DvSum DataPARC®

Quick Start: User Guide

Updated: Q3 2016

Table of Contents

1. Navigation & Dashboard

1.1 User-specific Dashboard (Home Screen)

The DataPARC Dashboard is the starting point for reviewing the current status of the data quality. The system and views are user-sensitive, so you will only see the Rules and the scope of data you are responsible for, or have permission to view. Here, you can see the key metrics in the Tests Scorecard and the Readiness Scorecard, and navigate to the various functions.

The Tests Scorecard gives a summary of the total number of audits in the system, how many are passing, how many are failing and the number of 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.
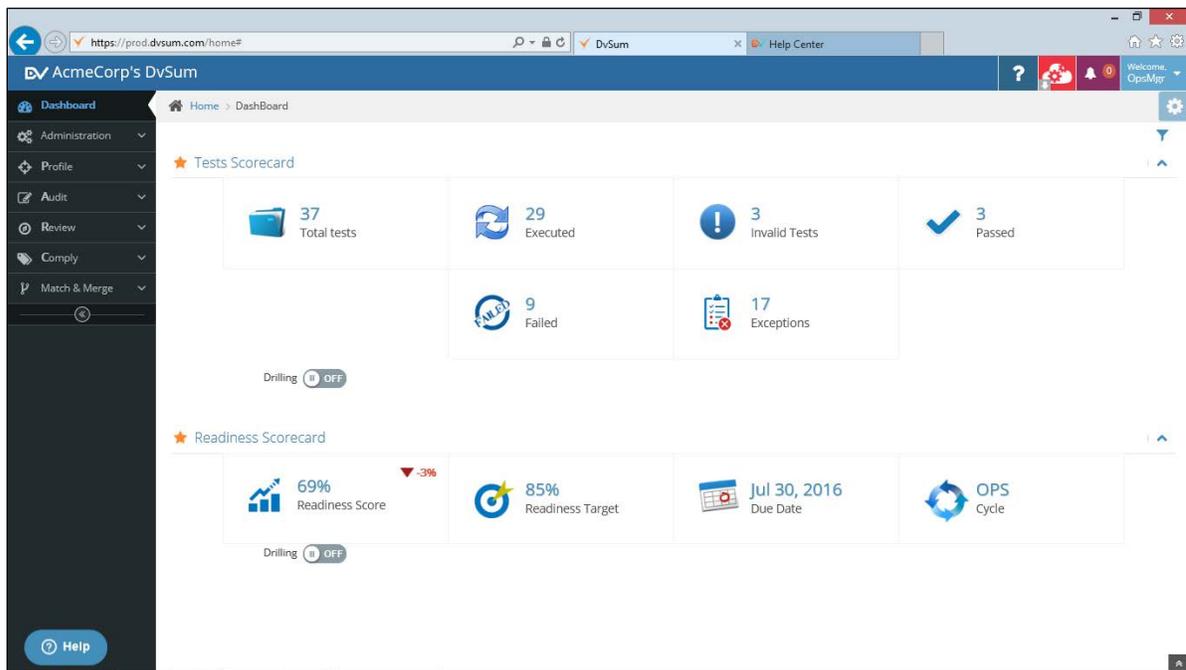
Depending on your role and scope of data, you can also see a breakdown of the data quality status by data source, data type and user group.
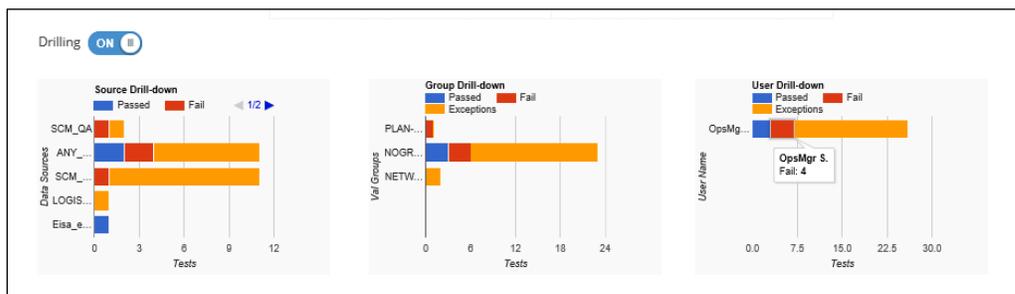
1.2 Navigation Menu
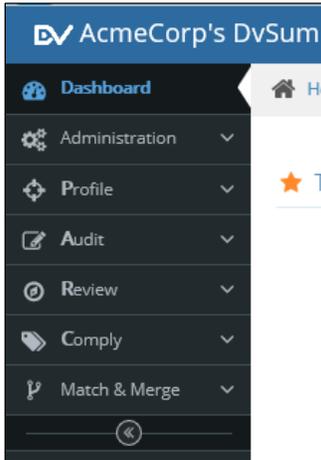
Primary navigation for DvSum is through the left-hand menu. Depending on your role, you may not see all the items shown, such as Administration, Comply or Match & Merge. The core menu items are built around the DataPARC process:

