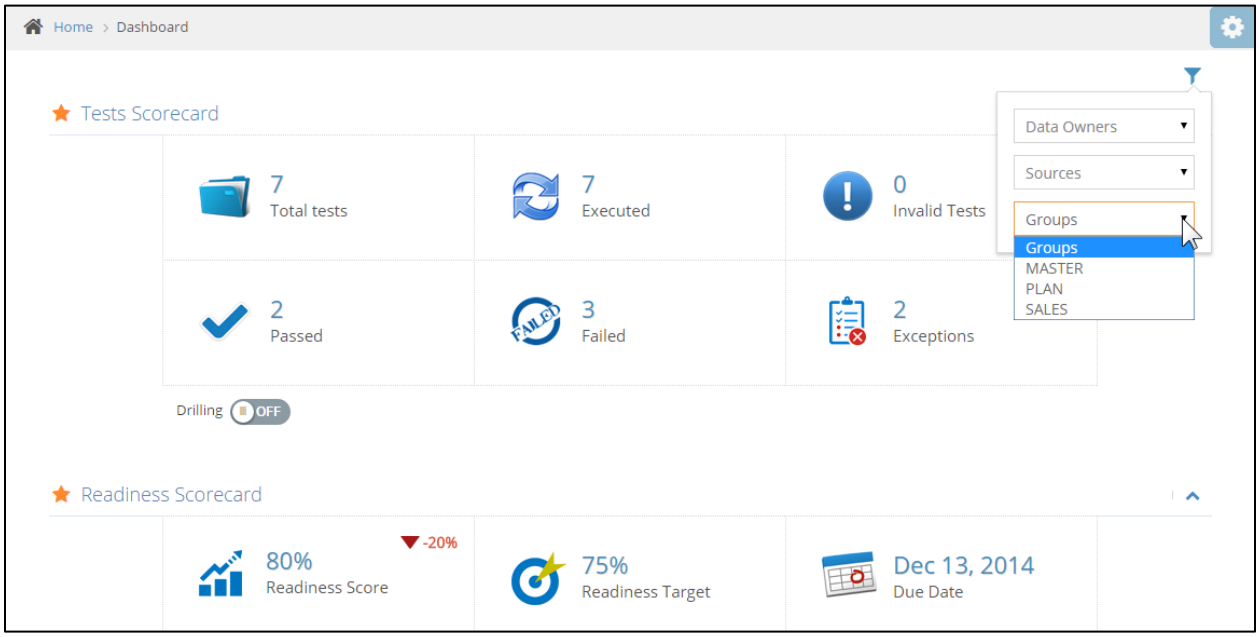


	For owner, administrator and super-user role, the summary information and the detailed audit rules and results visibility is across all data sources and all tests. For user role, the summary and the detailed information is specific to tests that were created by them or currently assigned to them.

You can do a quick drill-down by clicking on the "Drilling buttons on the Tests scorecard or Readiness Scorecard. This gives you a break-down of the tests scorecard or the readiness scorecard by data source, validation group or by user.

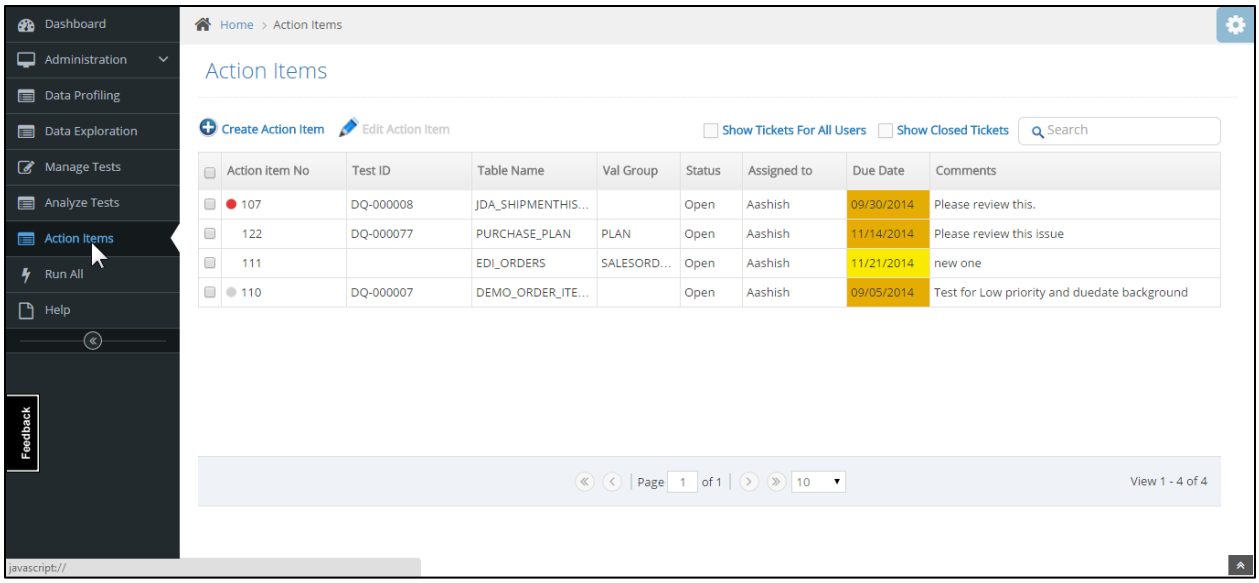


Once you identify the area that requires more focus, you can filter the dashboard and the details for that selection by using the filter on the top right of the dashboard.



15.3 Drive Accountability and action items

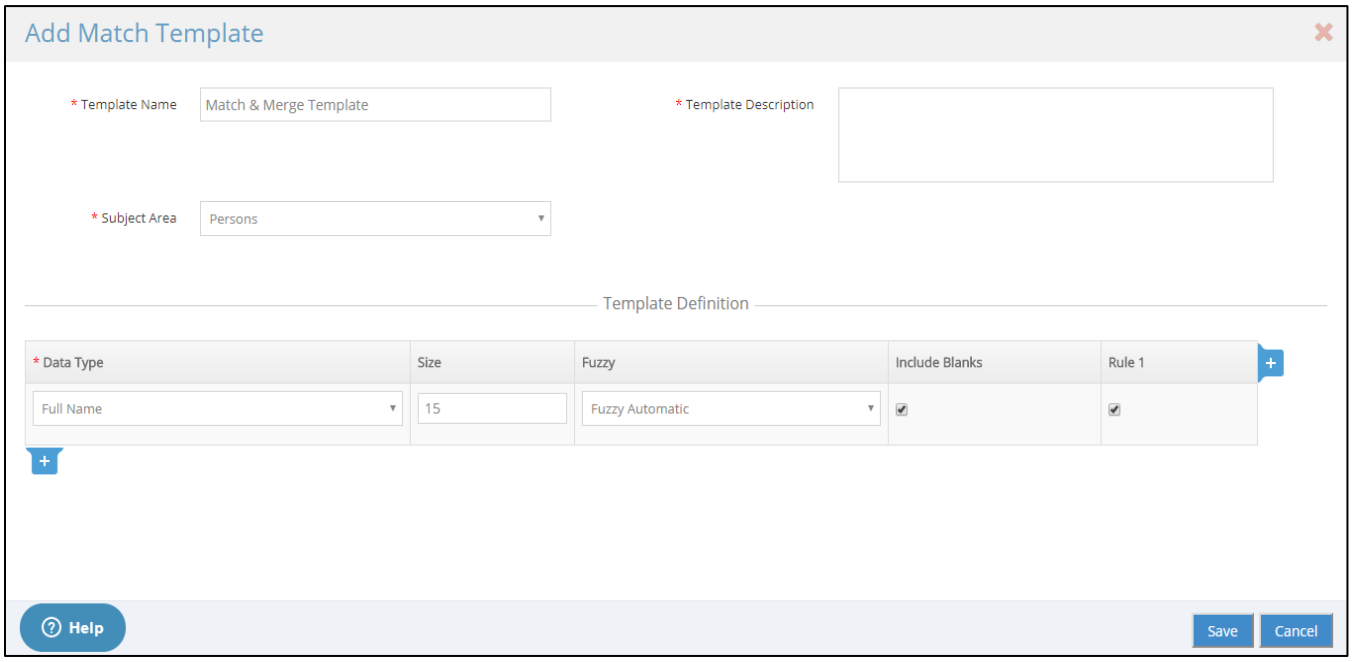
Once you have identified the areas that require focus, you can drive specific actions by using the action item workflow. Action item workflow allows you to easily assign action items related to specific data failures or general data quality process improvements to specific users. The users get notified by email when actions are assigned to them and super-users can monitor the status of action items.



Additionally, you can request a data quality digest to be sent to you which summarizes the data quality activities on a weekly basis. It contains details on activities supporting the data quality trend and performance. Information like how many execution actions were acted upon by different users, how many actions items they worked on, how much time they spent on the data quality tools and analysis. This information can be useful to help guide changes to the areas of focus and priority and tailor it for each user.

16 Match & Merge

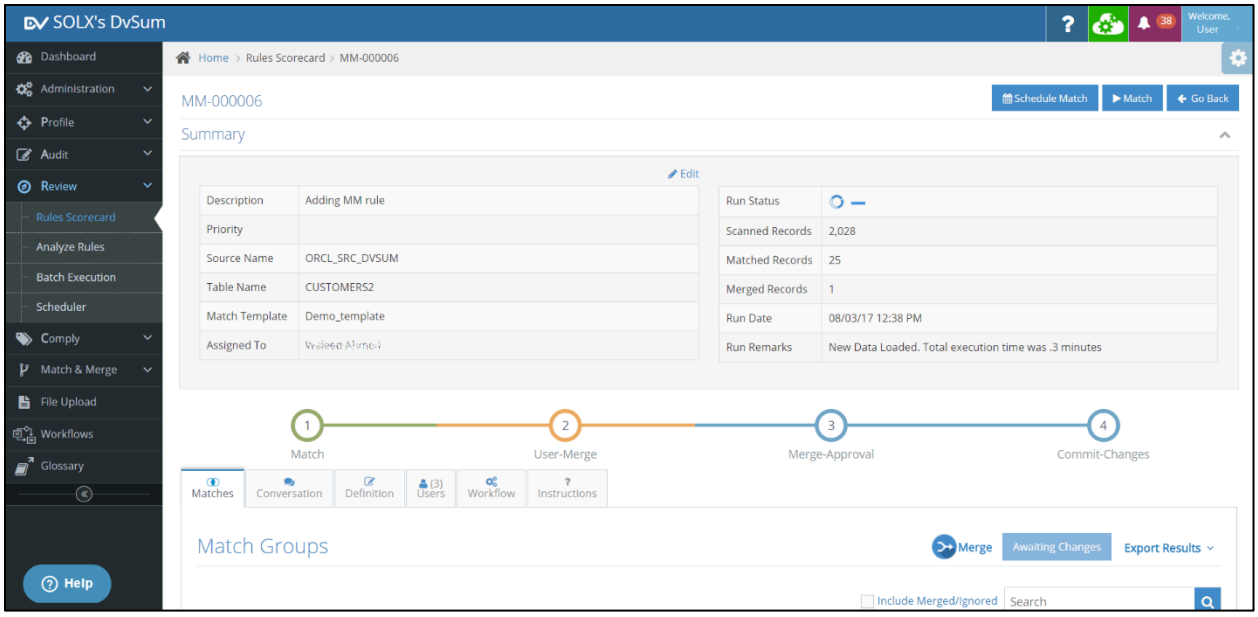
From Match & Merge, you can create a template and you can match your Source Table fields with it. From Match & Merge tab, click on Add Match Template. Select a subject area to define the template. Under the definition you can choose the Data Type, Size, Exact or Fuzzy Logic and the rules.



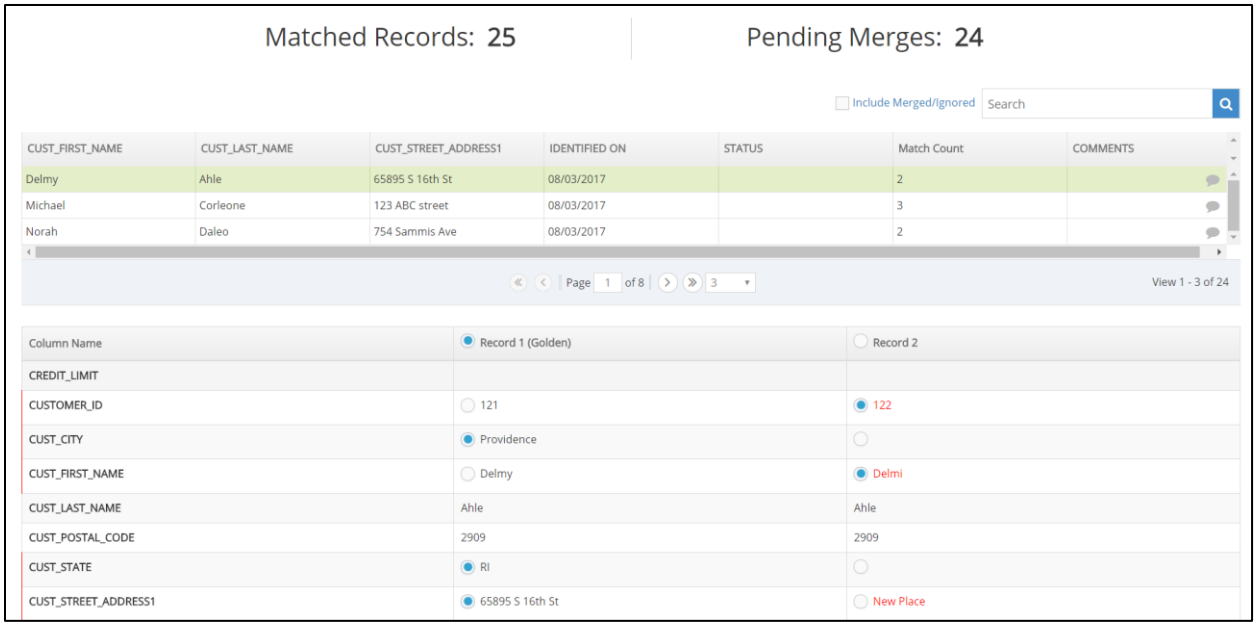
Once the template is created, you can add an audit for it from Review tab >> Rules Scorecard.

Select the Subject area and the Source Table to create a Match rule on. Click on add Rule and select the desired template. This creates a Match and Merge rule.

Execute the rule to see if there are any exceptions. Just like DQ Workflow, you can then create a Match and Merge workflow to cleanse your data.



The person assigned for User-Merge can cleanse data from Matches tab in the MM rule summary. Clicking on Merge opens a cleansing workbench where you can select the values and click on Merge.



The person assigned for approval will approve the changes and commit them to the Data Source.

17 File Upload

File upload requires user to create a folder where a file can be dropped. The idea behind file upload is the reuse-ability of rules on a dataset. In a real-life scenario, a user uploads a file and create a source in DvSum which uses that file. That file is then profiled and certain rules are made and executed on it. The report generated by these rules provides important stats to make day-to-day business decisions.

When the data in the file changes, a user simply uploads the new file to the folder and runs the same rules again. This saves user the time of repeating the process of profiling and creating rules again.

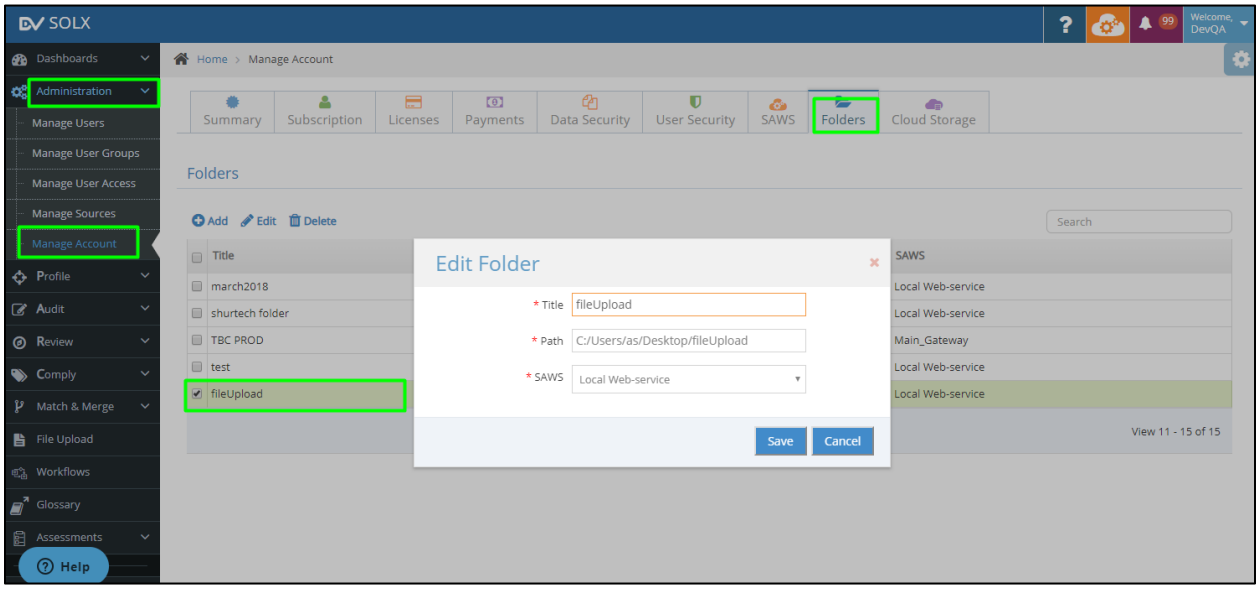
Note: This folder should be in the same machine as the SAWS associated with it.

Follow the steps below to upload file and get started:

Step 1: Create Folder

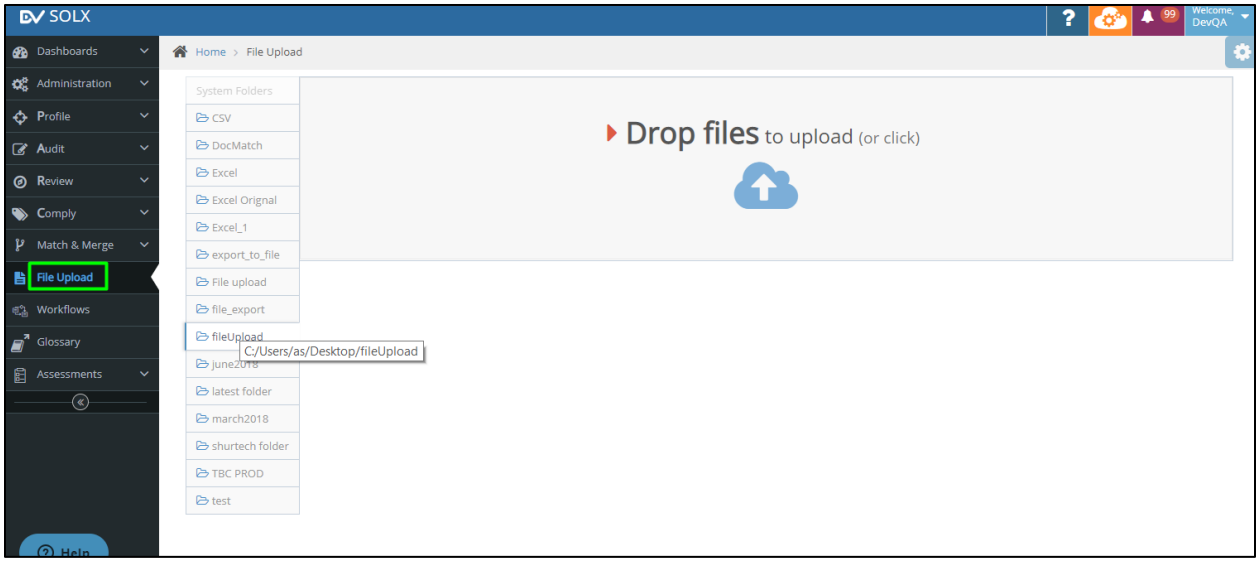
From Administration tab, go to Manage Account > Folders and add a folder. You need to specify folder name, folder path and the associated SAWS.

