USER GUIDE

Data Profiling

Release 10.0-MP3 (October 2020)

StiboSystems

MASTER DATA MANAGEMENT

Table of Contents

Data Profiling

STEP offers a set of tools dedicated to measuring and improving the quality of data in the system:

- **Data Profiles** - The data profile functionality provides detailed overviews of data in specific branches of the hierarchy in Tree. Each profile contains information about the profiled data and provides easy access to correcting data errors. See the **Data Profiles** topic.
- **Attribute Value Profiles** - Attribute value profiles enable users to see where and how an attribute is used. The profile provides an overview of how attribute values are distributed across the current context and workspace, and enables users to easily clean up data. See the **Attribute Value Profiles** topic.
- **Metric Score Indicators** - Metric score indicators visualize the quality of profiled objects based on scores users set for attributes, reference types, and product-to-classification link types. See the **Metric Visualization** section of the **System Setup** documentation.

Some features are privilege controlled. For more information see **Action Sets** in the System Setup / Super User Guide documentation.

Important: Users should be very cautious of having calculated attributes in the attributes / attribute groups being profiled. Calculated attributes are calculated when included in profiling, and this may potentially impact system performance.

Data Profiles

The Category / Data Profile functionality provides a detailed overview of data in a specific branch of the hierarchy in Tree. Each profile contains information about the profiled data and provides easy access to correcting data errors.

Note: For entities, classifications, and collections, these profiles are called 'Data Profiles.' For products, they are called 'Category Profiles.'

With a data profile, users can view information about the profiled data such as:

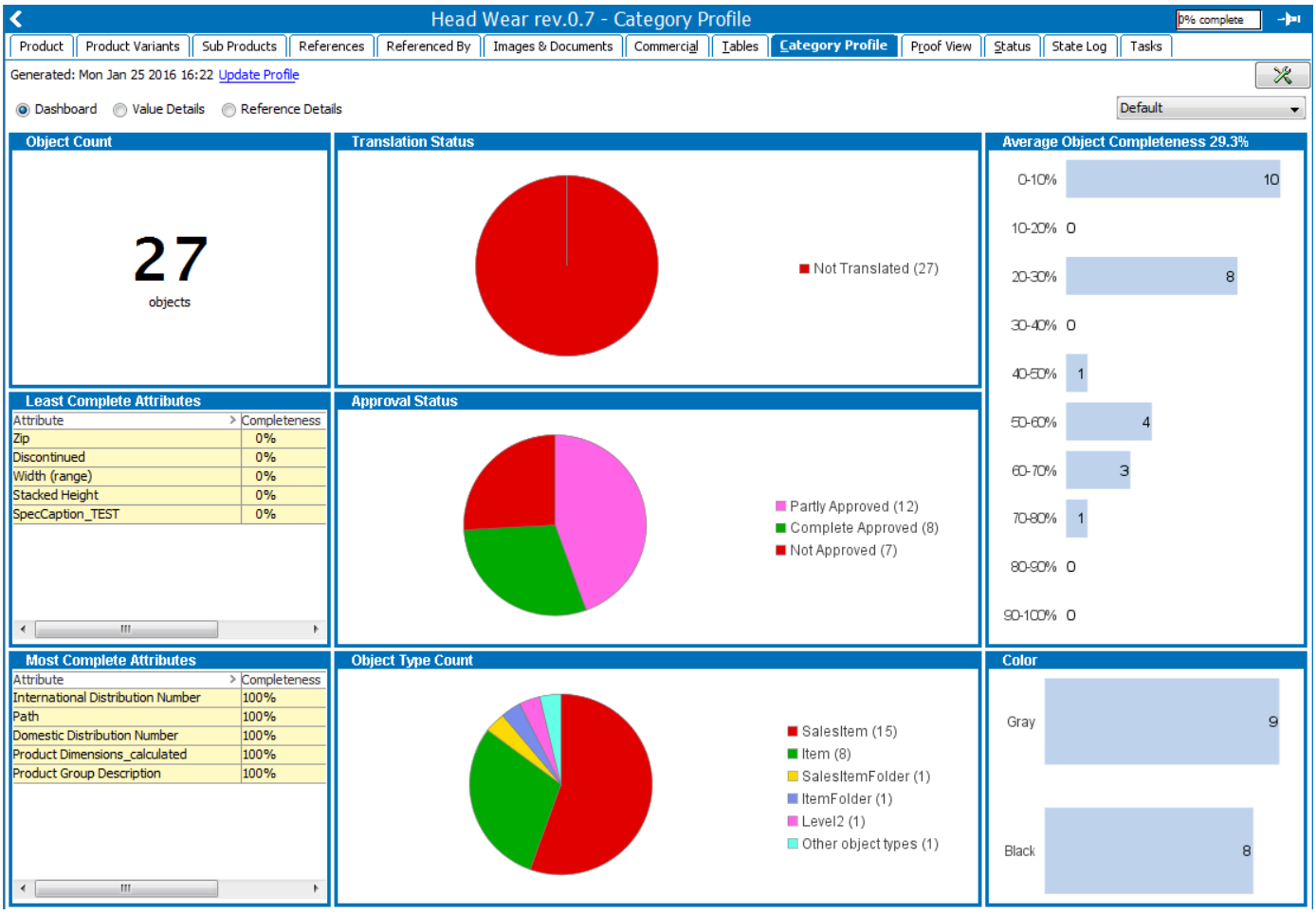
- The degree to which the attributes used in the hierarchy are populated.
- The most / least frequently used values for a given attribute, value ranges, used characters, and more.
- The approval and translation status of the profiled objects.
- The degree to which reference types are populated.
- The average product completeness (the degree to which important attributes, references, and product to classification links exist and have values.)