- **Profile:** Discover and explore your data to see statistics, distribution of attributes and determine initial rules
- **Audit:** Create and Manage Rules, and import Rules from the Rules Library
- **Review:** Analyze results of Audits, the data Exceptions by Rule, and fix the data
- **Comply:** Assign action items and manage readiness targets and dates

2. Creating Rules

Rules are created and added to Audits to be run together with other rules, or can be run individually. They can be scheduled or run ad-hoc. With DvSum, you define a data audit rule once, which can then be run repeatedly.

Rules in DvSum can be created from a number of different locations within the tool, covered in this section. They can also be added from the Rules Library, or if you have existing rules for your data, they can also be upload to DvSum.

When a Rule is created it is assigned a unique ID, such as DQ-000167. This links to the Rule page which includes the Rule Definition, Summary, Exceptions from the latest Audit run, and trends.

2.1 Standard Rule Creation

The majority of Rules are created through the Rules Wizards provided within the tool. From the left navigation pane, access Audit > Manage Rules, and click "+ Add Test."
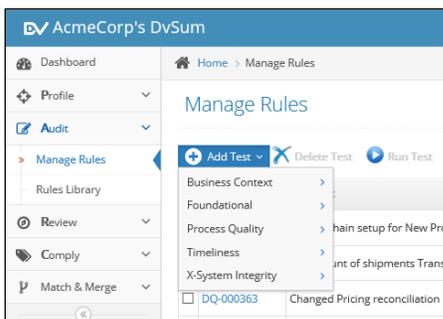
2.2 Types of Rules Wizards

We encourage customers to begin by identifying whatever data errors are causing impacts to their business, and start with rules to identify those exceptions. Some of the Rule types are straight forward, while others were created for specific customer use cases, and the potential application to your environment may not be readily apparent. Over time, potential rules and use cases become more apparent as you build out a more holistic data governance approach.

The Rules wizards you can configure are organized in the following categories:

Foundational	Basic, logical checks you would typically do in any environment, without necessarily knowing the contents of the data
Business Context	Validations and checks based on a specific business understanding of what the data should be
Process Quality	Is the data valid or what is expected based on the process or purpose
Timeliness	Alerts if data not updated or process not completed by a set time or duration
X-System Integrity	Compare and reconcile details and records across systems – without having to consolidate or pull the data.

Within these categories are various rule types that can be configured to validate data and generate the exceptions. There may also be different approaches to start with different rule types that generate the Exceptions you're looking for. We encourage users to explore different rules and approaches, as there is some art to determining what works best within your environment and business processes.

Type of Audit Rule	Definition	Example (s)
Foundational		
BLANKS	Availability and completeness of data. Are there missing values?	All the records where address field is missing in customer ship-to information.
ORPHAN RECORD	Record exists in one system's set of Records, but not in another system's set of what should be the same Records.	Order A exists in ERP's list of Orders, but not in Transportation System's list of Planned Orders
ORPHAN KEY	Group of Records of a certain type exist in one system but not in another. Summarized by type	Shipments for Customer X, not found in list of Shipments by Customer Product Category B, not found in list of Products.