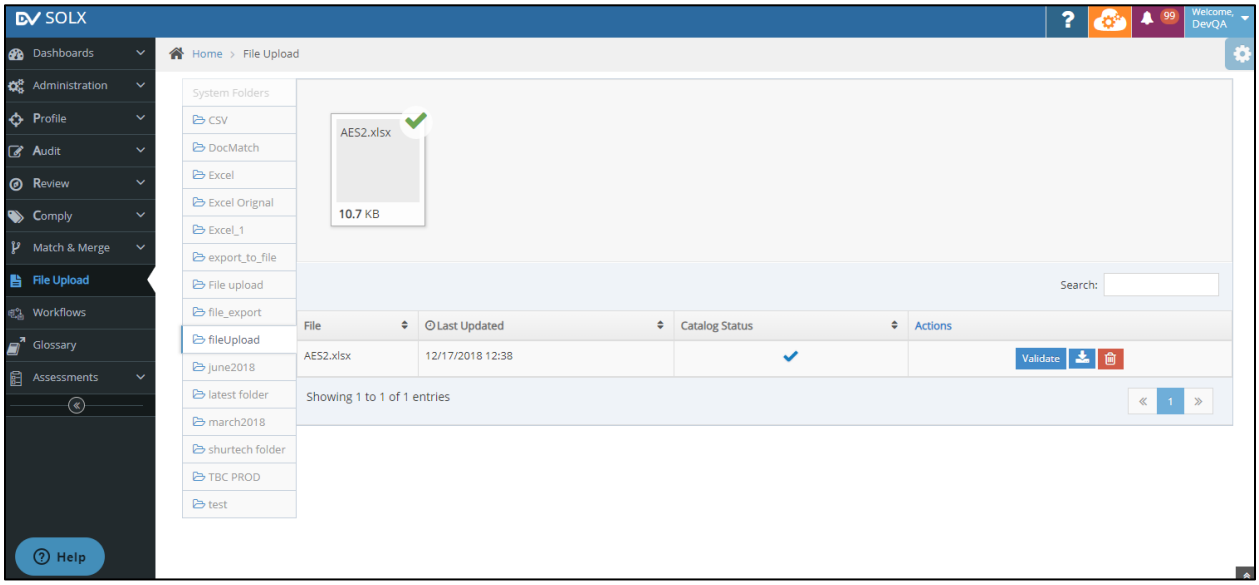
Once a folder is added, it will show on the grid and on the file upload page.

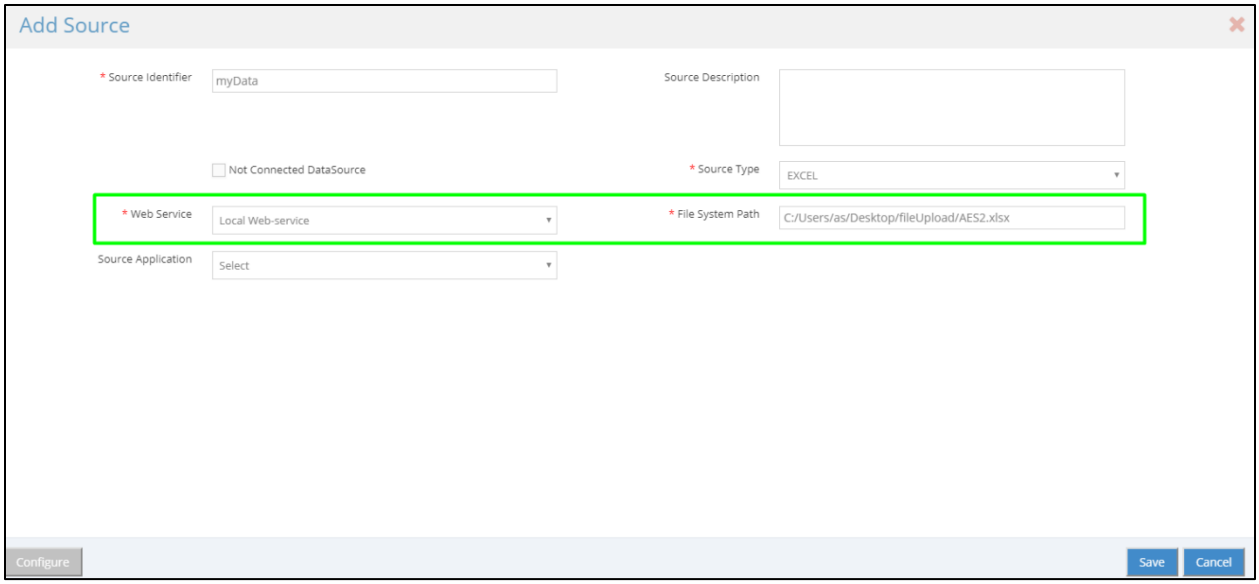


Step 2: Upload File

File Upload page shows the folders added by all users. If you hover over the folder you created, it will show you the path set for it. Simply add the files in drop-zone and they will be uploaded to the folder. (only excel, csv and txt files are allowed)


SOLX
Welcome, DevQA
Dashboards
Administration
Profile
Audit
Review
Comply
Match & Merge
File Upload
Workflows
Glossary
Assessments
Help
Home > File Upload
System Folders
CSV
DocMatch
Excel
Excel Orignal
Excel_1
export_to_file
File upload
file_export
fileUpload
C:/Users/as/Desktop/fileUpload
june2018
latest folder
march2018
shurtech folder
TBC PROD
test
Drop files to upload (or click)


SOLX
Welcome, DevQA
Dashboards
Administration
Profile
Audit
Review
Comply
Match & Merge
File Upload
Workflows
Glossary
Assessments
Help
Home > File Upload
System Folders
CSV
DocMatch
Excel
Excel Orignal
Excel_1
export_to_file
File upload
file_export
fileUpload
june2018
latest folder
march2018
shurtech folder
TBC PROD
test
AES2.xlsx
10.7 KB
Search:
File
Last Updated
Catalog Status
Actions
AES2.xlsx
12/17/2018 12:38
Validate
Showing 1 to 1 of 1 entries
1



Once the file has been uploaded, in order to configure the CSV or Excel File, the user needs to create a new source from Manage Sources. Here the system file path needs to be specified (which is placed in the folder created). After Configuration and Profiling of the newly added source, audits can be executed on it.

- If user uploads the file that already exists (same name), the old file gets replaced by the new one.
- When the user "Validate" a files, DvSum application checks the structure of file with same or similar name. If a similar files exits on which audits have been created and running, then current uploaded file is "Valid" and all audits will execute on the updated/currently uploaded file.

17.1 Cleansing with EXCEL Pre-process Properties

This is an advanced option if a user wants to apply preprocess properties to file in order to cleanse it before creating rules on it.

17.1.1 EXCEL Default Cleansing

When a file is uploaded, it is automatically cleansed to make it compatible for DvSum use. DvSum always reads the first row as "header" and rest of the rows as "data".

The default cleansing includes removing blank rows, replacing white spaces within a column name with underscore e.g. First Column will be changed to First_Column. Special characters from column name are removed. Also, the duplicate columns are dropped.

In some cases, the default cleansing is not enough. Sometimes a user needs to tell DvSum to skip rows which do not contain actual data. Or specify the format for number columns where the data automatically truncates the values up to required decimal precision. For that purpose, a properties file associated with every uploaded file is available for user to edit.

17.1.2 EXCEL Pre-process properties

When an excel file is uploaded to folder, properties file is created with the same name as the uploaded file name. A user can manually set the pre-process options to change file formatting. So whenever the same file is uploaded again to the folder, DvSum will automatically apply the pre-process options and cleanse the file accordingly.

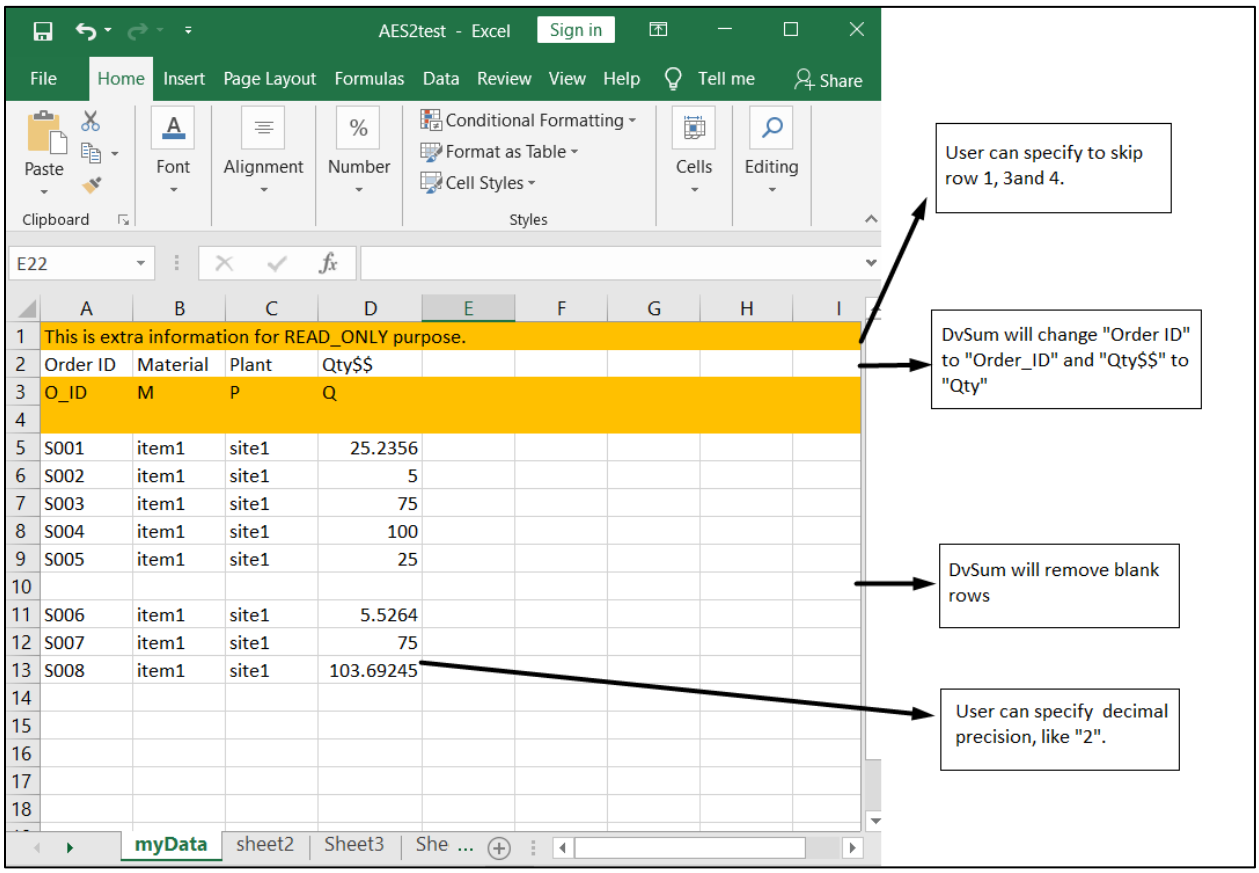
For example, below is the un-processed file.

1. There is some extra information in row 1, 3 and 4 which cannot be read by DvSum. User manually needs to inform DvSum to skip these.

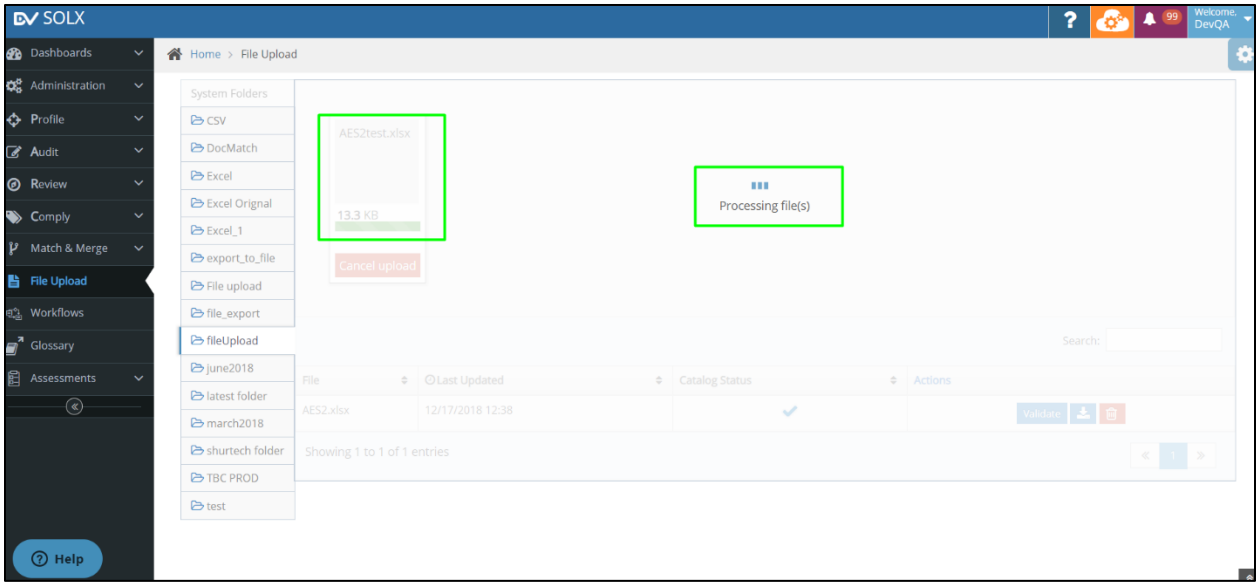
2. Any spaces in between column names will be replaced by underscore and special characters will be removed automatically.

3. Any blanks rows within the data will be removed automatically.

4. For number fields, User can inform DvSum to cleanse the values up to a specific decimal precision.



Once this file is uploaded to folder, basic cleansing (as mentioned before) will be applied by DvSum and pre-process properties for upload file will be created.



Open the pre-process properties file with the name as the original file (AES2test)

AES2test_preProcess.properties - Notepad

File Edit Format View Help

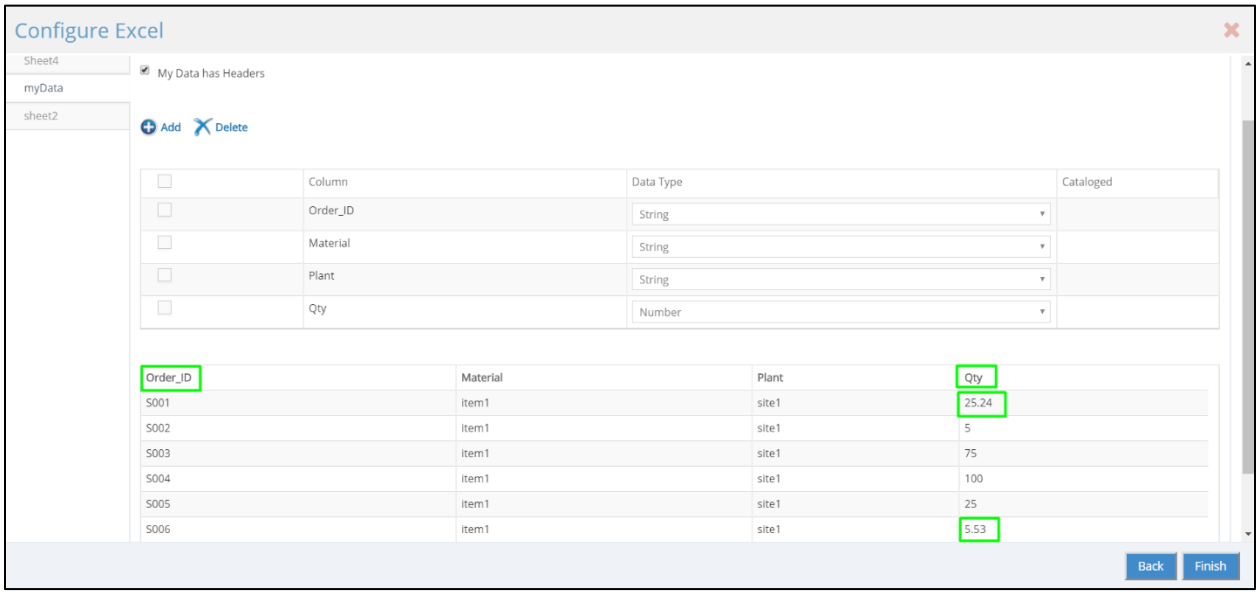
```
{
    "AES2test.xlsx": {
        "preProcess": "True",
        "tabs": {
            "tab_myData": {
                "tabName": "myData"
            },
            "myData": {
                "id": "myData",
                "modifiedName": "myData",
                "skipRows": "1,3,4",
                "columns": [
                    {
                        "originalName": "Order ID",
                        "modifiedName": "Order_ID",
                        "dataType": "object",
                        "index": "0",
                        "decimals": ""
                    },
                    {
                        "originalName": "Material",
                        "modifiedName": "Material",
                        "dataType": "object",
                        "index": "1",
                        "decimals": ""
                    },
                    {
                        "originalName": "Plant",
                        "modifiedName": "Plant",
                        "dataType": "object",
                        "index": "2",
                        "decimals": ""
                    },
                    {
                        "originalName": "Qty$$",
                        "modifiedName": "Qty",
                        "dataType": "object",
                        "index": "3",
                        "decimals": "2"
                    }
                ]
            },
```

Specify rows to be skipped (comma separated or range i.e. 1,2,3 or 1-3)

If column name is to be editted, user can specify here

decimal precision is set as 2

Now the file is cleansed and ready to use in DvSum. If you configure this excel source from Manage Sources page, it will show you the cleansed file.



18 Data Management Work Flows

DvSum DM Workflow is the orchestration and execution of procedures in order to manage process lifecycle. WF tasks are passed from a responsible person/group to another for action according to the needs of an enterprise.