These profiles are generated on demand, and are context and workspace specific.

For an example of using data profiles, see the **Configuration Example - Basic** topic in the **Matching, Linking, and Merging** documentation.

View Data Profiles

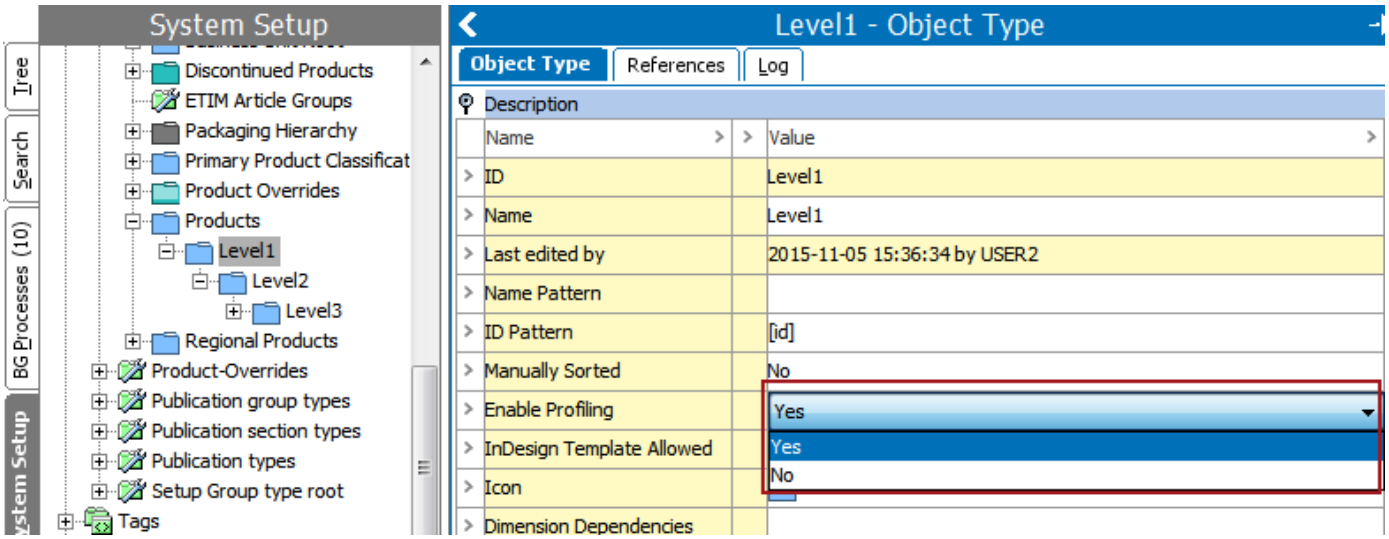
1. In the **Tree**, select the relevant hierarchy, and then click the **Category Profile** / **Data Profile** tab.
2. If no profile exists, click **Generate Profile**, otherwise click **Update Profile**