Business Context

UNIQUENESS	Are the values within a data element unique in the dataset?	Duplicate customers in customer table. Data is not unique
UNIQUE VALUES	Are the Count of Unique Values within a range or equal to a defined number	Expect unique locations to be between 95-100 based on 100 ship-from locations. Over 100 indicates error, under 95, something may be wrong
VALUE RANGE	Are the values in data within the expected range? Either within a continuous range, within a Pick List, or one of a specific value.	Manufacturing yield should always be between 0.01 and 1 Order type in sales orders should only be one of ZOST, ZOCO, ZOFR. Item_category in shipments should only be one of values in the item_category reference
CUSTOM QUERRY	Ability to write Custom SQL (Non-SAP)	Any Custom SQL Query

Process Quality

Address Validation	Validates existing address against Google Maps. Also enables standardization	123 Main Rd, should actually be 123 W Main St.
MODEL COMPLETENESS	Does valid data exist in reference systems to execute a process	Does each SKU have a valid Bill-of-Distribution? Does each Item have weight/measure populated?
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

Timeliness

RUNTIME CUTOFF	Monitoring if Data not present or updated by a certain time, send alert	If Orders not uploaded by 5pm. If Updated Forecast not received by Plan Generation at 12am.
RUNTIME ELAPSED	Monitors Data to determine if process from event 1 to event 2 exceeds	If Product attributes not populated within 1 week of New Product being created in system If Booking Confirmation not received within 3hrs of Booking Request.

X-System Integrity

COMPARE COUNT	Compare the count of records between 2 different data sources at same or different granularity	Compare total Orders or Shipments in ERP at transactional level to Order or Shipment summary loaded in data warehouse at aggregated level
COMPARE CUSTOM QUERRY	Create your own custom validation rule within one data set and compare to the same query in a separate data	Custom

	set or system.	
COMPARE METRIC	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
DOC-MATCHING	Compares Records at a Field level and highlights discrepancies	For Planned Orders in ERP vs Planned Orders in Warehouse System, validate dates, quantity and price and show discrepancies

1.3 Completing Rules Wizards

For each of the Rules wizards, information is completed within the following sections. Depending on the type of Rule, options will differ. If the Rule is applicable to a single data source, Reference Inputs will not apply. Doc-matching rules will also require different information.


For some rule types, under Basic Inputs, the Variance Check check-box indicates this is a rule that is comparing to a previous run and you will be asked to provide a Variance Threshold or %.

1.4 Creating Rules from within Profiling – Data Quality

When in Profiling, by clicking the Data Quality tab, you have the ability to create Rules and Rulesets for the data you are reviewing.
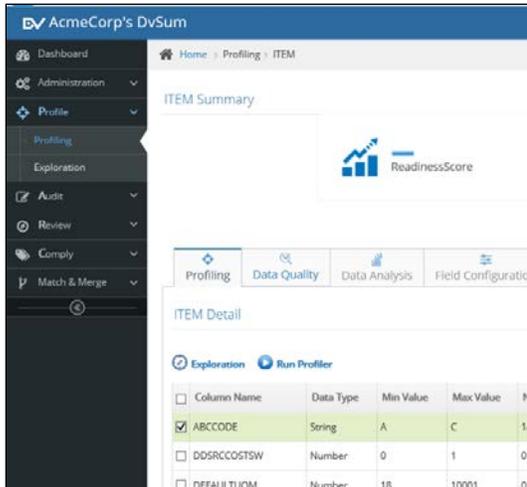

You must first click "Enable Ruleset," complete the required information and click "Save" before defining the specific checks.
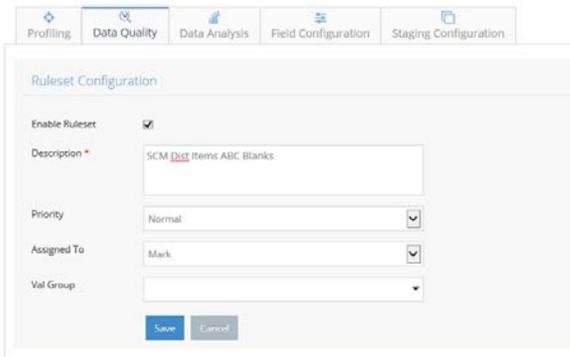

Rules created from the Data Quality tab within Profiling are limited to Completeness and Value Range checks. For example, by selecting the "Blanks Check" check box, the system automatically creates a rule to find and return blanks for that selected field.