These tasks can be further classified as;

- Simple tasks – Informational tasks
- Validation tasks – DvSum DQ audits execution
- Upload File tasks – File transfer/provision

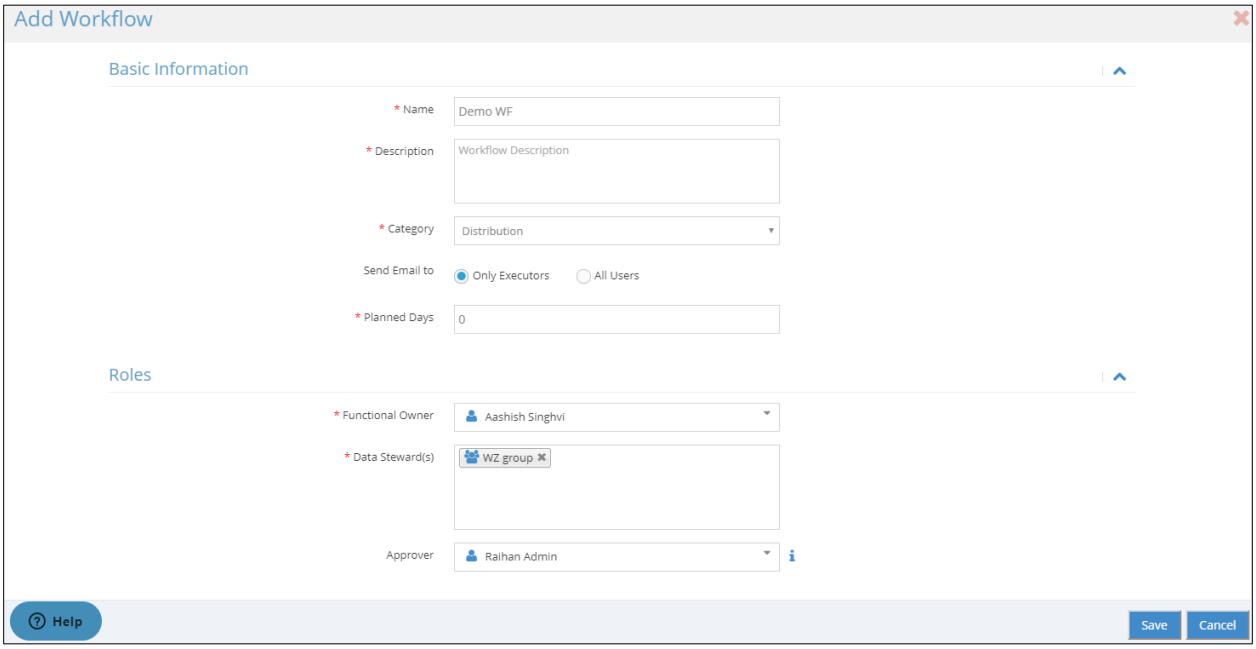
The WF management allows the functional owner to monitor the performance of participants with the goal of improving the efficiency, responsiveness and productivity of a business.

18.1 Manage Categories

Manage Categories tab will be visible to only Administrators and Owners. An admin can add/edit business categories. These categories are used to create workflows for easy access and management.

18.2 Manage Workflows

Manage Workflows tab allows user to create a workflow. In the Add Workflow pop up, among the mandatory fields a user must choose a Category, Functional Owner and Data Stewards. Approver (optional) which can be a single user or a user group. Adding watchers is also optional.



18.3 Workflow Designer Window

Once the Workflow is created successfully, it is added in the "All Workflows" list and the user is navigated to the Workflow Designer Page. A user can create functions/activity for a workflow here.

There are 4 types of available boxes:

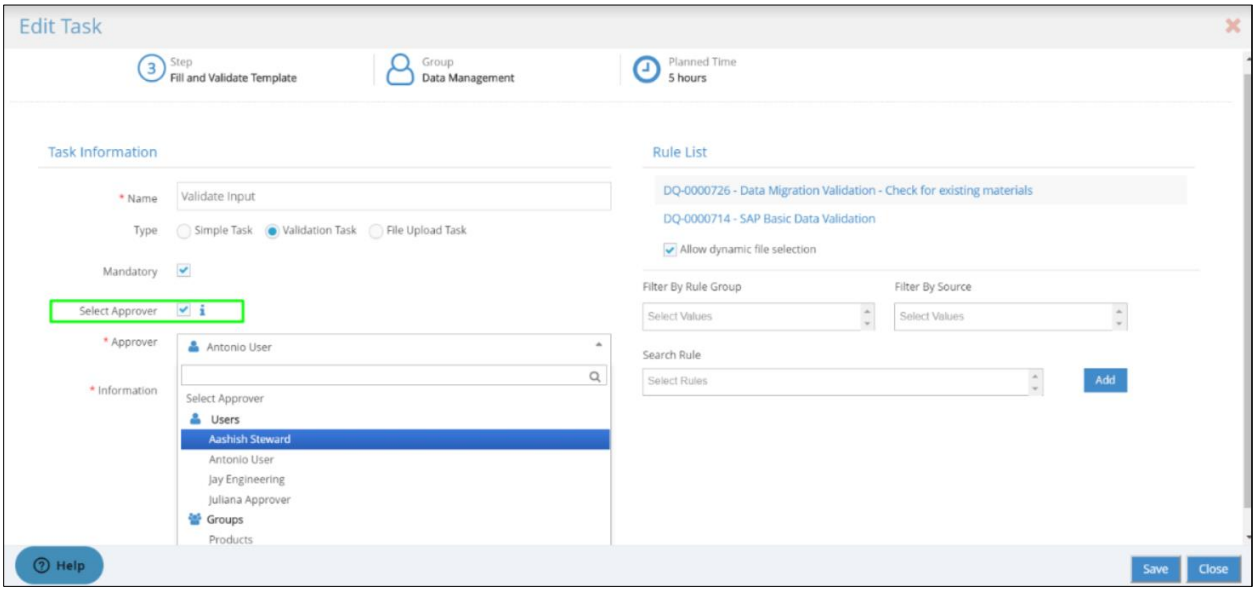
1. Terminator – is an oval box used to define the "Start" and "End" of a workflow. There must be 2 terminators in a workflow
2. Process – is a rectangular box used to define the activities or tasks. Process details window allows a user to Add Task of the following type:

   2.1 Simple Task – The user can provide contextual information about the required task to be completed

   2.2 Validation Task – Data Quality Audits can be added that are associated with the WF. If there are exceptions after DQ rule execution, the task is considered as failed. If the audits pass, then the task can be marked as complete.

   2.3 Upload Task – User is required to upload a specific named file to the folder that is set by the administrator

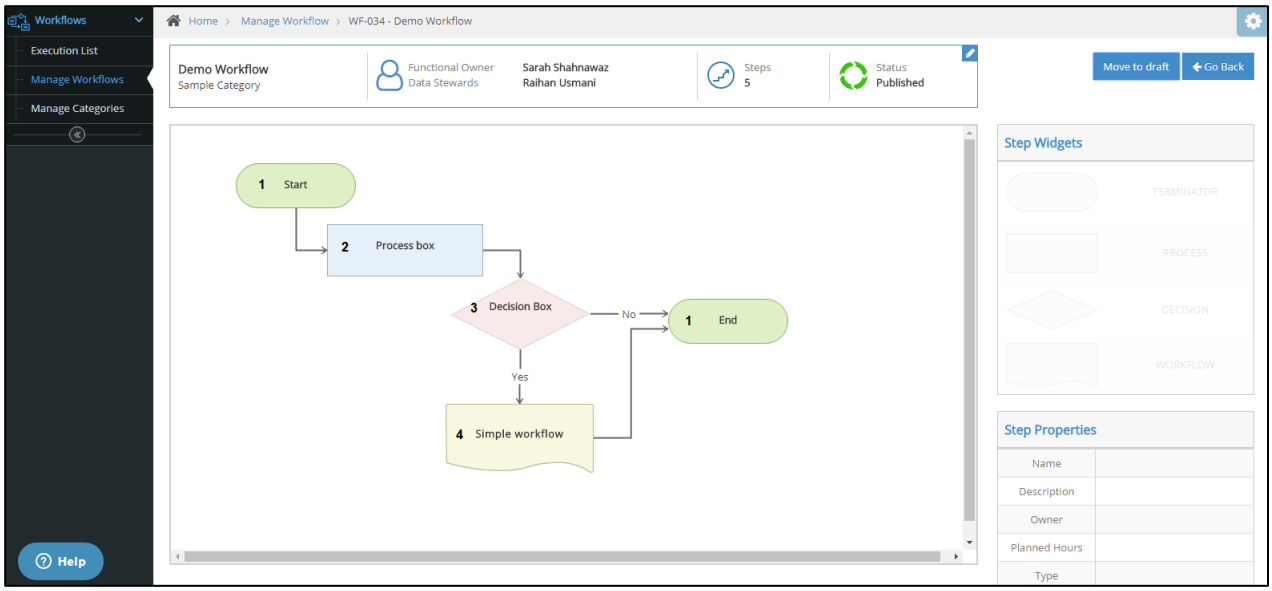
Note: You can mark all tasks as mandatory. But for Validation tasks, if they are not completed, user will assign the step to approver. Approver decides whether to mark step as complete or not.



Validation Task with DQ rules list and an Approver

3. Decision – is a diamond shaped box used where a decision is necessary (Yes/No). There are two arrows corresponding to "Yes" and "No" respectively. Depending on the decision output the next process is executed.

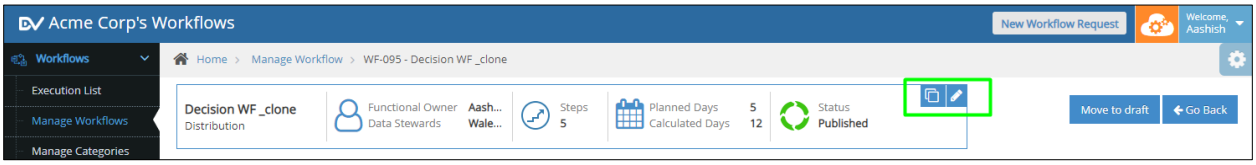
4. Workflow – Is a box shaped as a document used to embed a workflow within a workflow. This will include all the steps of the embedded workflow as a part of the new workflow.



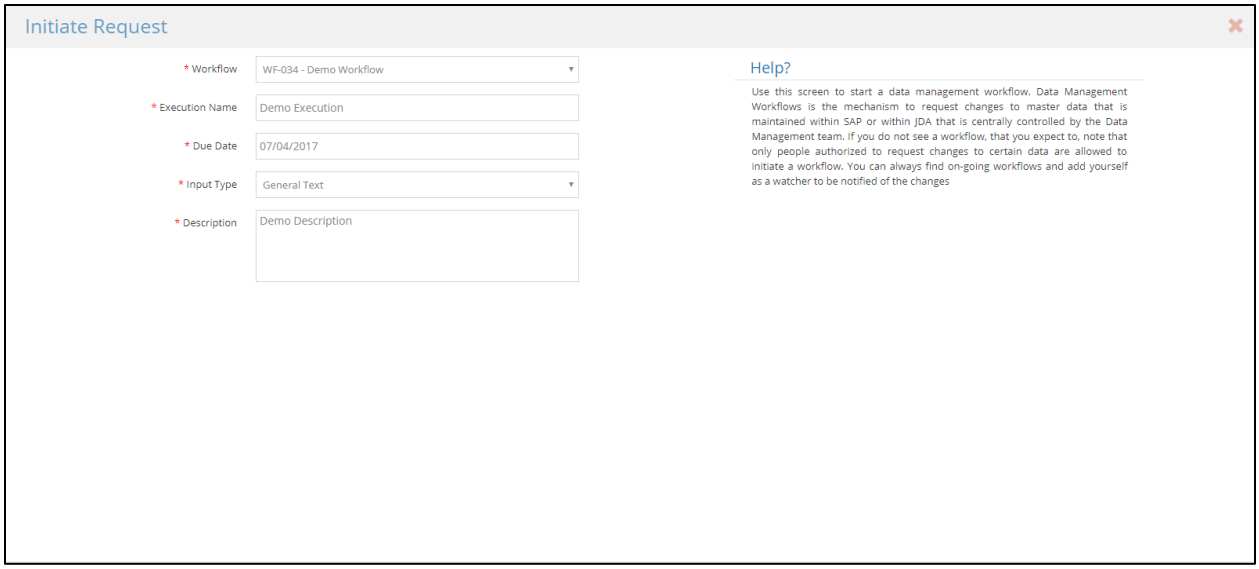
Workflow in Published State after the tasks are defined

The boxes are connected together via an arrow which represents the direction of process flow. After the Workflow content is complete, it is to be moved in the "Publish" state for the WF processes to be triggered on execution (as shown in the figure above).

Note: You can edit or clone a workflow from the header appearing at the top of designer window



In order to execute a workflow, WF request has to be initiated. After the user provides an Execution Name and Due Date, a confirmation message notifies the beginning of a WF.

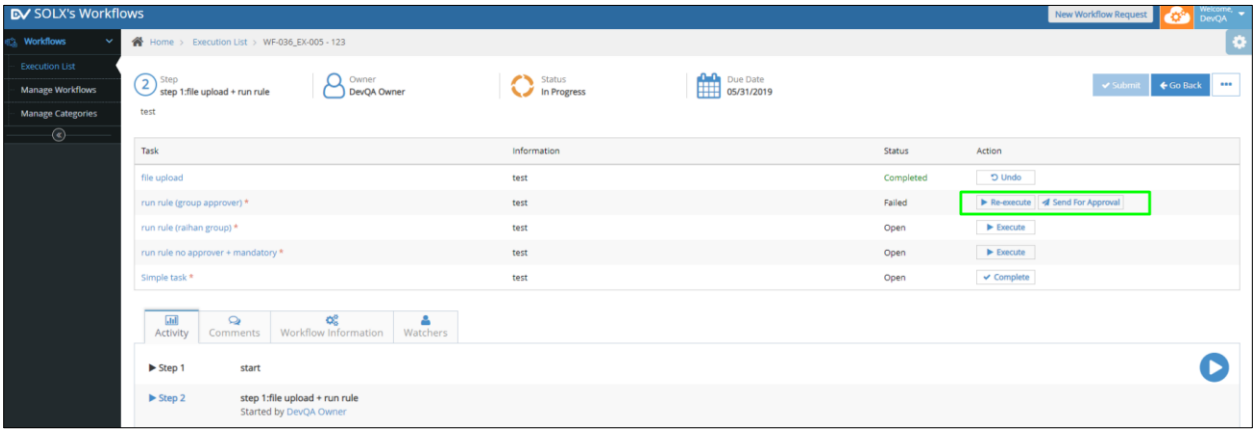


18.4 Execution List

The execution window shows the activity of a WF and the user assigned to perform the certain task over a Step.

A Step can have following Status:

- Open – When the task is awaiting user acceptance
- In Progress – When the user "Start Progress" on a Step and needs to perform the required action according to the type of task i.e. "Execute" and "Complete" a task.
- Pending for Approval – This option is only for mandatory Validation task. If the task fails and there are exceptions and a user requires to move forward without fixing all exceptions, he can ask for "Approvers" permission.

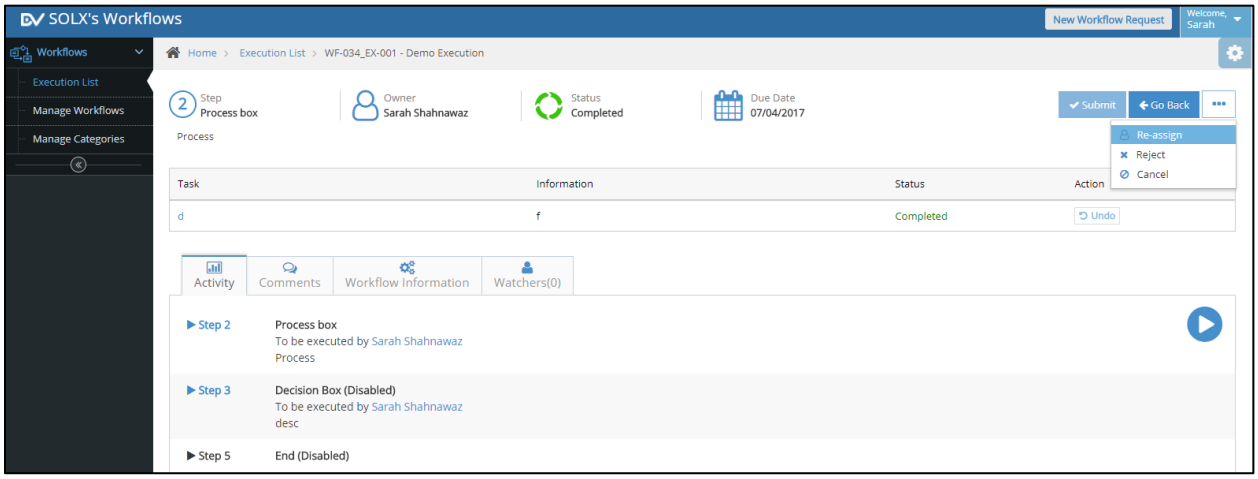


- Completed – After completing the task, "Submit" button completes a Step

There are three more actions a user can perform over a Step:

- Re-assign – When a user wants to reassign the Step responsibilities to another participant
- Reject – When a user wants to reject a workflow progress up to a certain step

- Cancel – When a user wants to cancel the workflow execution



Workflow with active execution showing Step Owner, Task State and Step state

18.5 Adding Watchers

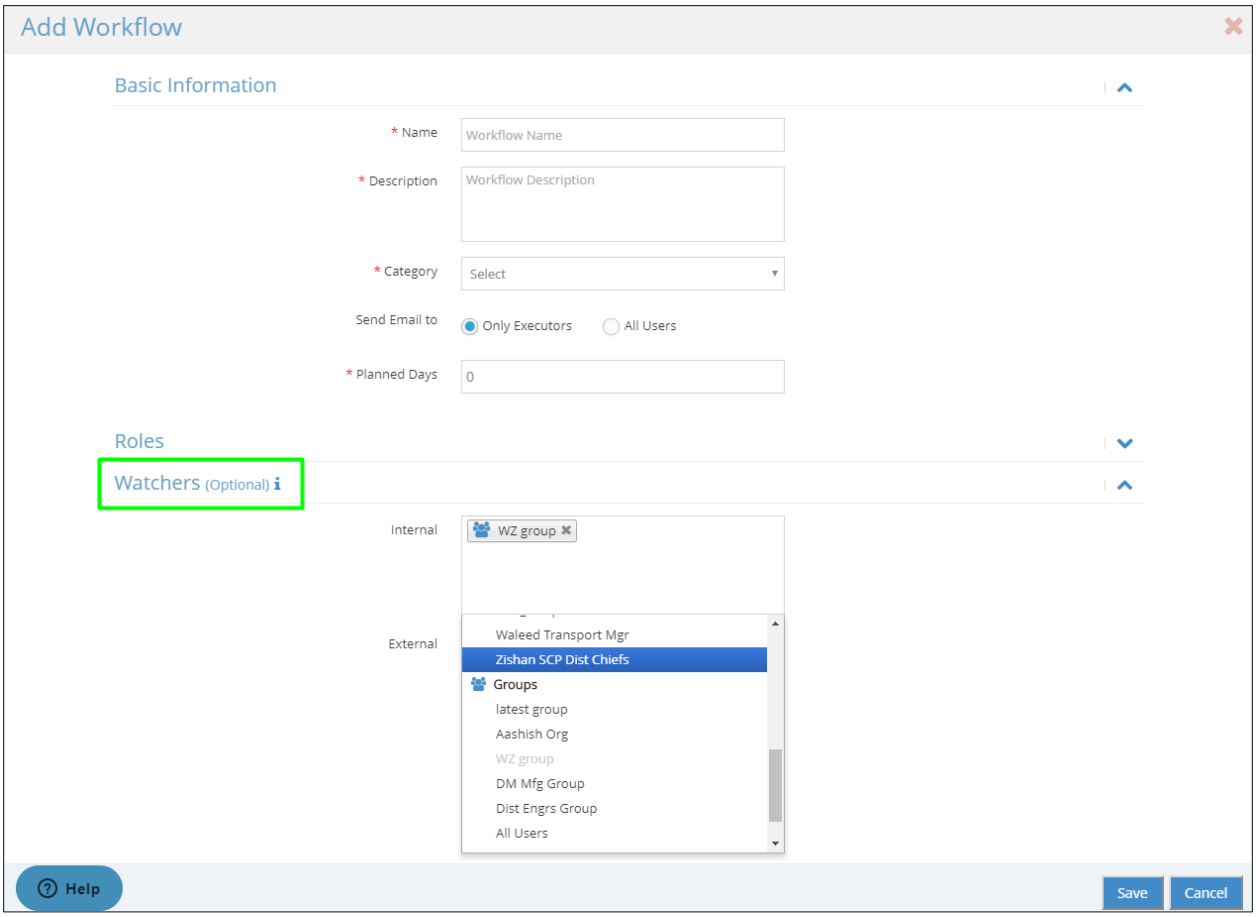
Any user in DvSum, who is not participating in a workflow execution can be added as a "watcher". A watcher receives all emails of an execution to monitor the WF activity. It could be multiple users or even User Groups.