Enable Profiling on an Object

Objects can be flagged for profiling via object type configuration.

1. In **System Setup**, expand **Object Types & Structures**.
2. Select the relevant object type, then click the **Object Type** tab. Next, in the **Enable Profiling** field, select **Yes**. Note that you can only make changes if you have the proper privileges. For more information see the **Action Sets** section of the **System Setup / Super User Guide** documentation.

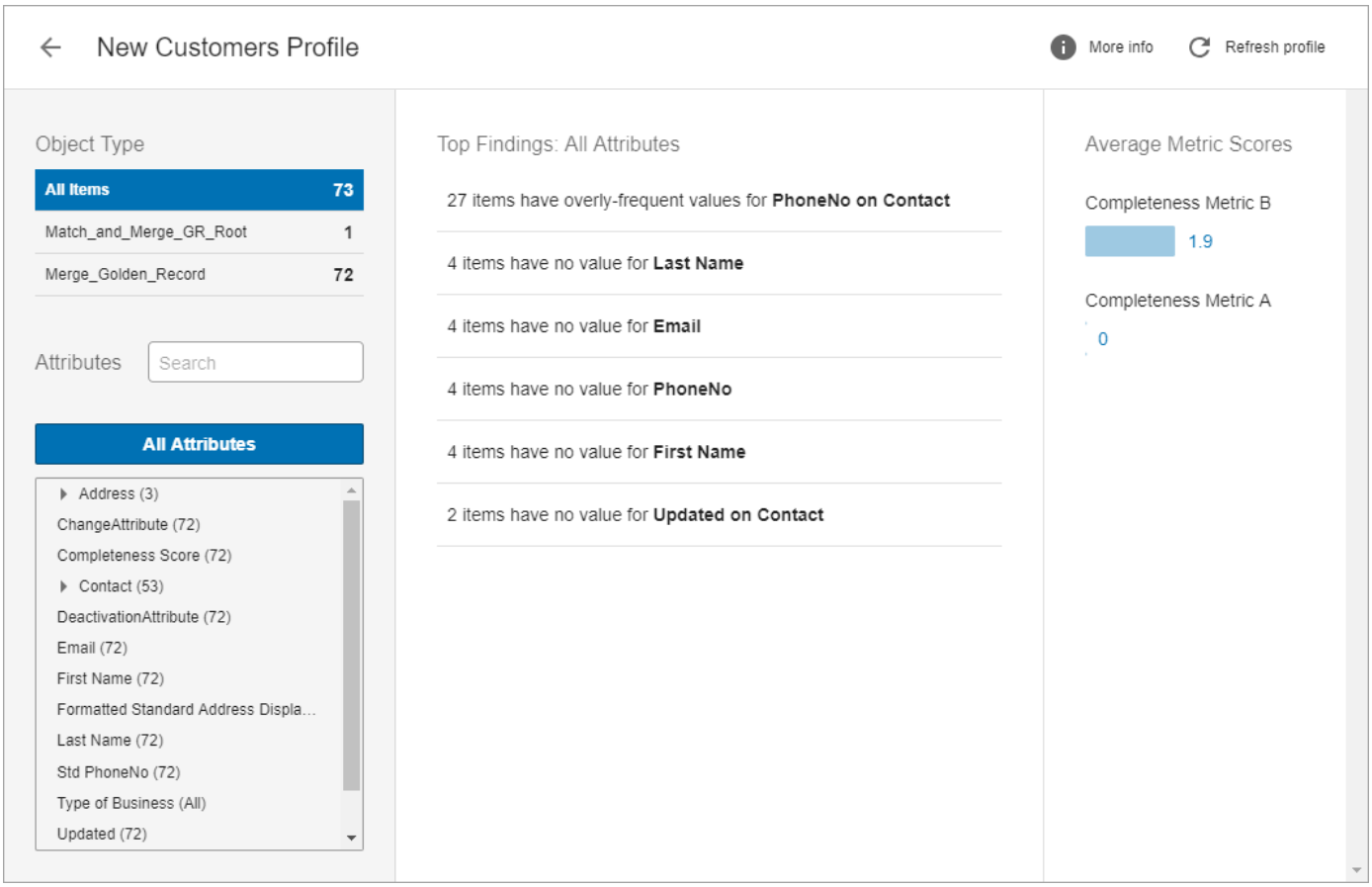


Because this configuration is made via object type, data / category profiles can be made on any level in the hierarchy.

Note: Information about product override objects is not included in category profiles.

Data Profiles in Web UI