You also have the opportunity to establish Filter conditions, to eliminate the scenarios where a blank might be an acceptable value.
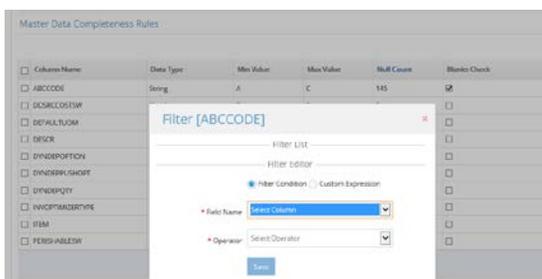
Within the same Ruleset, or as a standalone Rule, you can also define Value Range Checks. These can be defined based on a Min/Max Threshold, or based on select values from a Picklist within the Field or Reference Dictionary.
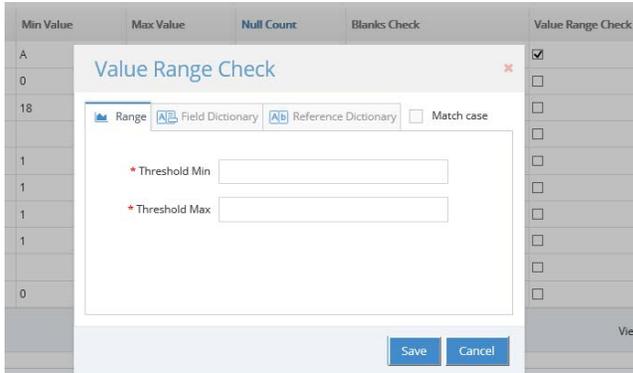

Based on what you define for the Blank Checks and Value Range Checks, this is automatically saved as a Rule, and can be viewed and edited under Audit>Manage Rules.

1.5 Creating Rules from within Profiling – Data Analysis

More complex, data combination Rules can be created from within Profiling, under the Data Analysis tab. When exploring data through the filtering analysis, you may identify combinations of data that you want to define as Exceptions or that you want to create a query for to review the record details.