You can also add non-registered users as watchers too. These are added as External Users in DvSum application. External Users cannot participate or log on to DvSum. They will only receive all workflow execution emails to keep them informed about the progress of WF.

Watchers can be added from 2 pages;

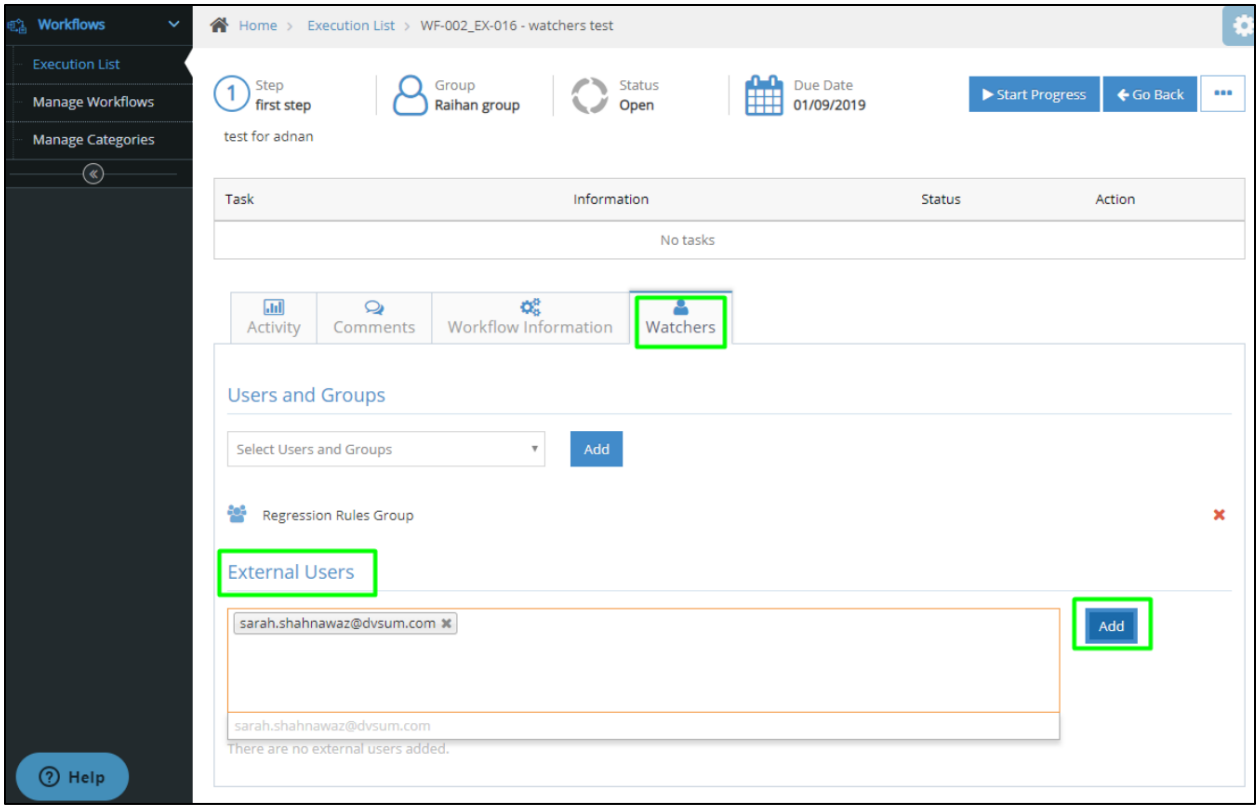
18.5.1 Add Watchers while creating a Workflow

When you create a new WF from Manage Workflows, you can see the section for watchers at the bottom of page



18.5.2 Add Watchers from Workflow Execution

Watchers can also be added from Watchers tab on WF execution detail page. “Users and Groups” section will show DvSum users and the groups.



"External User" section will allow you to add non-registered users as watchers. Note that their email domains should be white-listed before adding them.

Read more details on [How to Add External Watchers in WF Execution]

18.6 Workflow Alerts

When there are group of people involved in a Workflow, it is hard to keep track of WF step progress. We have added automatic alerts which notify users via email;

a. When a workflow execution starts, initiator and first step owner will receive email.

b. When a workflow step is completed, email alert goes to the person who completed current step and the person to whom next step is assigned to.

c. When a workflow execution is completed, email alert goes to the person who completed last step and to initiator of workflow.

d. When a workflow step is rejected up to a certain previous step, both the step owners are notified (rejected-from-step owner and rejected-to-step owner)

e. When a workflow step is re-assigned to someone, both the step owners are notified. (ex-step owner and new owner)

f. When a workflow execution is cancelled, initiator is notified.

g. Daily Email alert when Workflow execution completion is suspected to cross it's due date or has already crossed the due date because of a delay in 1 or more steps.

Steps information in the Workflow detail page now shows you when steps were open, when work was started, what is the expected completion and when it actually got completed.

Note: "Watchers" of a workflow which can be external users as well, receive all notification emails mentioned above.

Learn more on how to control Workflow Email Alerts and Workflow Digest Emails from User Profile page.

19 Business Glossary

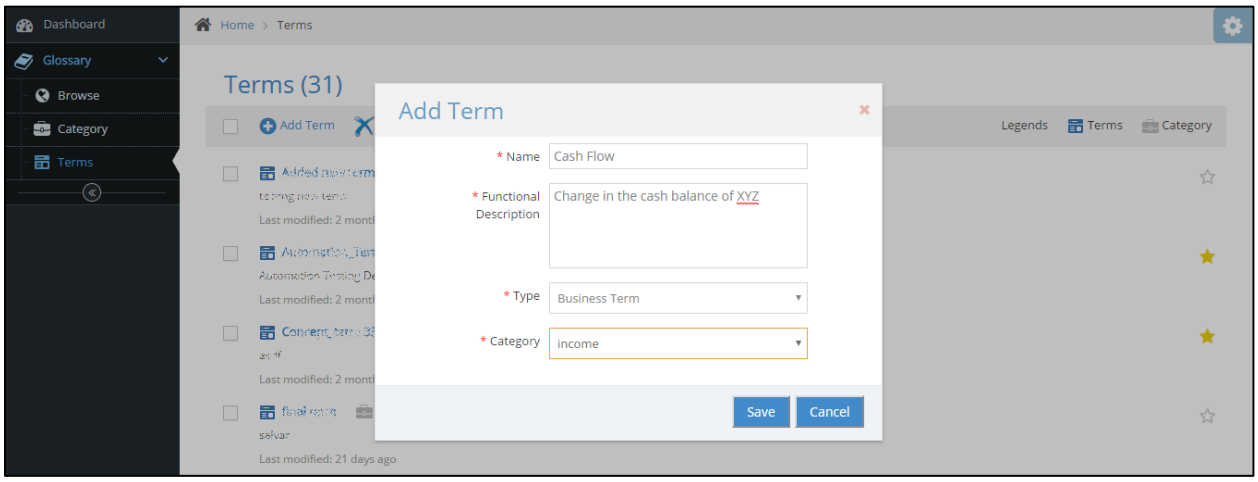
DvSum allows data stewards to create a vocabulary framework of business concepts and terminologies along with their associations (synonyms and related terms). This common shared business vocabulary enhances data governance, data quality and helps manage metadata. Metadata management is the significant input to capturing an enterprise data flow and representing *"data lineage"*.

19.1 Glossary Dashboard

Here you can see the over view of your organization's business glossary. The dashboard tells you recently viewed terms by you, list of data stewards, total number of terms, categories and data sources.

19.2 Terms

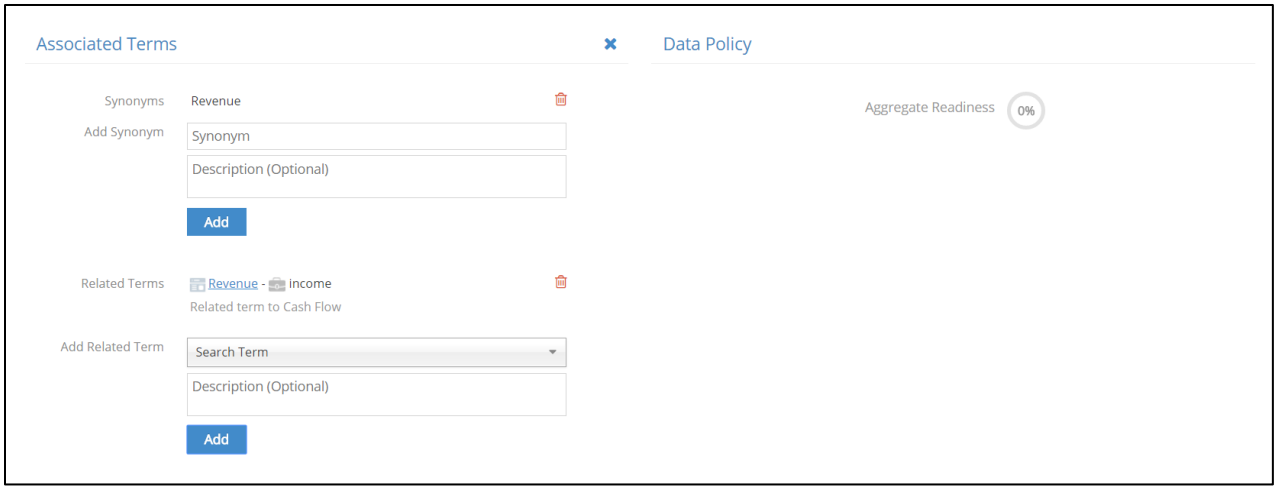
A user can "Add a term" from the dashboard or from Terms Page. Under the Type drop down, you can choose the term to be added as "Concept" or "Business Term" and select the required category. Categories for terms can only be defined by DvSum Administrators.



Once the term is created, a user is navigated to detailed view for the term where additional information like synonyms, related terms, roles can be added. For the Business Term, two additional fields are visible i.e. Data Policy and Assigned Assets. Assigned assets is the section from where user can create Data Lineage flowcharts.

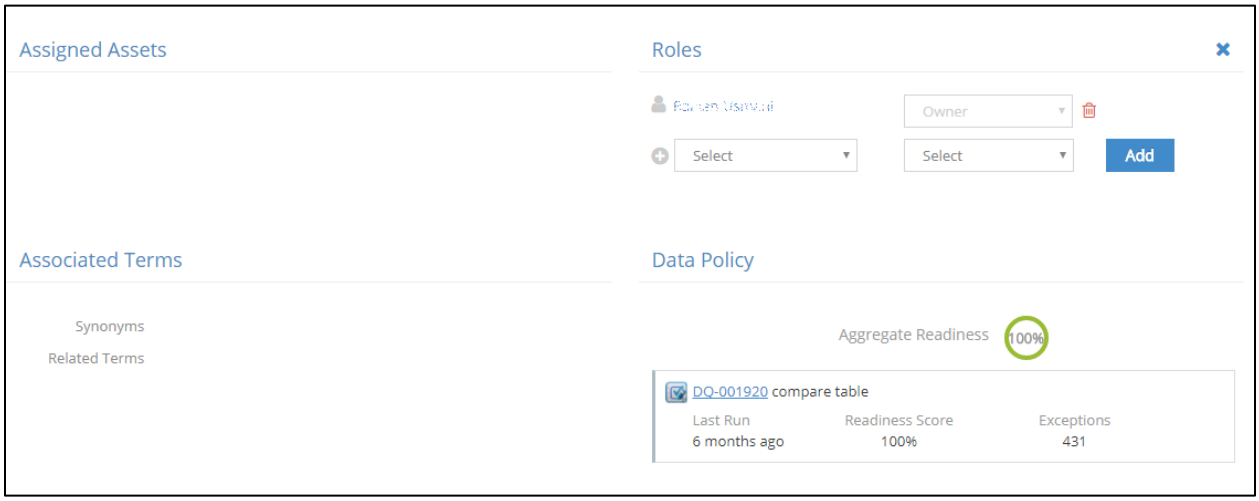
19.2.1 Associated Terms

Let's say we for a newly added Business Term "Cash Flow", "Revenue" can be added as the related term as shown in the figure below.



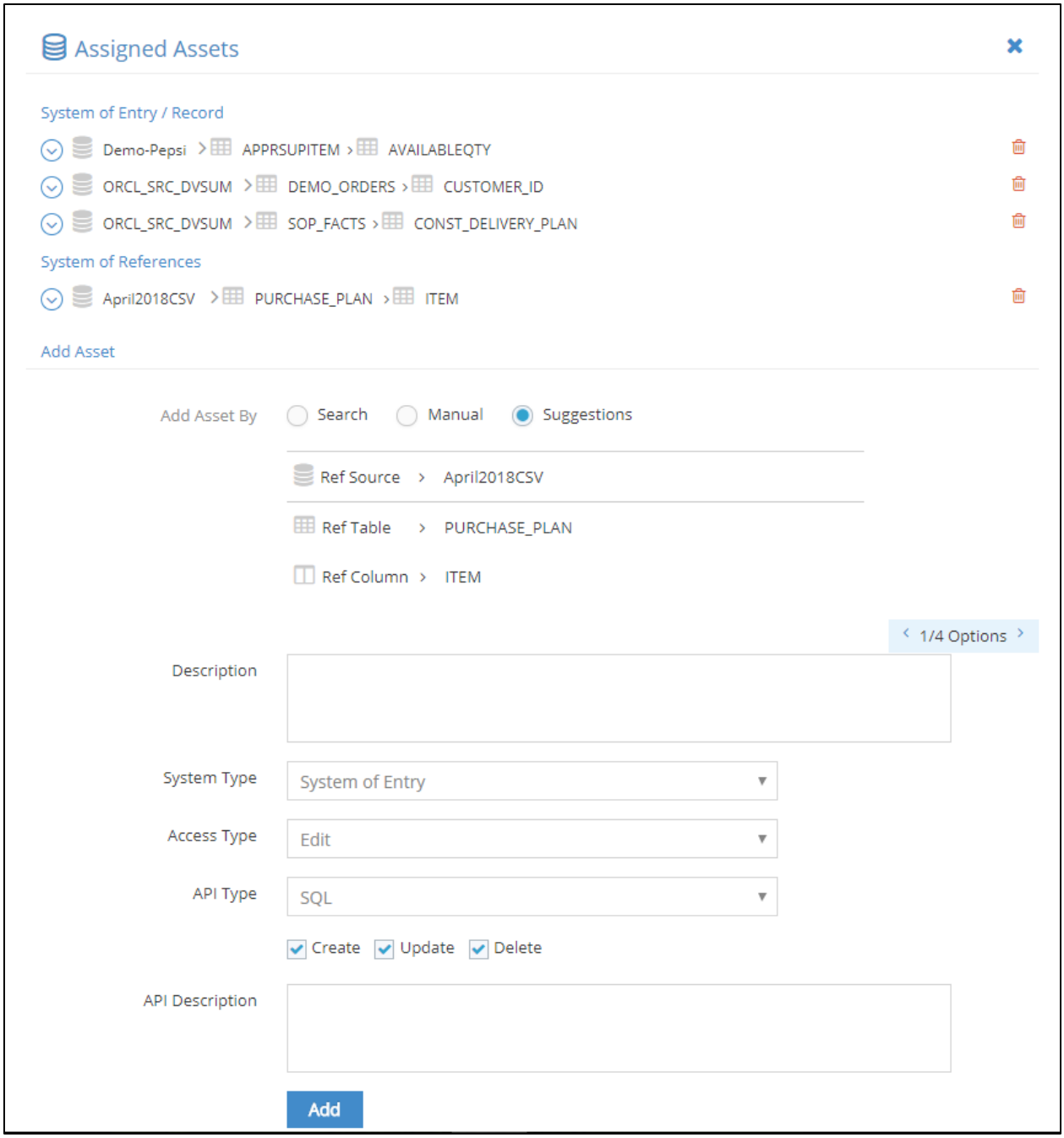
19.2.2 Roles and Data Policy

Under the Roles section, you can assign the roles of Owner or Data stewards to users. Under the Data Policy section, the DQ audits defined in the DataPARC can be associated with a or you can create a policy of your own by adding textual data alone.



19.2.3 Assigned Assets

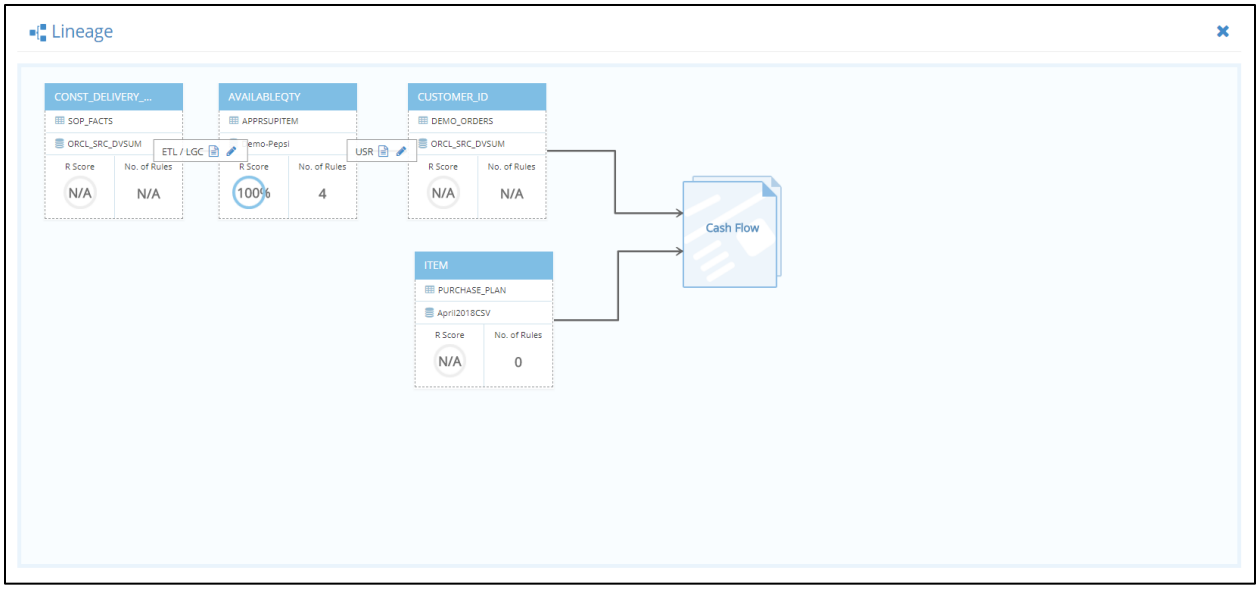
Under the Assigned assets section, add the assets. You can define System of Entry / Record and System of References here. The Assets added will be shown below in the lineage designer window.