Data profiles can also be viewed in the Web UI via the **Data Profile** component screen. This screen displays much of the same information found in the workbench and provides a list of 'Top Findings.' The 'Top Findings' list highlights outliers in the data such as missing attribute values, irregular value patterns, overly-frequent values, and more.



For more information on using data profiles in Web UI, see the **Data Profiling in Web UI** section of the **Data Profiling** documentation.

Generating, Updating, and Scheduling a Data Profile

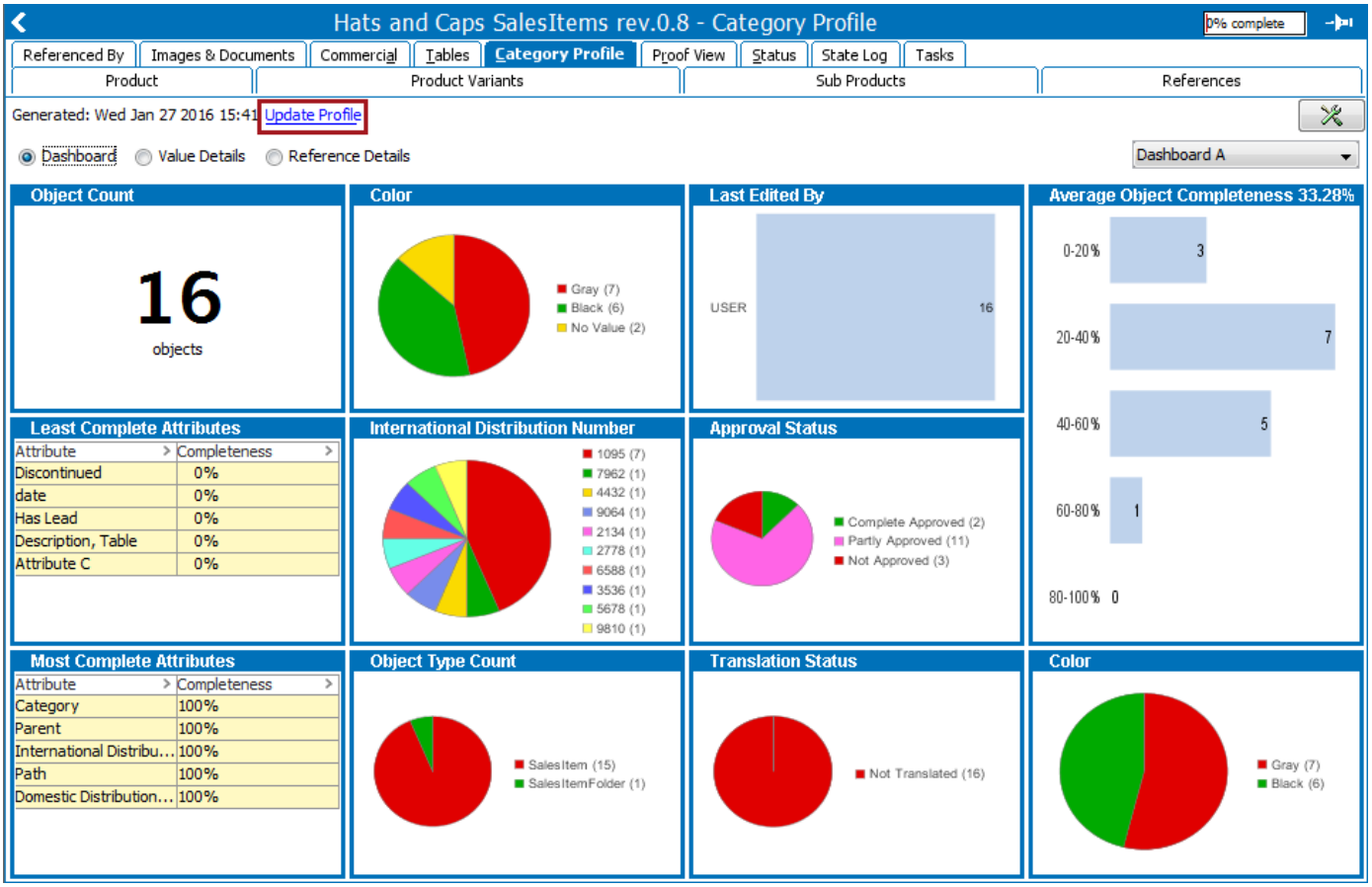
Before the data in a category / data profile can be viewed a profile must be generated. Generating a profile takes a snapshot of the data and does not actively monitor the profile for changes. Because of this limitation, profiles must be manually updated whenever changes to the data are made (or a scheduled process must be configured).