For example if the combination is invalid and should not exist, a Rule can be created from this view. As seen below, for a Product/Item Category of Accessories (Accy) sourced from Asia, a user may know that the Plan Level (planning category or priority) should not be "4." By clicking "Add Business Rule," entering a Description, and clicking Create, a Rule template (or shell) is created, enabling you to create the specific rule.
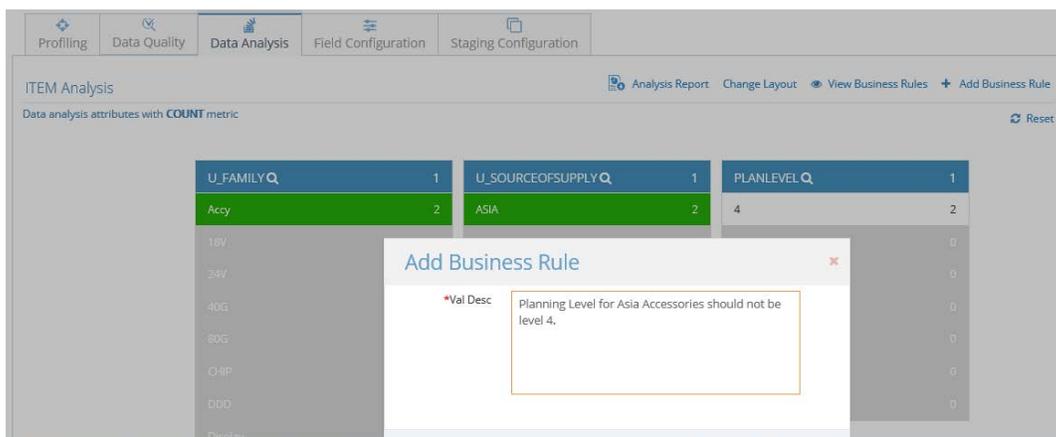
Clicking Create takes you Audit>Manage Rules with the Rule template created. From here you can Edit, add detail, Action Items, Create a Workflow, etc. before finalizing the Rule.
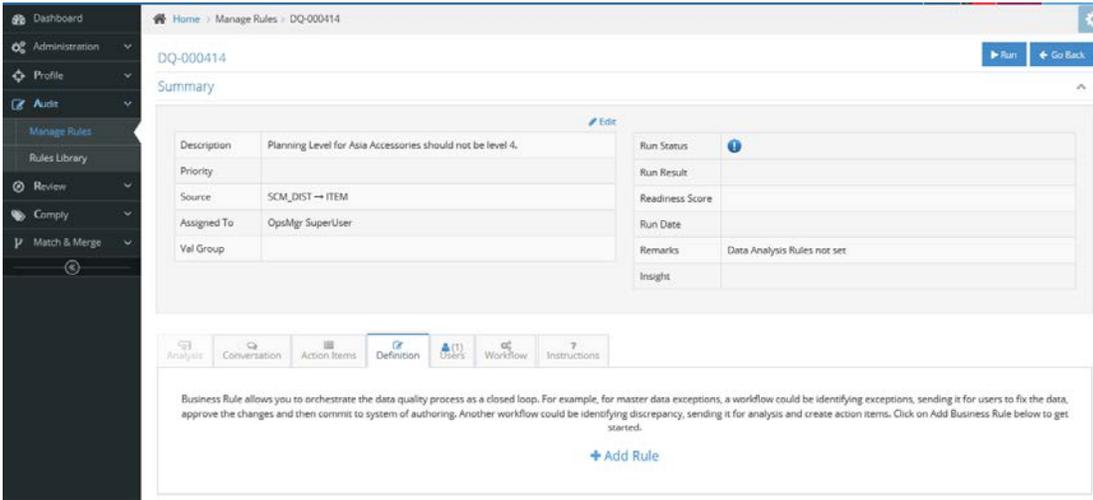
You may also simply choose to click, "Add Rule" which takes you to the Rule Editor. Here you will need to recreate the filtering conditions for the Rule you want to create and click, "Add Rule" which notates the condition under the "Exceptions Rule List."
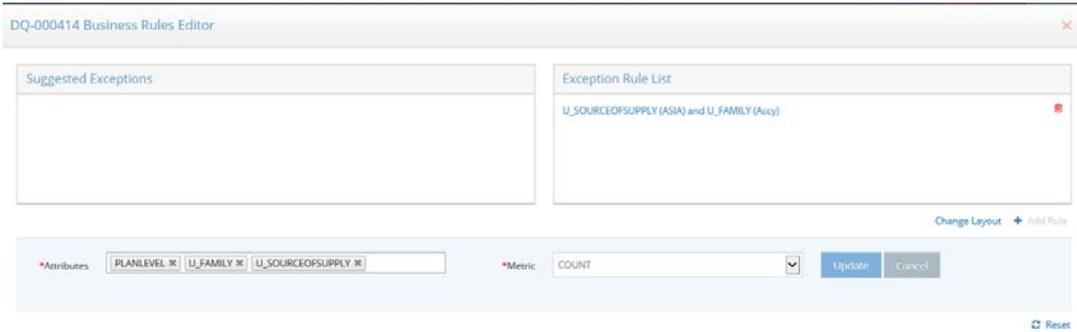
When creating Rules from this view, you are able to create multiple filtering conditions, within the selected Attributes, and also add them as Rules. This creates a compound Rule that returns all qualifying Exceptions or records within the query, and allows you to view them together.

3. Reviewing Exceptions & Fixing Data

After a Rule or Audit is run, you may be responsible for reviewing and fixing data exceptions. After each Run, the status, number of Exceptions and trend for the Rule is provided. This helps prioritize where to focus, and alerts you to other actions that may need to be taken.

3.1 Navigating to Data Exceptions

If you log into the system as part of your process, you can simply click on Exceptions from your Dashboard. This navigates to Review > Analyze Rules, where the list of your Rules and their Status can be seen. The specific records and details for the Exceptions can be found by clicking the specific Rule.
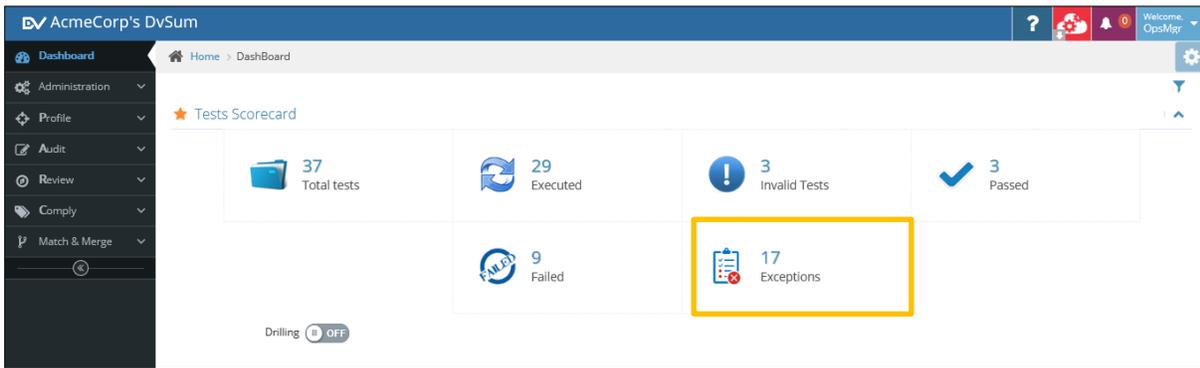