19.2.4 Data Lineage

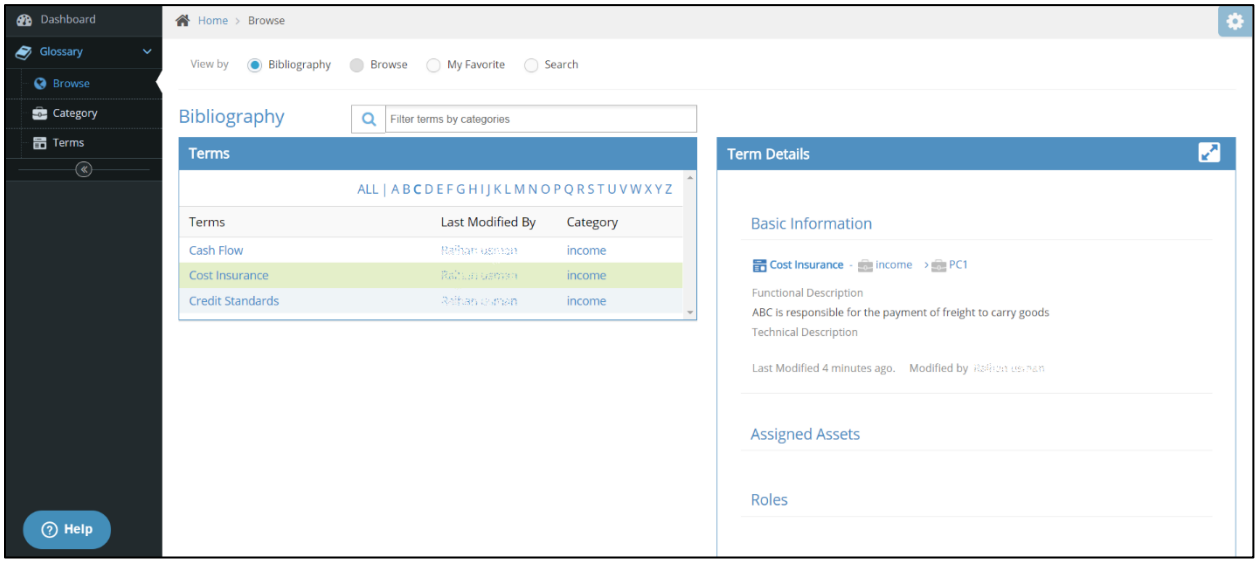
The data flow or data life cycle of a process within your organization can be illustrated by Data Lineage Flowchart.

In Data Lineage designer window, the “ETL” or “User Process” relation can be defined when the data flow connections are added between assets. These data flow connections represent the data flow of your organizational process from source to destination.



19.3 Browse

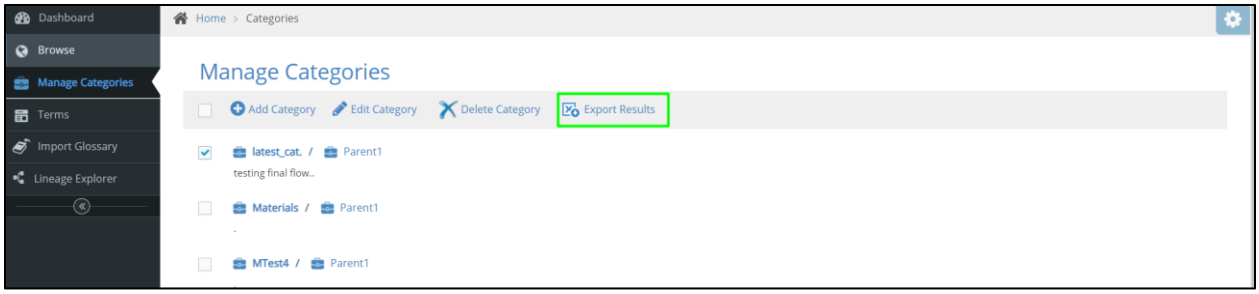
Since it's a common business vocabulary, DvSum provides the biblographic view of an organization's glossary so the terms are easily accessed by all users.



19.4 Glossary Categories

DvSum Administrators can add Glossary Categories for the terms. A category has to be defined under a parent category. When you add a term, you need to specify under which category this term will belong. Hence you can categorize your glossary.

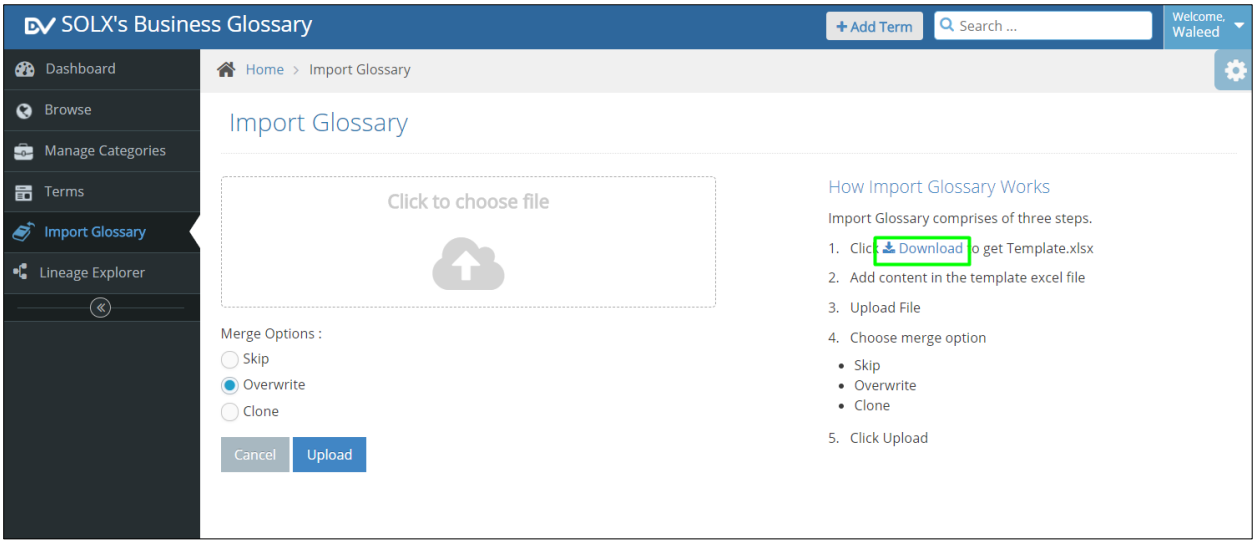
Another feature DvSum provides is of "Export" category. This downloads excel file containing all terms within the downloaded category.



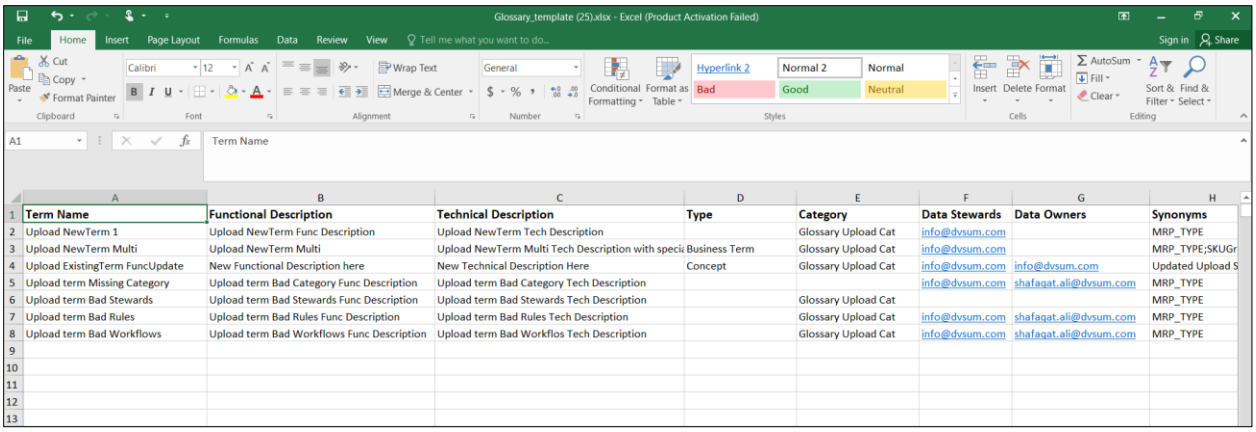
19.5 Import Glossary

Import Glossary allows you to add terms to DvSum Glossary in bulk. Here is a step-by-step tutorial to help you out.

Step 1: Download the "Glossary Template" from Import Glossary tab.



Step 2: Open the Excel file that is downloaded and add terms with their details. You must specify the term name, functional description and term category otherwise the term will be considered invalid.



Term Name	Functional Description	Technical Description	Type	Category	Data Stewards	Data Owners	Synonyms
Upload NewTerm 1	Upload NewTerm Func Description	Upload NewTerm Tech Description		Glossary Upload Cat	info@dvsum.com		MRP_TYPE
Upload NewTerm Multi	Upload NewTerm Multi	Upload NewTerm Multi Tech Description with speci	Business Term	Glossary Upload Cat	info@dvsum.com		MRP_TYPE;SKUGr
Upload ExistingTerm FuncUpdate	New Functional Description here	New Technical Description Here	Concept	Glossary Upload Cat	info@dvsum.com	info@dvsum.com	Updated Upload S
Upload term Missing Category	Upload term Bad Category Func Description	Upload term Bad Category Tech Description			info@dvsum.com	shafaqat.ali@dvsum.com	MRP_TYPE
Upload term Bad Stewards	Upload term Bad Stewards Func Description	Upload term Bad Stewards Tech Description		Glossary Upload Cat			MRP_TYPE
Upload term Bad Rules	Upload term Bad Rules Func Description	Upload term Bad Rules Tech Description		Glossary Upload Cat	info@dvsum.com	shafaqat.ali@dvsum.com	MRP_TYPE
Upload term Bad Workflows	Upload term Bad Workflows Func Description	Upload term Bad Workflos Tech Description		Glossary Upload Cat	info@dvsum.com	shafaqat.ali@dvsum.com	MRP_TYPE

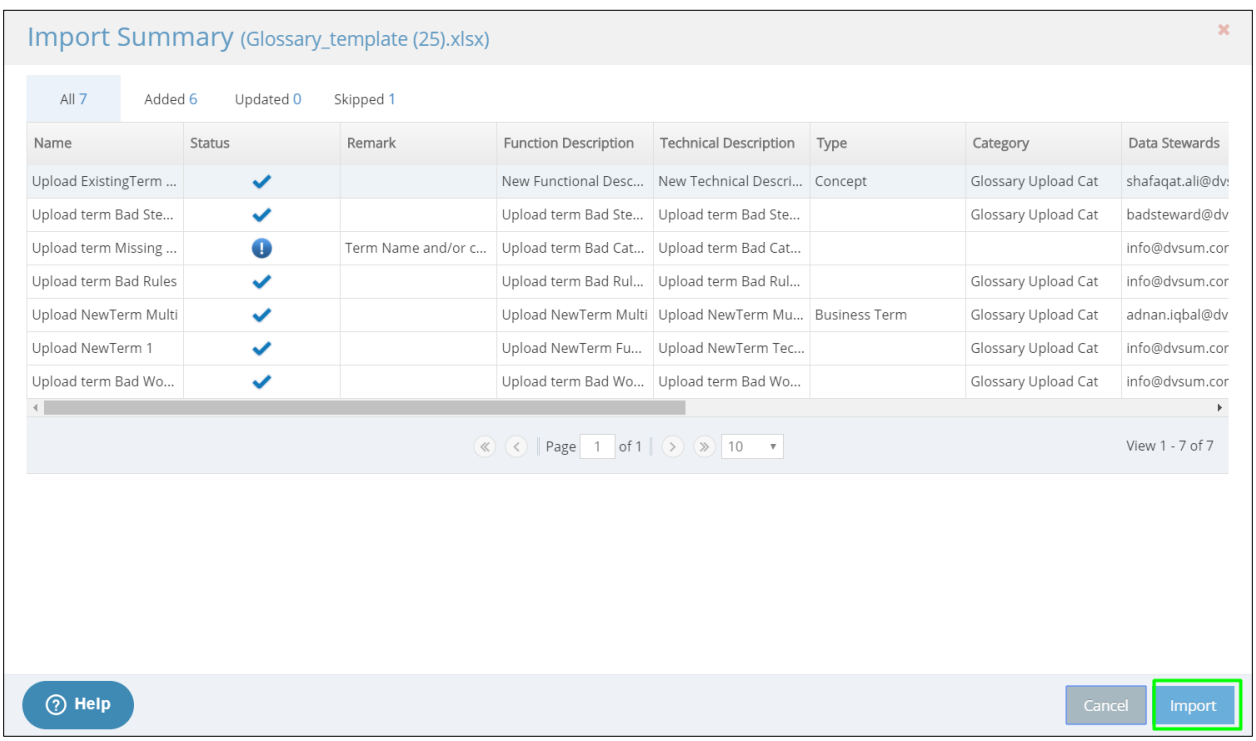
Step 3: Upload the "Glossary Template" in the upload section of this page and choose any of the merge options.

Skip - This option skips the invalid terms and the terms that are already added in DvSum Glossary.

Overwrite - If you want to replace the existing terms with new data, choose this option.

Clone - If you prefer keeping old terms and new terms, this function will create a clone of those terms.

Step 4: Click on Upload which will open the file for viewing.



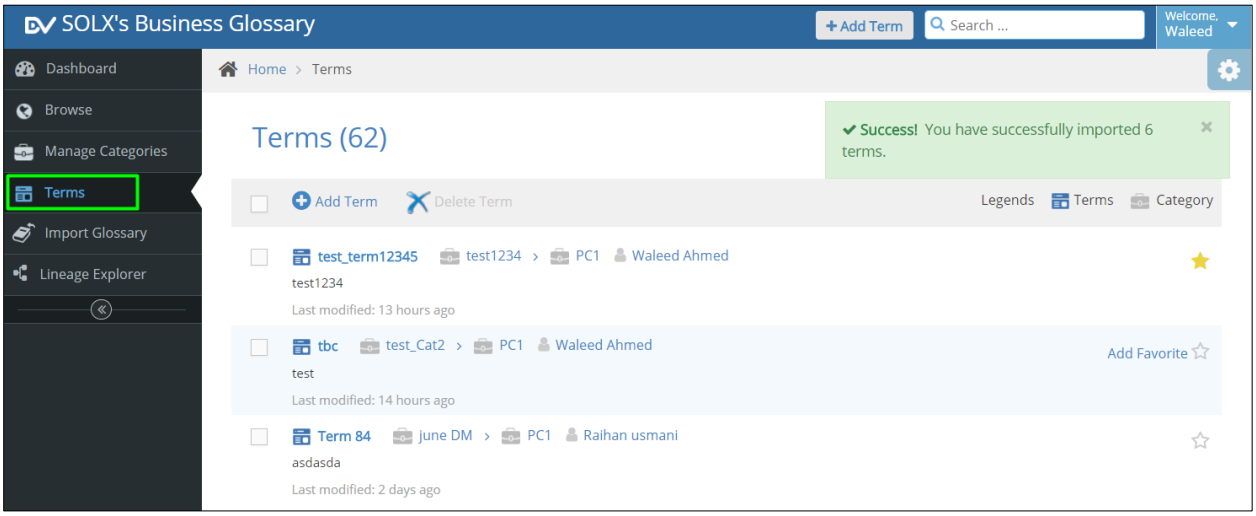
Among the list of all terms, you can see the are categorized for your ease.

Added - are all terms with valid status

Updated - are all terms which are to be overwritten (only appear if you choose the option to "Overwrite")

Skipped - are invalid or duplicate records (duplicates only appear if you choose the option to "Skip" the duplicates)

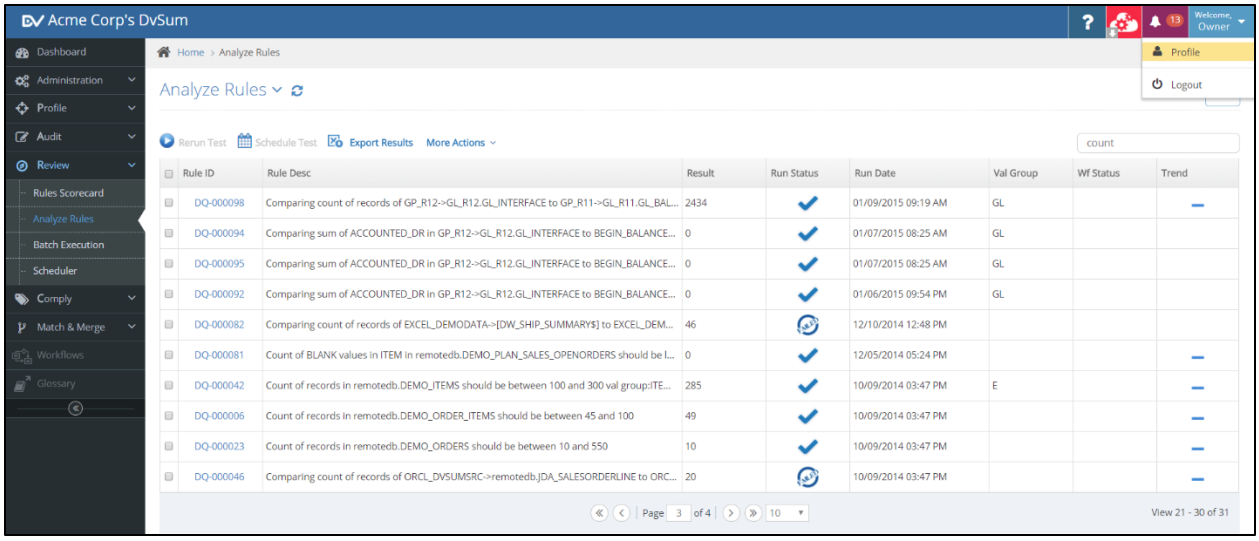
Step 5: Click on "Import" button to add and view these terms on Terms page.

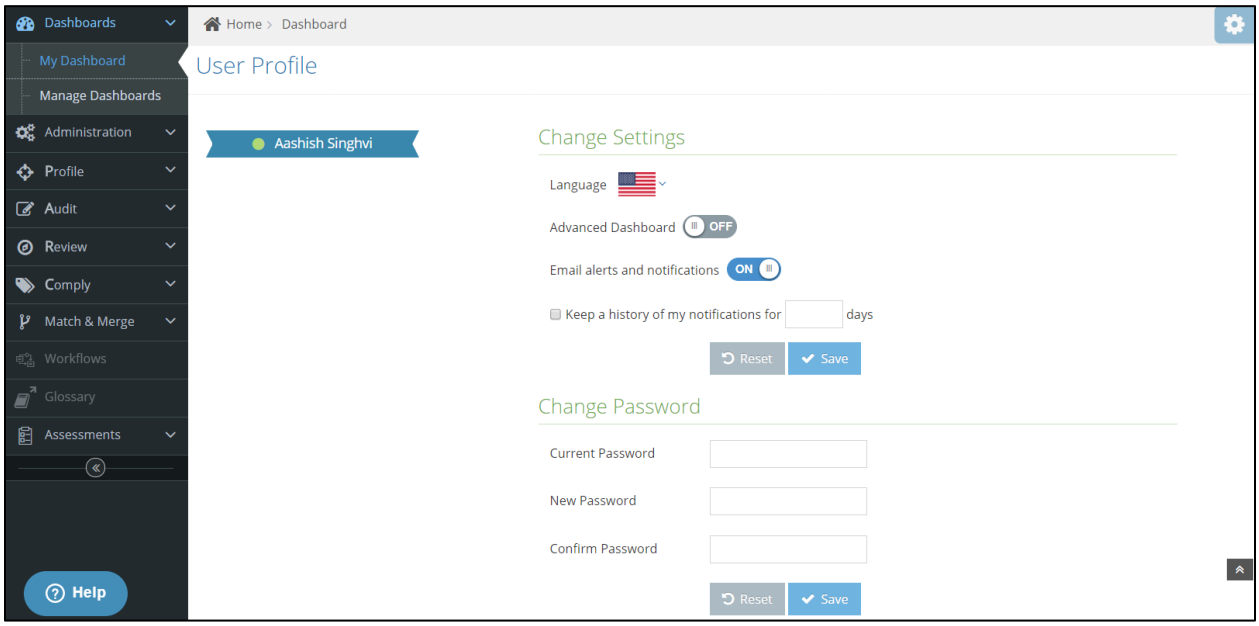


20 Updating your profile

20.1 Change Settings

DvSum supports language localization which means the whole application can be translated to your cultural preferences in any other language than the source language (English). You can change that from Profile Page.