Note: Only users with the proper privileges are able to generate or update a category profile.

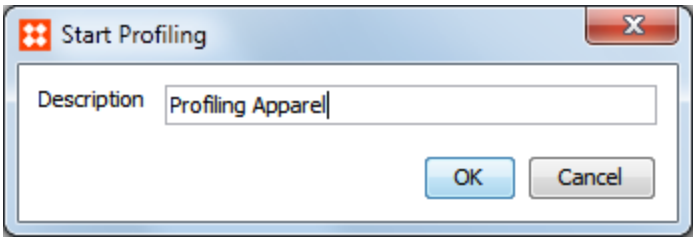
Generating / Updating a Data Profile

1. In the **Tree**, expand the applicable hierarchy, and then, select the relevant object group or object.
2. On the **Category / Data Profile** tab, click the **Generate Profile** link. The **Start Profiling** dialog appears.

   If updating an existing profile the link is called **Update Profile**.



3. Click **OK**. A background process starts. Once finished, the profile is displayed.



Important: The background process runs using super user privileges, and the profiles are identical regardless of who started the process. Also, if a user has profile viewing privileges, all data is displayed in the profile no matter which view limitations otherwise apply to the user. By configuring the privileges, it is possible to control who can view, configure, and update profiles.

Scheduling Data Profiling

Category / Data profiles can be scheduled to update automatically using the **Schedule Category Profile Update** wizard.