You may also receive an email alert when an Audit is run, which includes a link to the Rule or Ruleset you are being asked to Review.
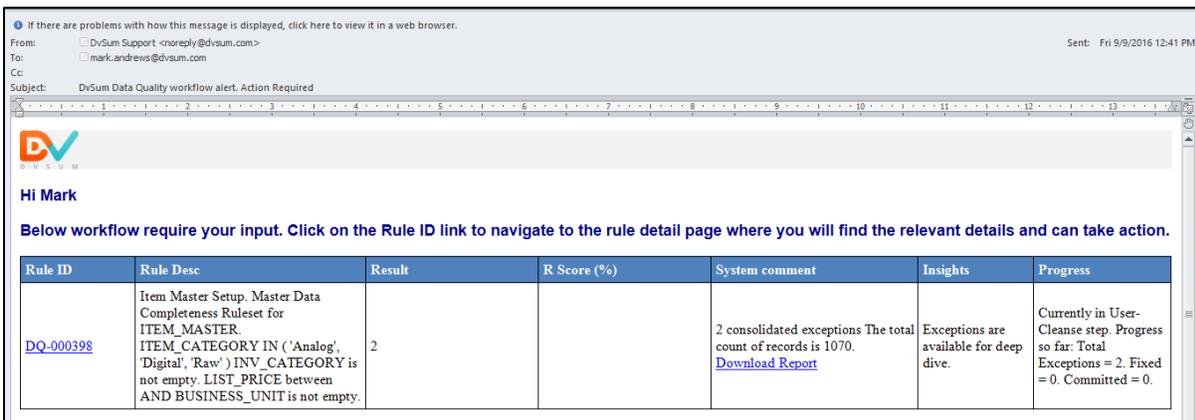
At any time during your process, you can also navigate to Review > Analyze Rules, to view the latest statuses for the Rules you're responsible for, and note which ones have remaining exceptions based on the Status icon.

From this view, you can prioritize which Exceptions to review, based on the High Priority Exceptions notated with the red dot, or based on the number of Exceptions under Result, or the trend indicated under Insights.
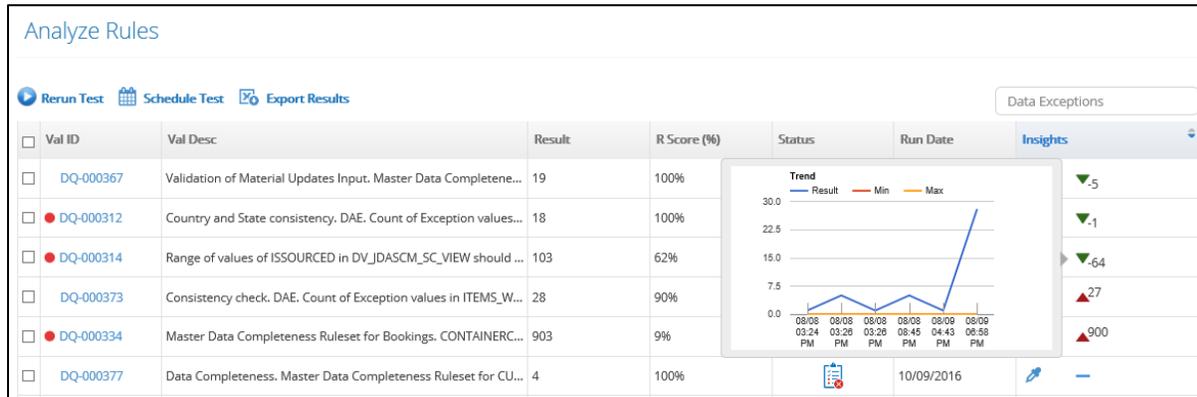
3.2 Analysis of Rule & Audit Runs

Once the Audits are executed, information is provided on the Status, Result and trends associated with each Rule. The Insights column provides additional information to help with drilling-down to the specific data quality failures or exceptions and to identify the root-causes.