Our Advanced Dashboards are OFF by default. A new user gets to see the "basic dashboard". You can switch to advanced dashboard once you get familiar with the application.

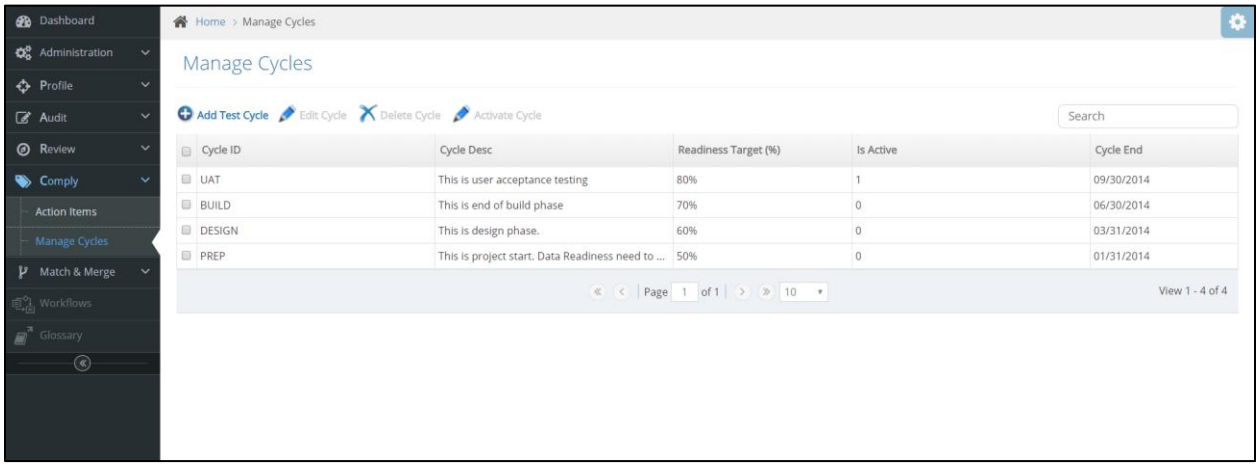
Action notifications are sent to users online on top of the page, except for account notifications which are always sent by email. If you would like to receive notifications by email as well, you can do that by changing the email preferences on the Profile Page.

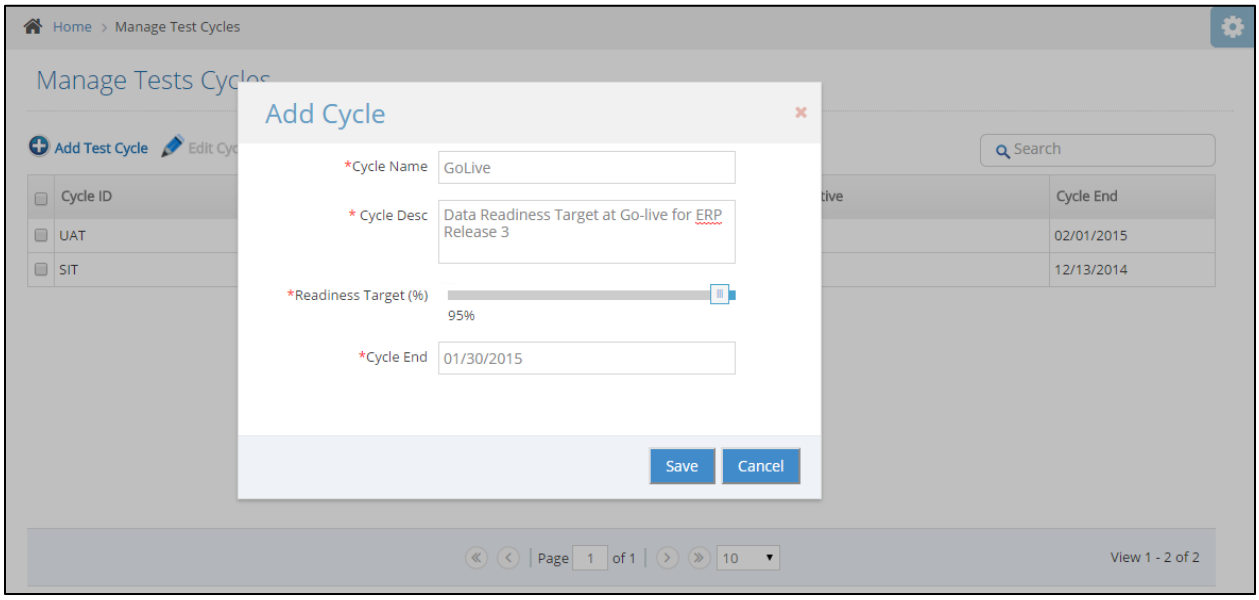
20.2 Change password

You can also change your account password from this page.

21 Manage Cycles

You can set the Data Readiness goals using the Manage Cycles workflow. This is only accessible by owner and administrator roles.





5 Demo Data

When you setup a new account, Acme Corporation demo dataset is automatically added to your account. This will allow you to start test-driving the solution immediately. Because of this demo dataset you will see data sources, data profiles, audits, test results and action items populated in your account.

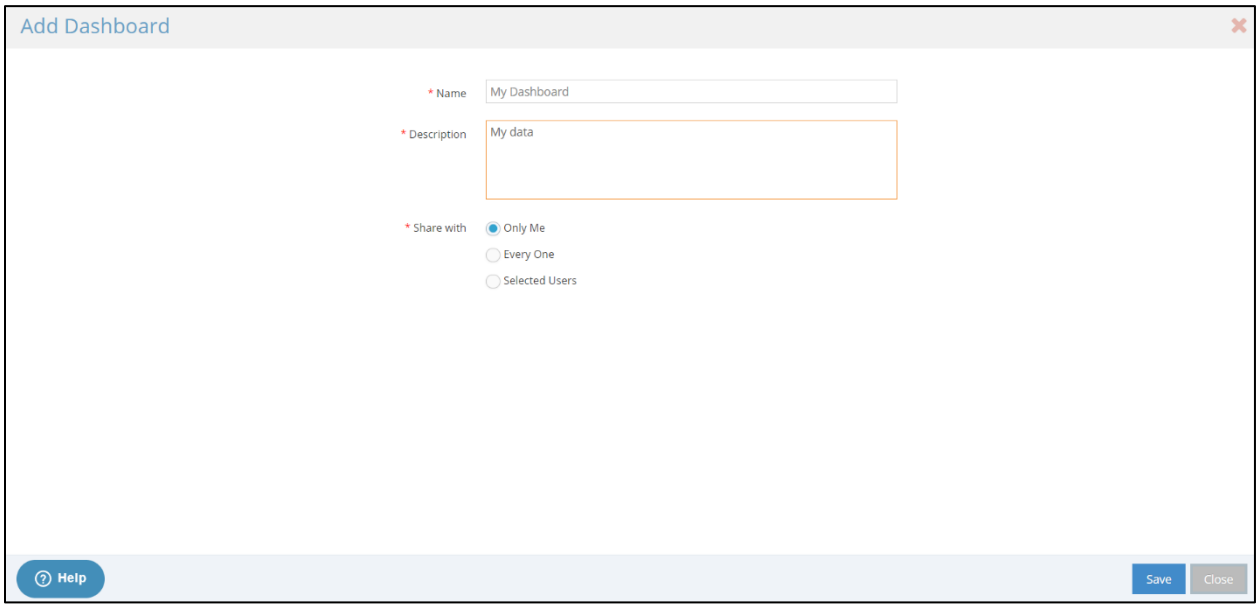
Applicability: All users

22 Demo data details

22.1 Dashboard

Step 1: Create a Dashboard

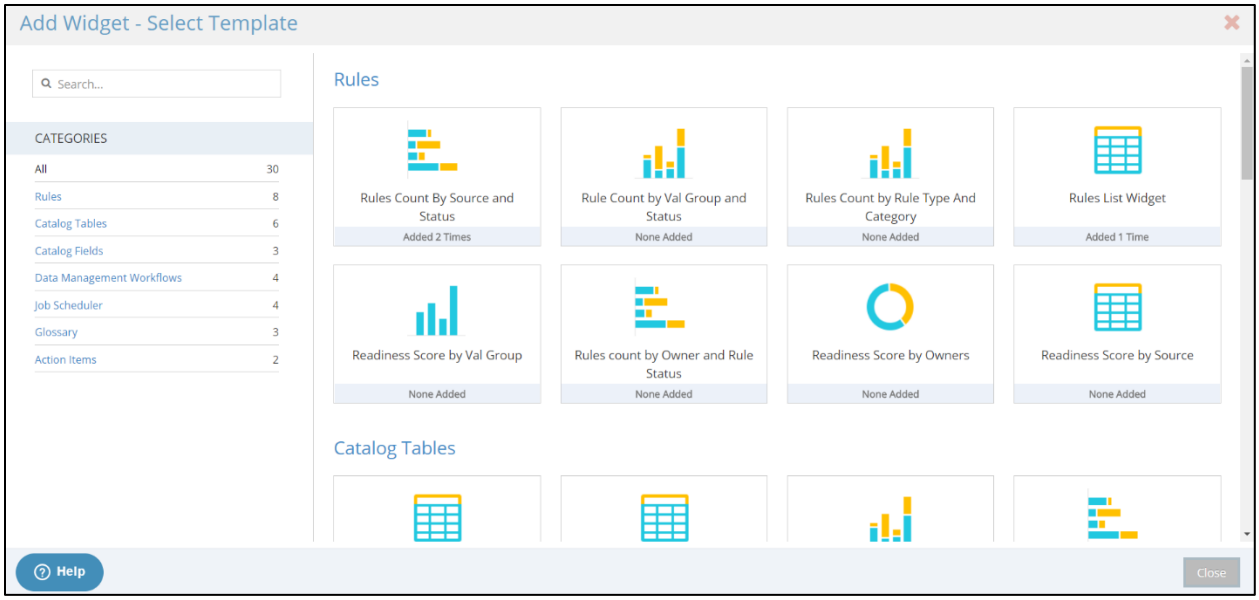
Go to Dashboards >> Manage Dashboards Page. Click on "Add" button. Specify a name and description. You can also choose to **Share** your Dashboard with others or keep it to yourself.



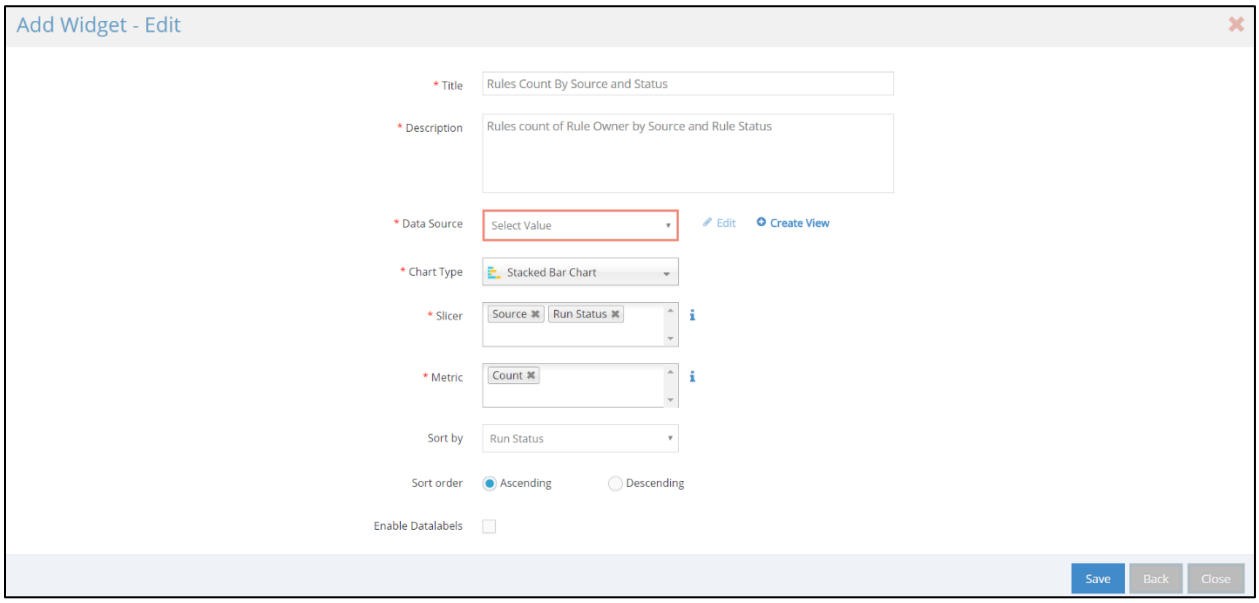
When you "Save" this Dashboard, you will be navigated to "My Dashboard" tab where the newly created dashboard opens. Since you need to add widgets to your dashboard, an "Add Widget" dialogue will appear.

Step 2: Add widgets

You can add template widgets to your dashboard.



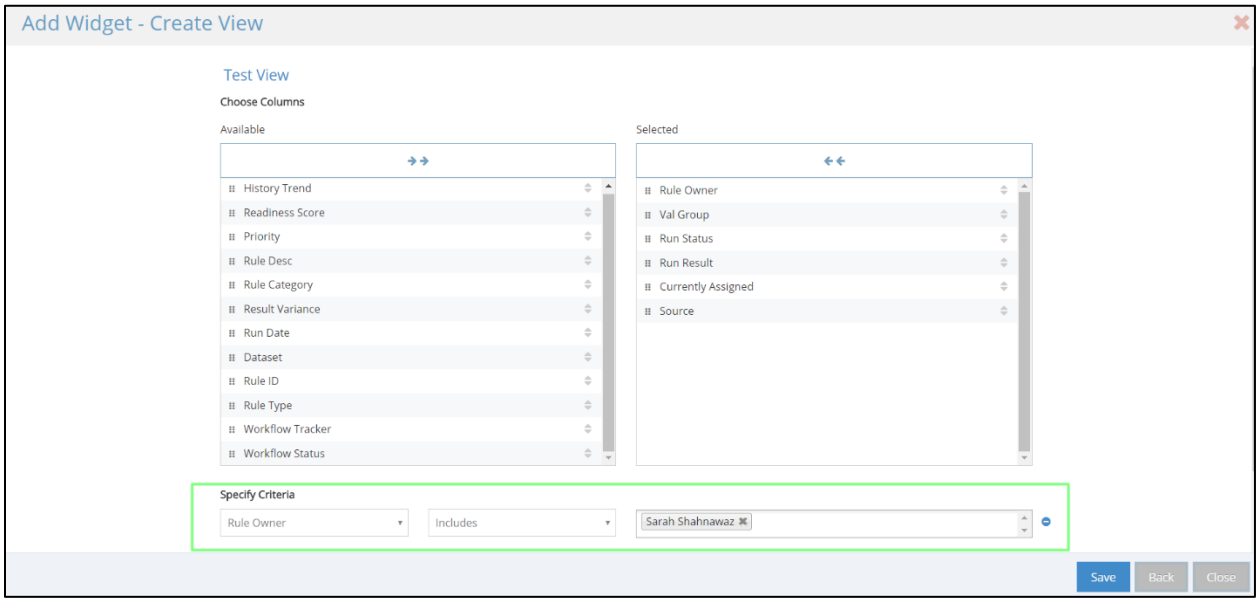
Upon selecting a template widget, it opens the widget dialogue which allows you to specify on which data source do you want the template widget is to display the information (*explained in next step*). These template widgets auto-populate the preferred choice for **chart type** related to **slicers** (Y-axis) and **metric** (X-axis). You can **sort** how the data appears on the chart and enable **Datalabels** (legends) on charts.



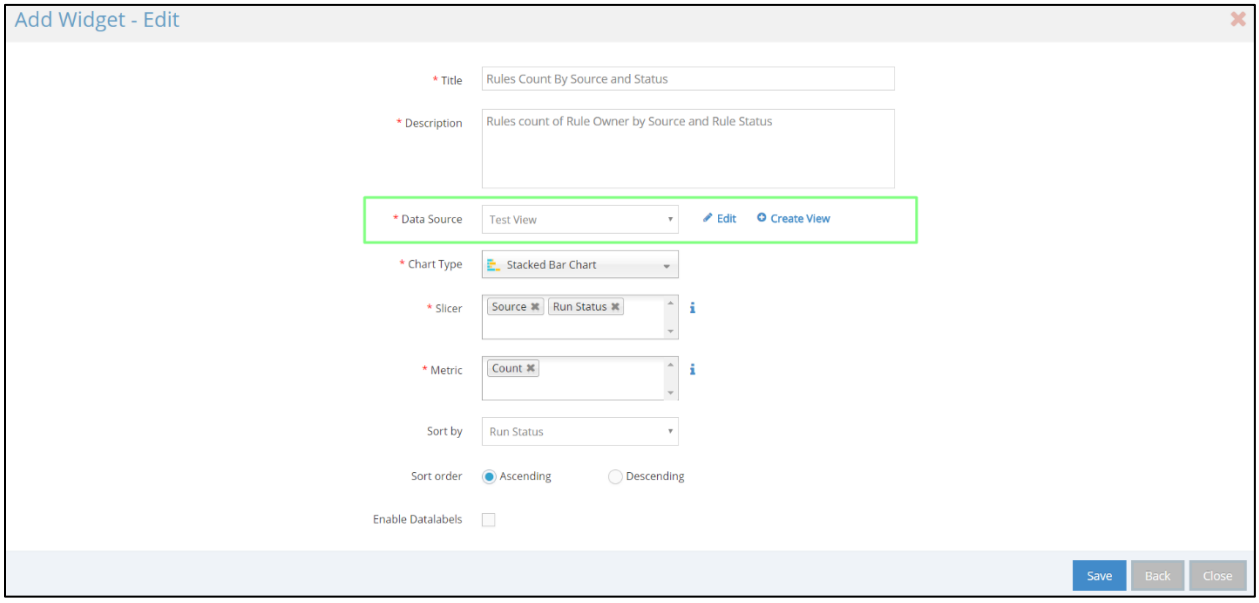
Step 3: Create a View

Click on "Create View". It opens an "Add Widget - Create View" dialogue. Based on this View, the widget will display the information.

Enter a Name for View. Specifying columns won't appear on the chart because that's what slicers and metric do. But that is required to create a view. However, the conditions that you apply on a view under the "Specify Criteria" section filters the data accordingly before displaying it on the dashboard.