In the **File** menu, select **Schedule Category Profiling** to open up the wizard:

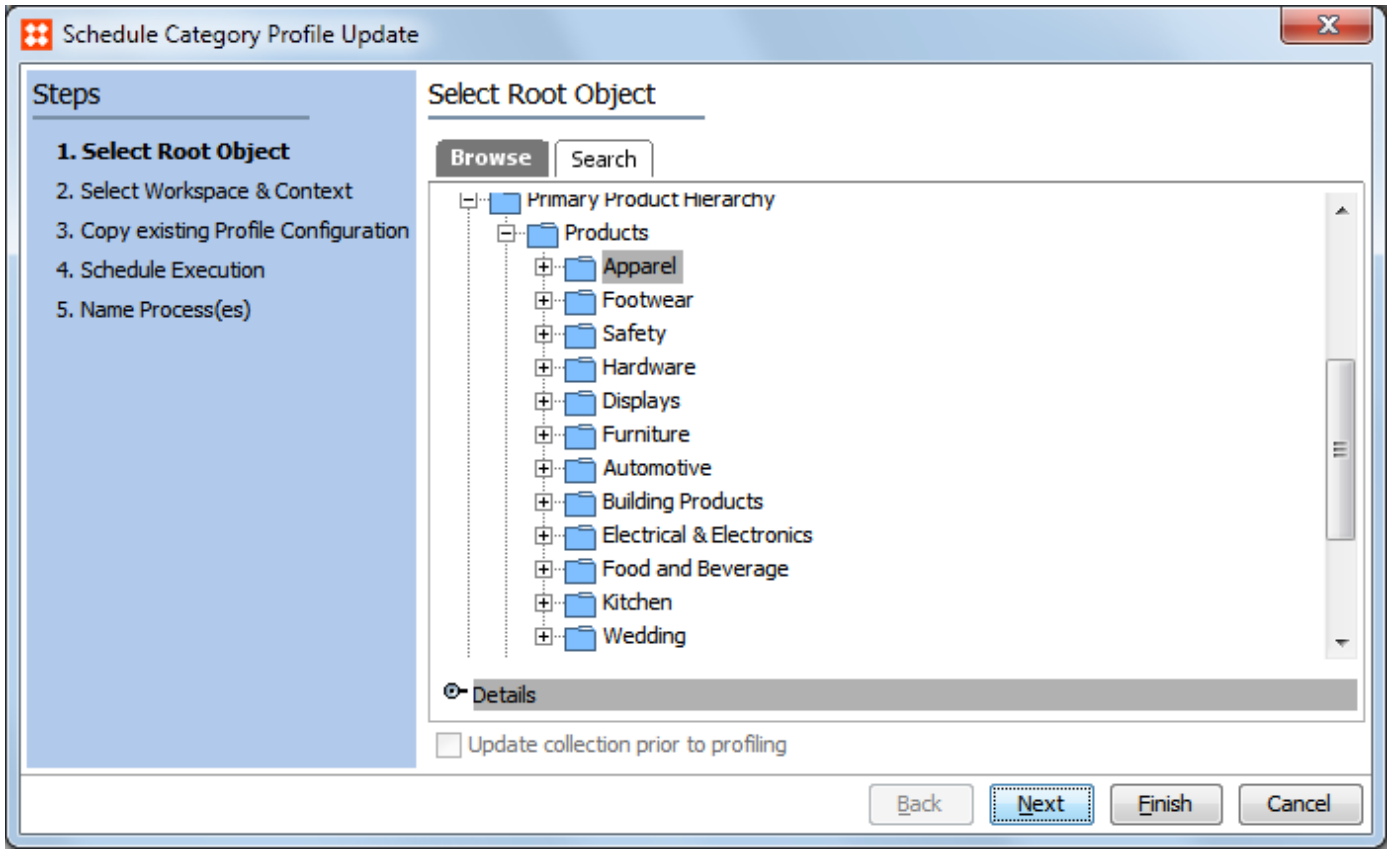
1. **Step 1 Select Root Object** - Browse or search for the preferred root object, and then, click **Next**.

Note: All category / data profiles underneath the selected node will also update.