2.2.1 Audit Run & Rule Status

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

Icon	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	Audit rule is not of type master data and the rule is failing	Run over run variance of qtyplanned in PURCHASE_PLAN is 5% for Supplier X which is more than the limit of 3%
	Exception	Audit rule is of type master data and there are 1 or more exceptions	There are 2 routing records where yield is not between 0 and 1

2.2.2 Run Status & 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 Result	Audit Result Icon	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	FAILED	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	FAILED	54% (3000 is 46% short of 5500). Same test as above, but data quality is lower.
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%	FAILED	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)

Following insights are available with your audit results

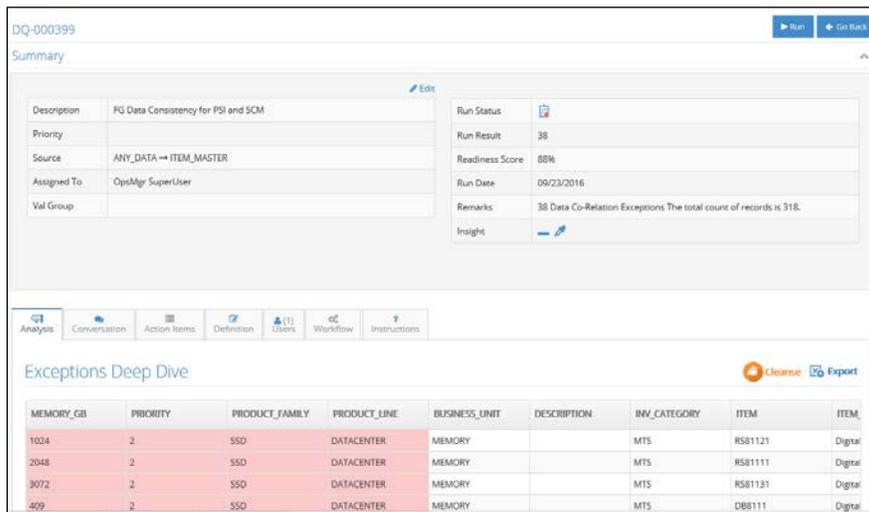| Icon	Insight Type	Definition
↻	History	History for all audits
🖊	Exception List	List of Exceptions for master data audits
⊥	Drill-down Analytical	Drill-down with Variance for aggregation audits

2.2.3 History Trend 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 or ▼105	**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	Things are improving even if the tests are failing. May not need to focus.

	high and now trending down.	
▼-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

2.3 Review Exception Detail

By clicking the specific Rule under Analyze Rules, or clicking the link within an alert, you will be taken to the unique Rule page, which shows the Rule Definition, Summary from the latest Audit run, and trends. Data Exceptions are highlighted with the individual record details.

2.4 Fixing Data Exceptions

To fix data exceptions highlighted under the Analysis tab, click the Cleanse icon which brings you to the cleansing workbench and allows you to click and edit the individual fields.

Depending on the field and type of Exception, edits may be free text entry, or selected from a pick list. These changes can be applied to the individual record field, mass updated to the same Exceptions within other Records in the view, or Applied as a Rule, to auto-correct all future Exceptions of this kind with the updated value.

Edited fields are highlighted in blue and tracked within the Change History box. Clicking the Save button saves the changes locally and then enables the Commit button.

2.5 Commit Changes: Options

Depending on your configuration, the Commit options may include, Email which sends a link to the changes, Write-back to Source, or if a workflow is enabled, the next user in the process is alerted that they have updates to review.

4. Data Profiling & Exploration

Data Profiling and Exploration allows you to discover 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 to discover more information, identify missing relationships, profile the data, and 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.

4.1 Data Profiling

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.