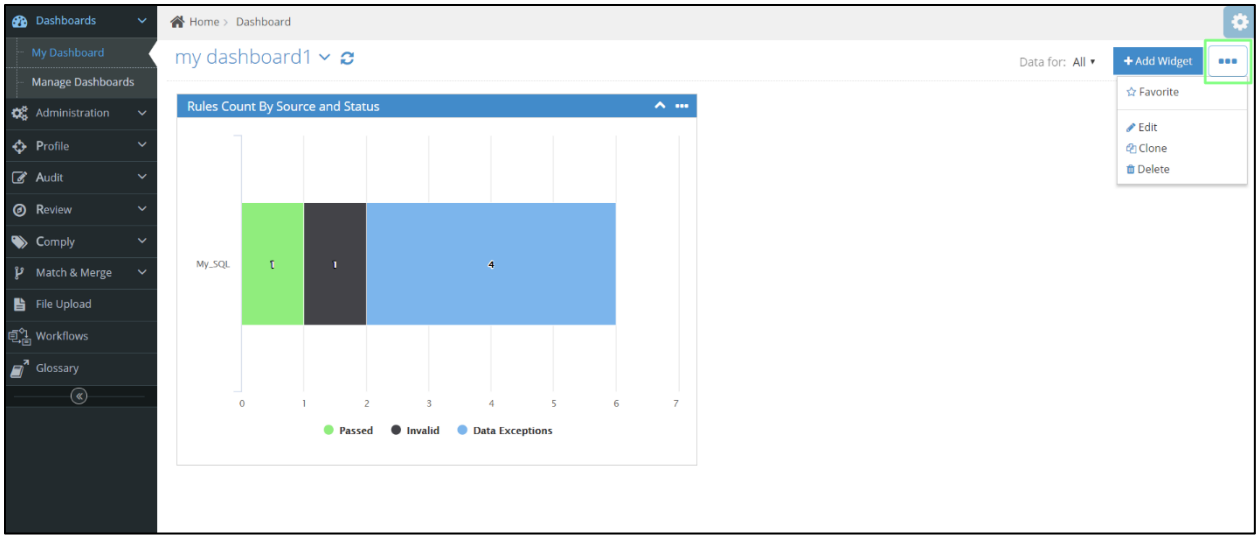
When you save this view, it will be populated in the "Add Widget" dialogue of step 2 (image shown below).



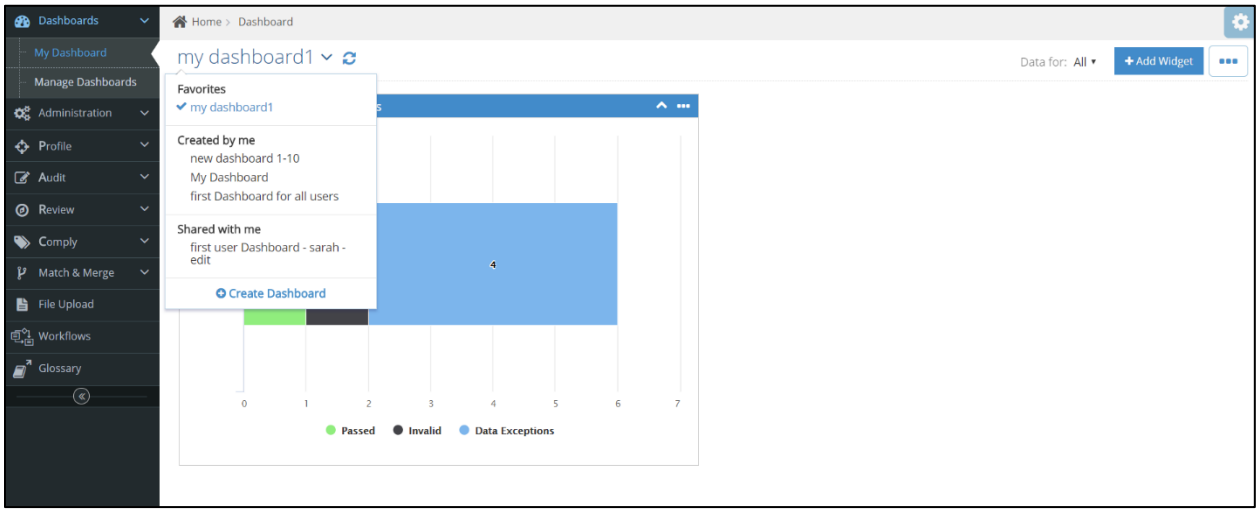
Save this widget to make it appear on Dashboard. This completes Step 2.

Step 4: Dashboard

The widget will start appearing on Dashboard. You can add more widget by clicking on "**Add Widget**" Button. You can mark this dashboard as **Favorite** or **Edit**, **Clone** and **Delete** it.



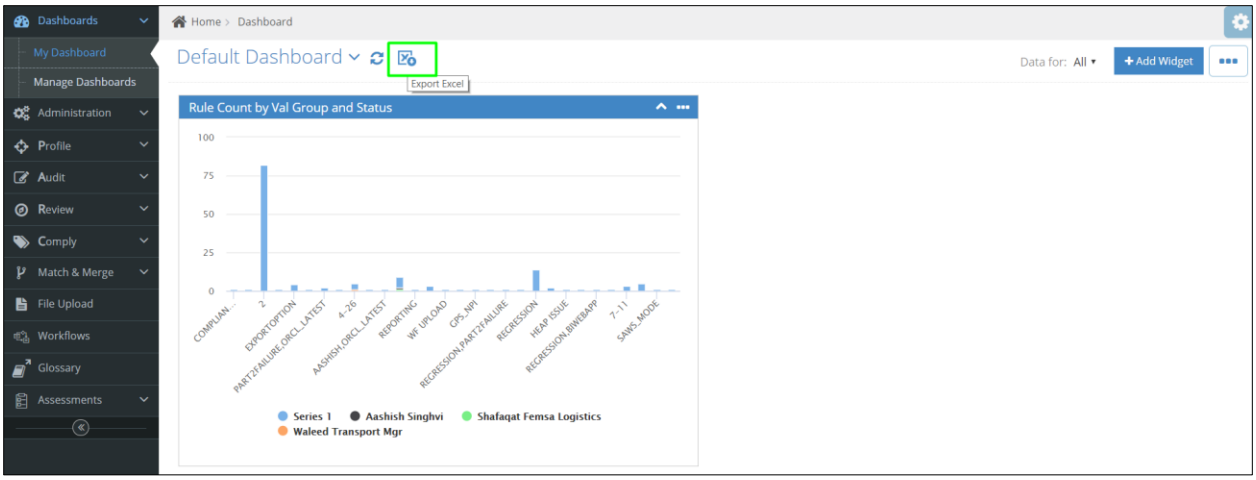
Once you mark the Dashboard as Favorite, it appears in the Dashboard drop-down under the "Favorites" section along with "Created by me" dashboards and "Shared with me" Dashboards.



Note: You can view the dashboards shared with you but you cannot edit them. You may however clone the dashboard and edit it according to your liking.

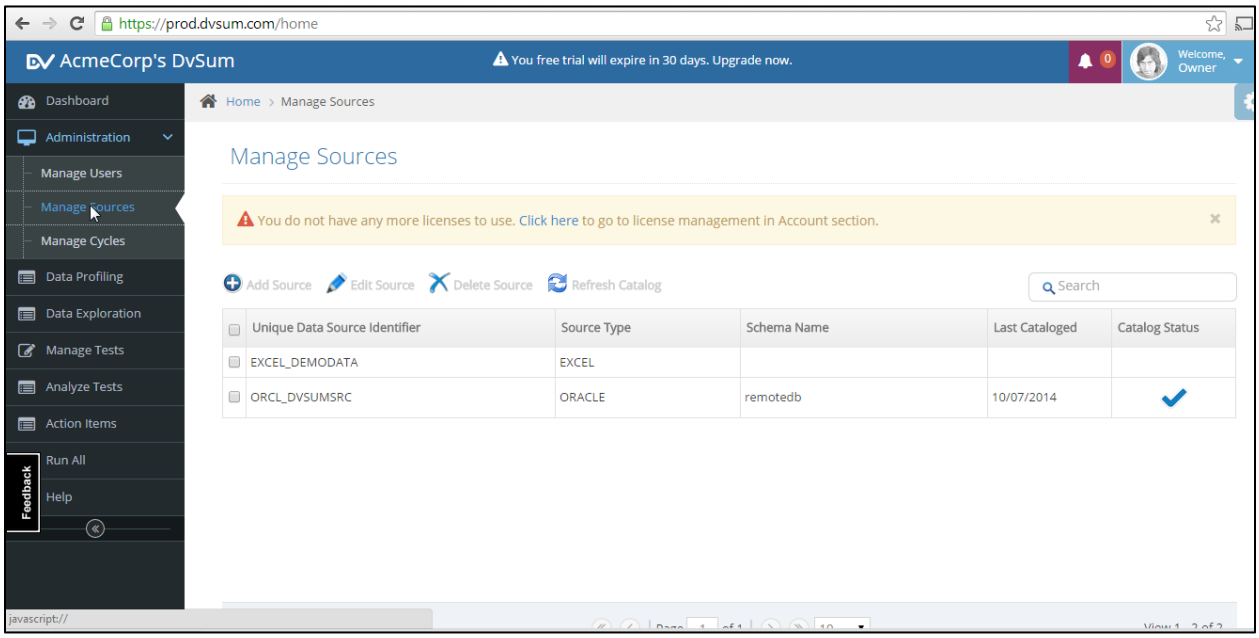
22.1.1 Export Dashboard

Export functionality allows you to download the data of widgets in a dashboard.



22.2 Data Sources

Two demo data sources are available in your account.



- **ORCL_DVSUMSRC** - Oracle data source that contains multiple data objects. Master data like demo_items, demo_customers, transactional data like demo_sales and planning information like demo_plan_production.

This data source is an oracle database schema that is hosted on a publicly accessible database server by DvSum. The credentials for this database are set but are not provided to you to prevent accidental changes to the demo server. You are able to run the full functionality of DvSum using this demo server with the provided setup.

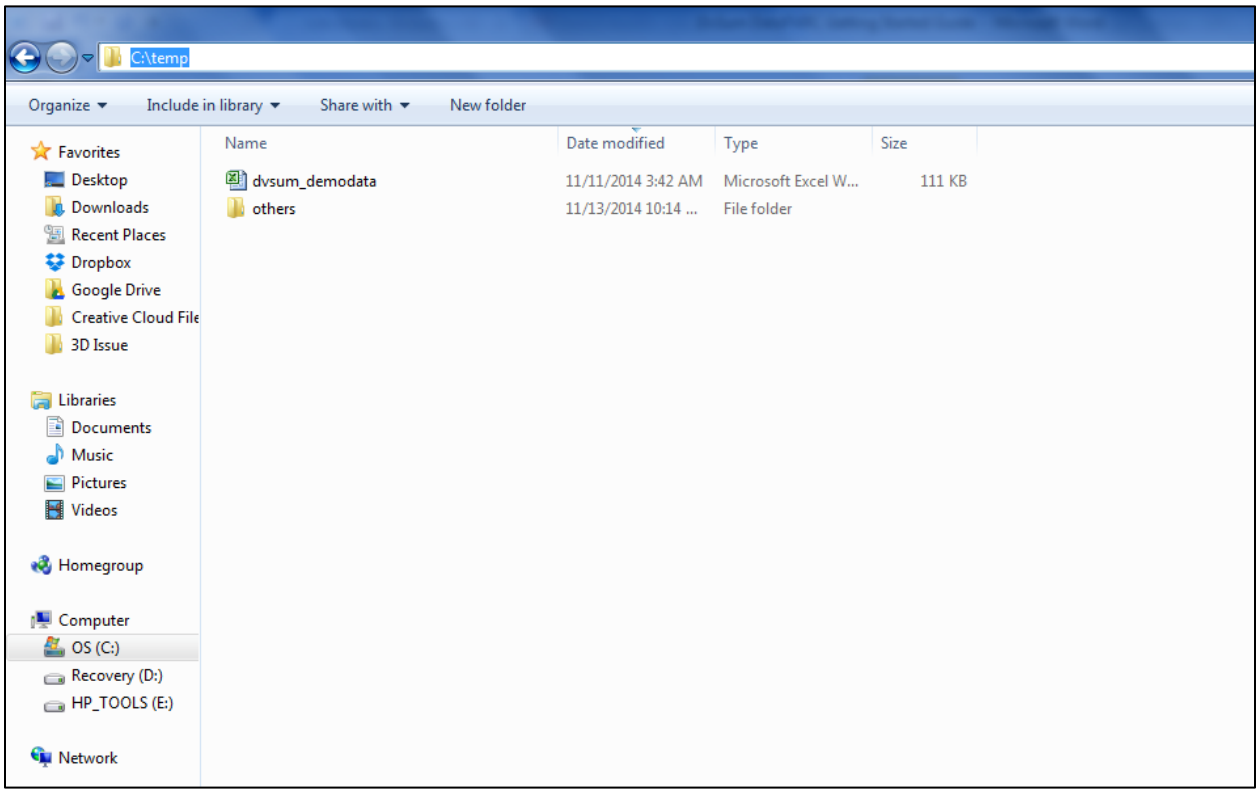
- **EXCEL_DEMODATA** – Excel data source that contains multiple objects.
    a. MFG_ROUTINGS – routing and cycle time information for manufacturing

b. SALES_FORECAST – monthly forecast by item, customer, plant
c. PURCHASE_PLAN – weekly purchase requirements by supplier, plant
d. ERP_SHIP_DETAIL – extract of shipment history at transactional level emulating data within a transactional system like ERP
e. DW_SHIP_SUMMARY – summarized shipment history at aggregated level emulating a data warehouse

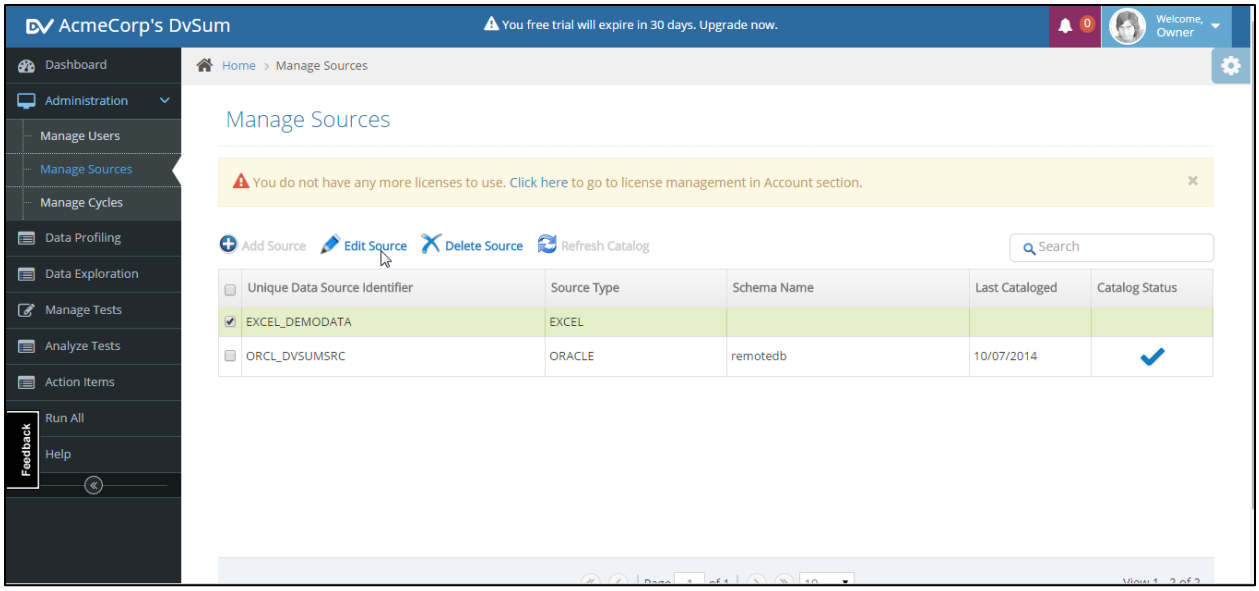
This is an excel data source. **Please download the excel source file from DvSum site here.** This is a mandatory step for you to run data profiling, audits on this data source.

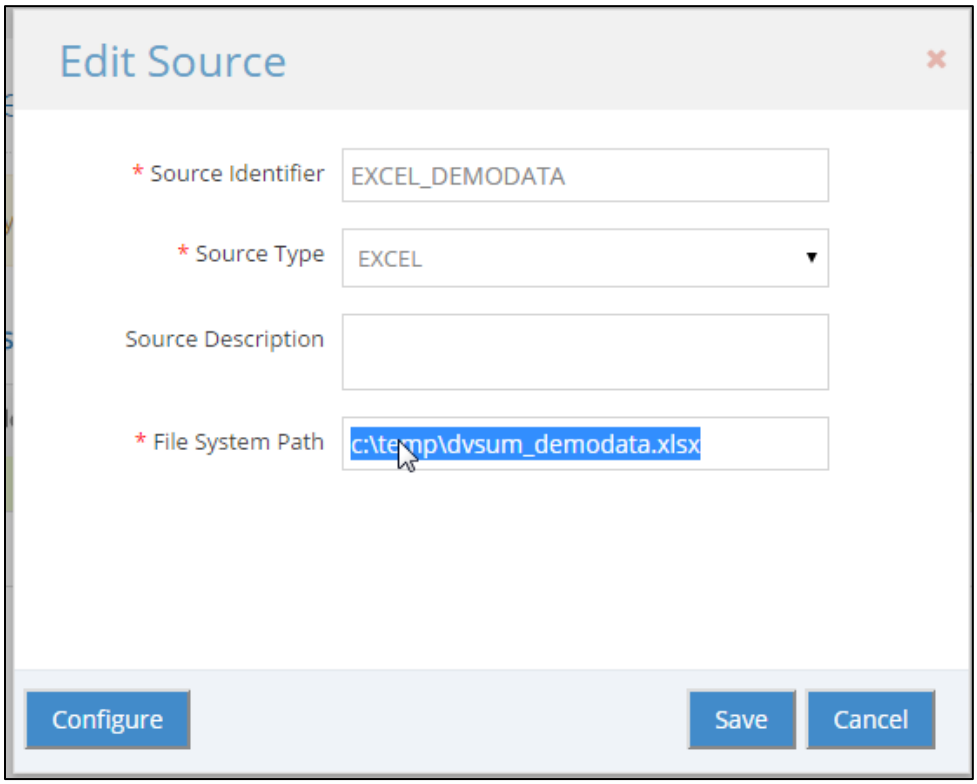
http://dvsum.com/wp-content/uploads/2014/11/dvsum_demodata.xlsx

The data source is by default configured to a location **c:\temp**. Therefore, please download the excel file and store it in c:\temp folder on your PC. If you do not have a temp folder, then go ahead and create one.