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

You can access Data Profiling from Left Navigation, and then selecting the Source and Tables you want to view, of simply selecting a View from your history that is automatically displayed. Profiling information is also available when reviewing nodes in the Data Exploration section.

The initial Table detail is a good view to identify basic data exceptions such as Blanks, or Values that may be out of the expected or desired range. As outlined above, Rules can be created within Profiling by clicking on the Data Quality and Data Analysis tabs.


Selecting the specific Column Name displays the distribution of the data to enable further analysis.
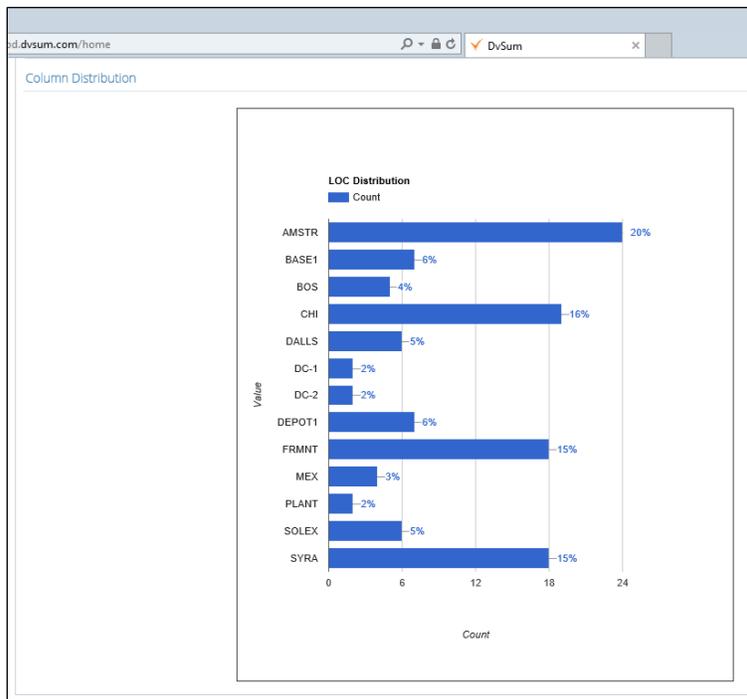
4.2 Data Analysis

Within Profiling, the Data Analysis tab provides a view to quickly explore the data. Select the Attributes and Metrics you want to view and then click on the boxes for quick filtering of the data. This enables you to discover counts of records and data combinations that are unexpected and warrant further exploration. Based on your business process and setup, you may see invalid or undesired combinations of data that warrant creating a Rule to generate the detailed Exceptions, as outlined above.
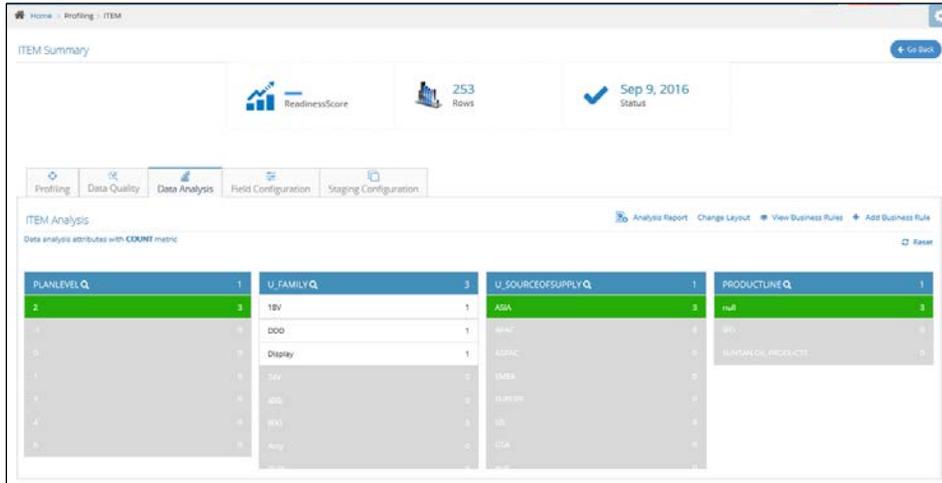
4.3 Data Exploration

Data Exploration is a workflow 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 statistics, identify missing relationships, synonyms and integrity across systems. Exploration can assist with

- Audit a particular data element within the company for regulation requirements. Find which databases, which schema where 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