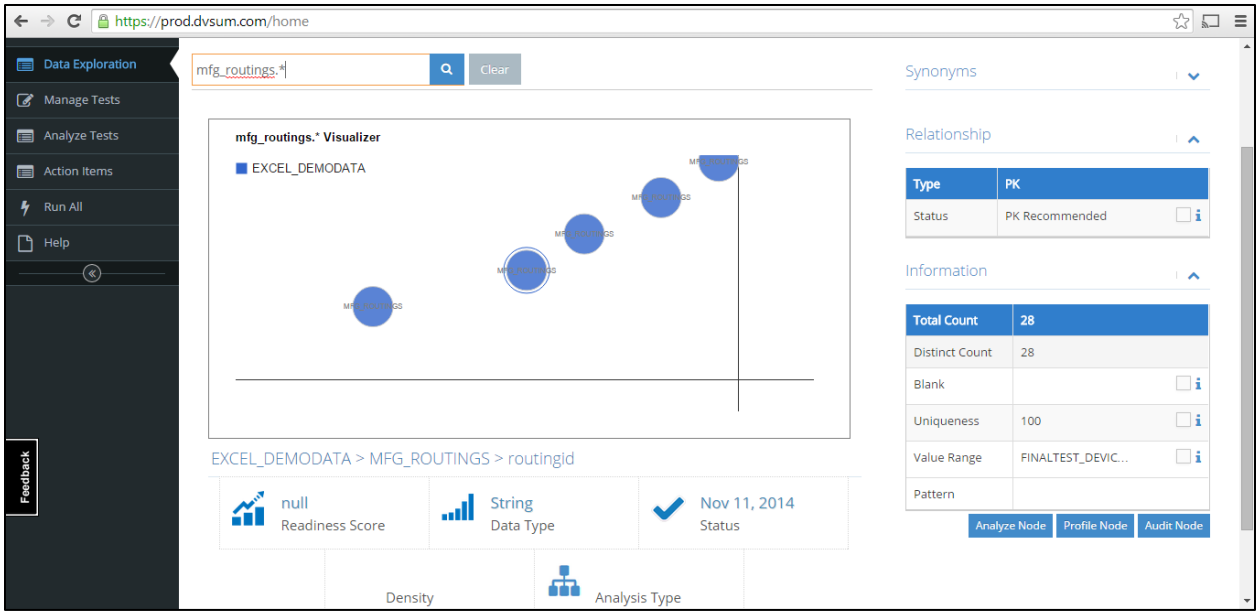
CAUTION	If you prefer to save the file in another location, you can do that but you will have to change the source configuration.





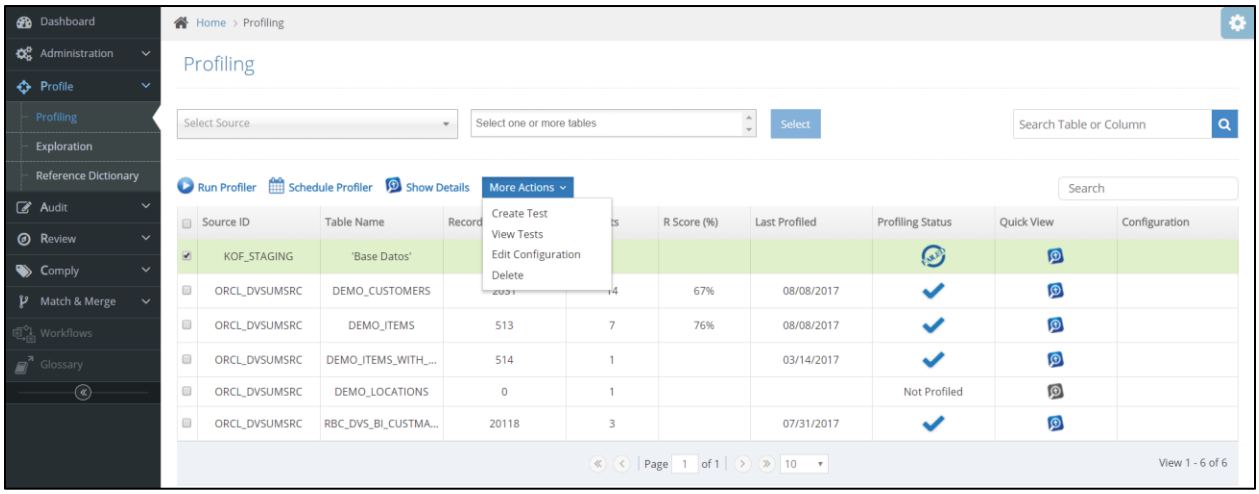
22.3 Data Catalog

Data Catalog for both these data sources is available for the demo data. You can navigate the data catalog using the Data Exploration workflow. If you are an administrator or owner of account, you can refresh the catalog using the refresh Catalog workflow under Manage sources.



22.4 Data Profiles

Data Profiling for both the data sources has been executed for many but not all the data objects. This is done intentionally, so you can see the difference in available statistics between data that is not profiled and what information is available after profiling. To know more about data profiles, you can use Data Exploration workflow or the Data Profiling workflow. All users can refresh the profiling of a data element by using the Run Profiler workflow on Data Exploration or Data Profiling section.

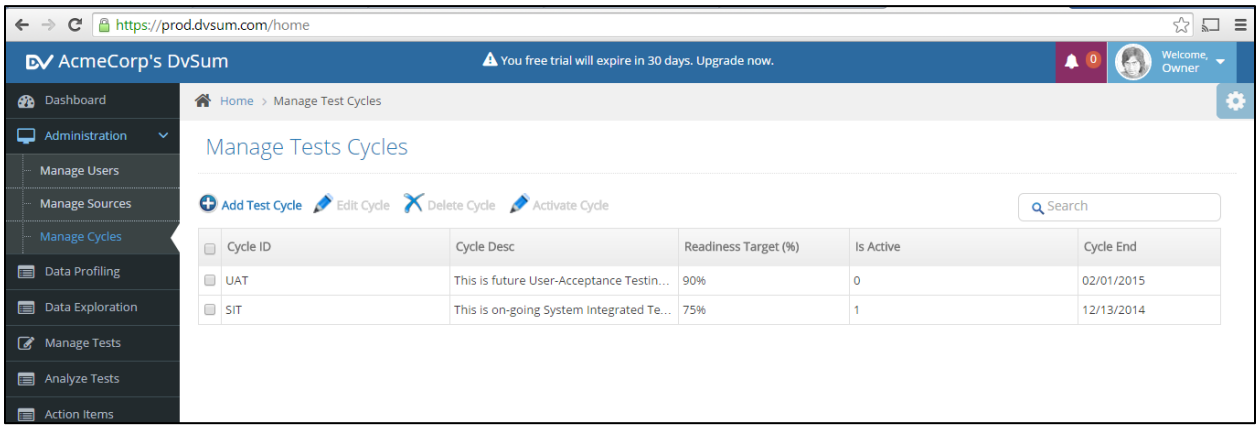


22.5 Data Quality Milestones

2 milestones are set as default as part of the demo dataset.

- SIT – System Integrated Testing which is the active milestone with a due date as 30 days from when you signed up

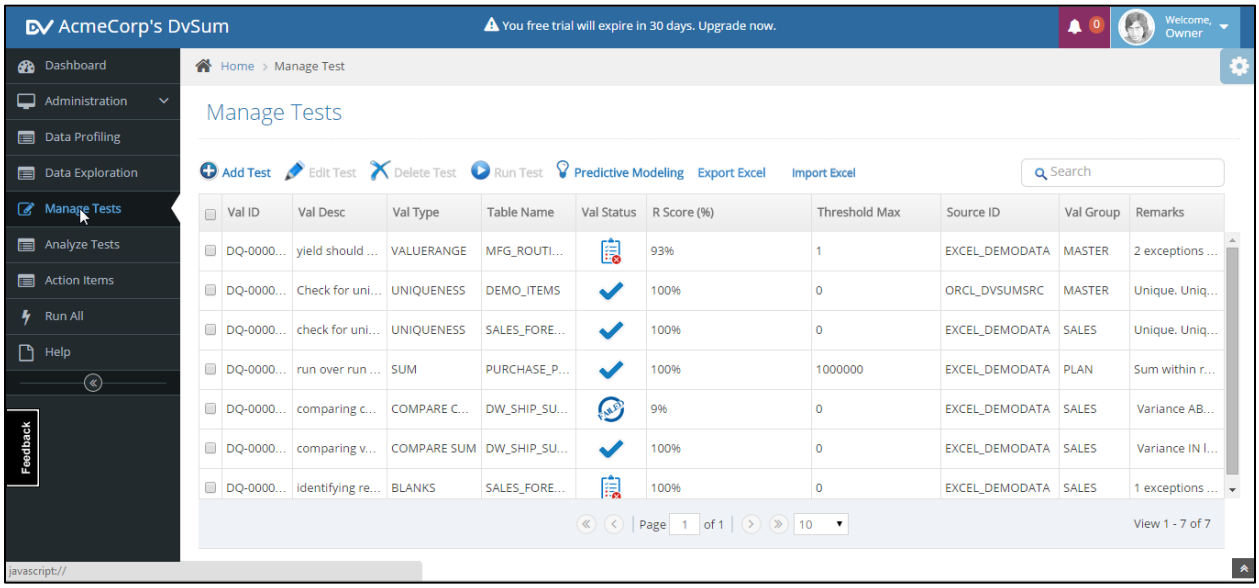
- UAT – User Acceptance Testing which is a future milestone with a date 10 week from when you signed up.



	These are sample milestones only. It is recommended that you replace these milestones that reflect the data quality initiative (project driven or on-going operations) of your organization

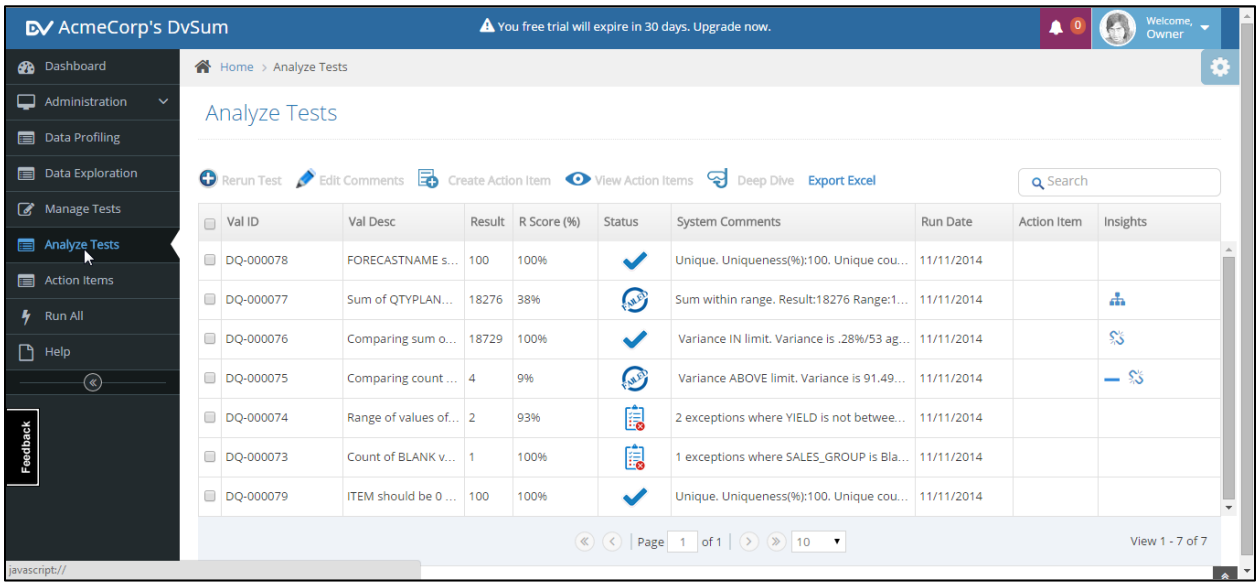
22.6 Data Audits

7 data quality audits have been created in the demo dataset. The definition of these audits can be accessed by navigating to Manage Tests on the left navigation.



22.7 Audit Results for analysis

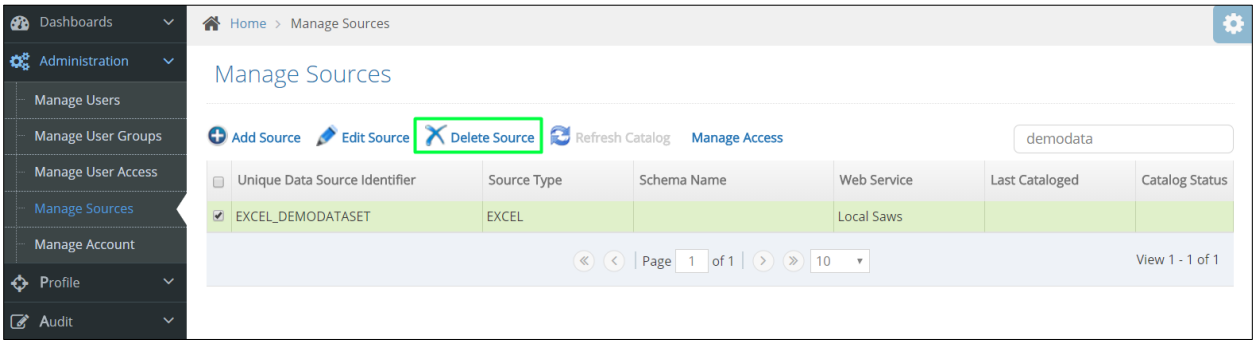
The above mentioned 7 data audits have been executed and you can find the audit results and deep-dive insights by navigating to the Analyze Tests section.

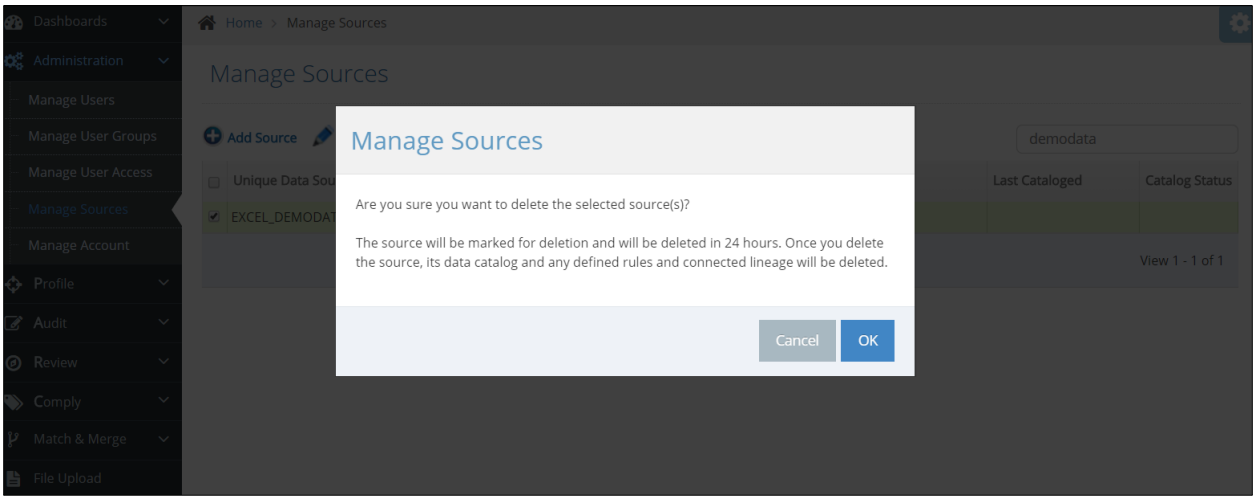


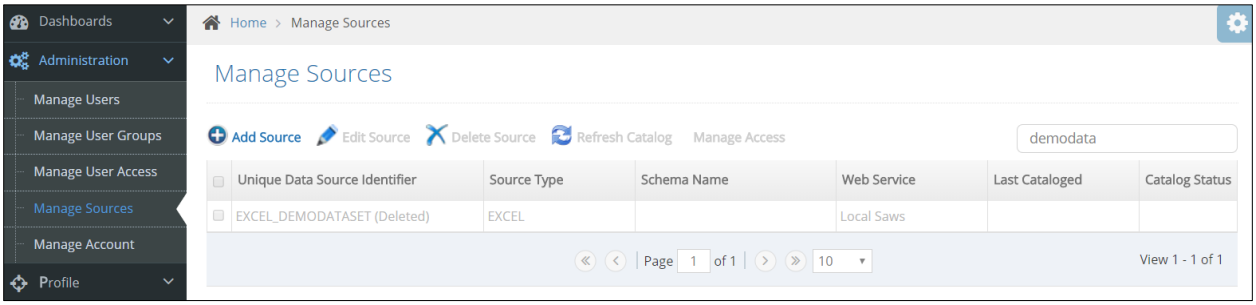
23 Deleting the demo dataset

Once you have setup the execution workflow using the applet or web-service and are ready to setup your own data sources, you can go ahead and delete the demo dataset from your account.

To delete, navigate to Administration → Manage Sources on the left Navigation. Select the data sources by name EXCEL_DEMODATASET and click on "Delete Source".





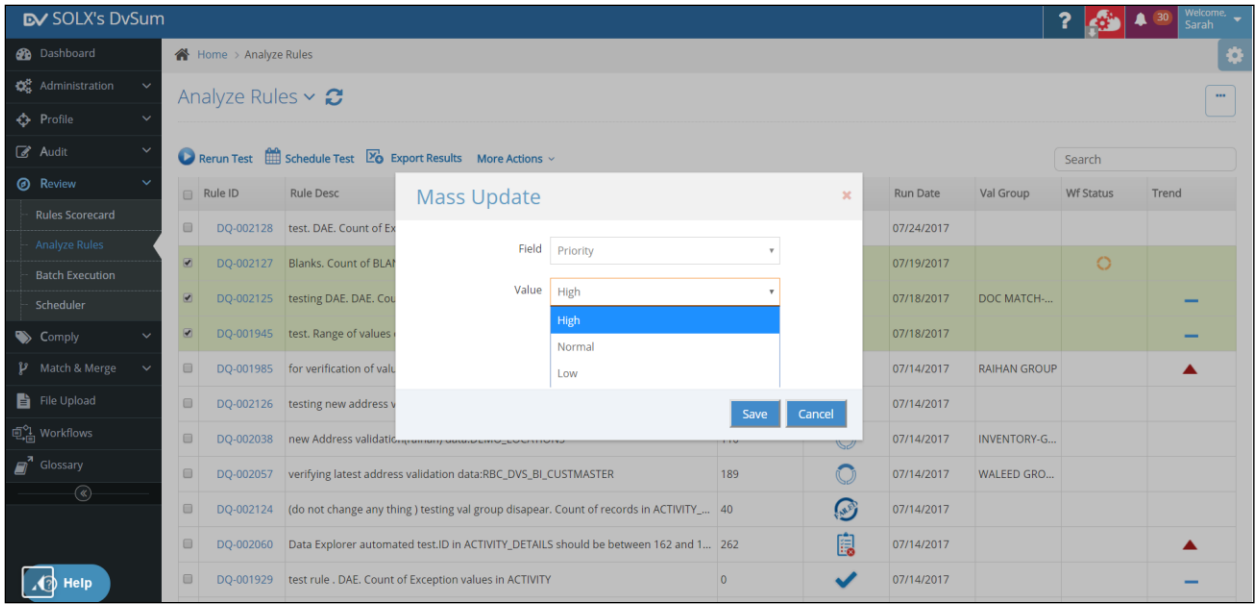


This action will mark the source as deleted but it will be cleaned from the Source list after 24 hours. However, once you delete the source, it will stop appearing anywhere in the application i.e. profiling tables, the rules created or the Workflows related to this source will be deleted.

	Administration section is only available for Owner and Administrator roles. If you are a user or Super-User, you will not be able to delete the demo dataset.
CAUTION	All data for the selected demo datasets will be deleted. This includes profiles, audits, results and action items. The data is also at an account level. So once the data sources are deleted, they are deleted at account level and no user can use it. It is recommended that you delete the demo dataset after confirming with any users in the system who might be using the demo dataset.

24 Mass Update

DvSum provides the handy feature where you can edit multiple rules at once from a list view. There are number of fields available for mass update from Manage Rules and Analyze Rules grid.



You can select the rules and click on Mass Update from More Actions. The fields available are: Priority, Val Group and Assigned to where you can reassign multiple rules to another user or user group.

For every field, you can select values. For e.g. for priority, a user can select high, medium or low. Once you click on Save, all the selected records are updated.