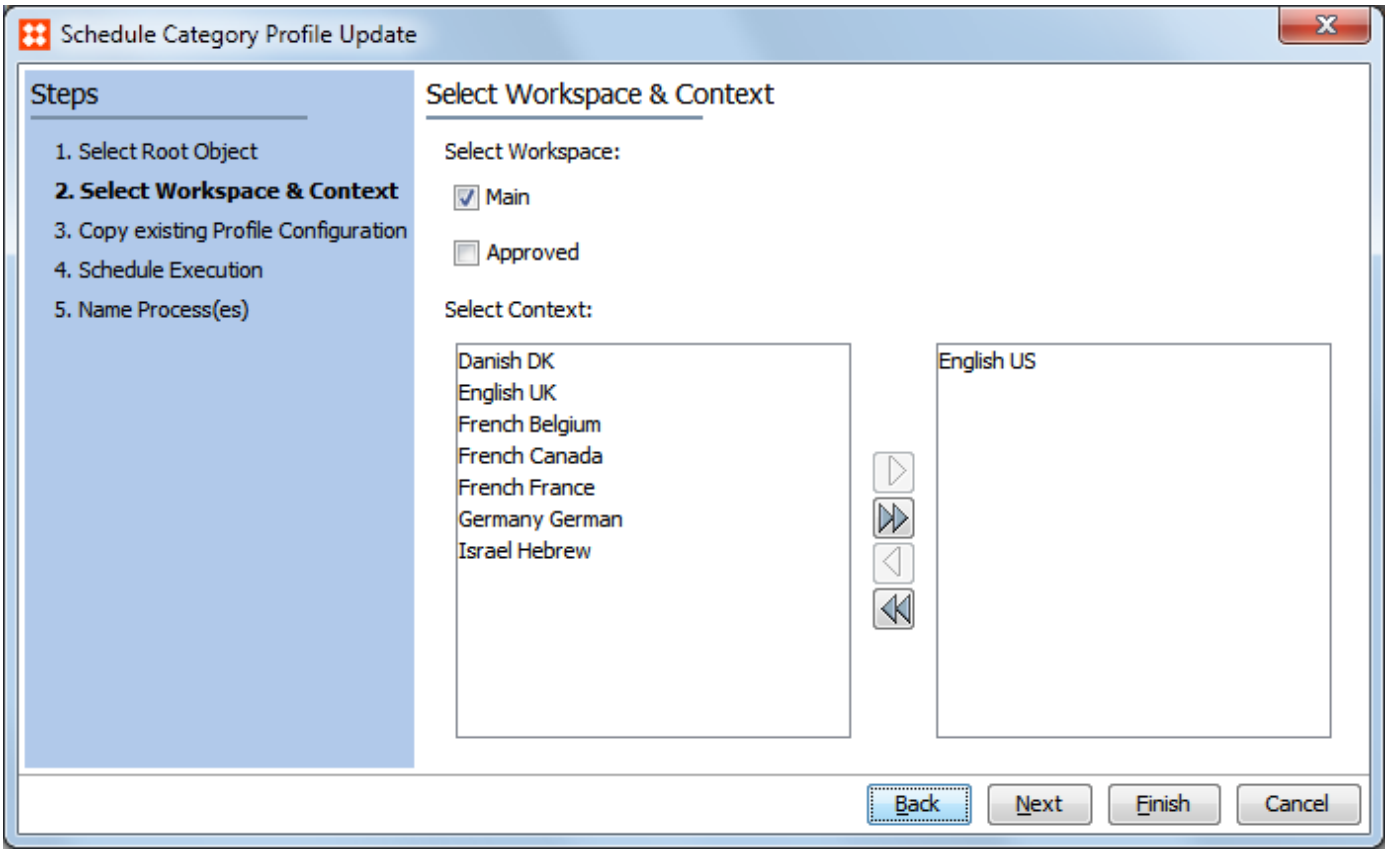
Valid selections include products, entities, classifications, and collections. If a collection is selected, profiles are generated or updated for all objects in the collection for which profiling is enabled.

Important: When a profile is generated via a scheduled event the objects within the collection are updated, not the collection itself. If you intend to generate profiles for the collection objects themselves (not just the products, entities, etc. within), the collection object in question must be included within another collection. Scheduling a profile update for the collection containing this collection object will update the collection object and the data within it (but not the collection containing it).

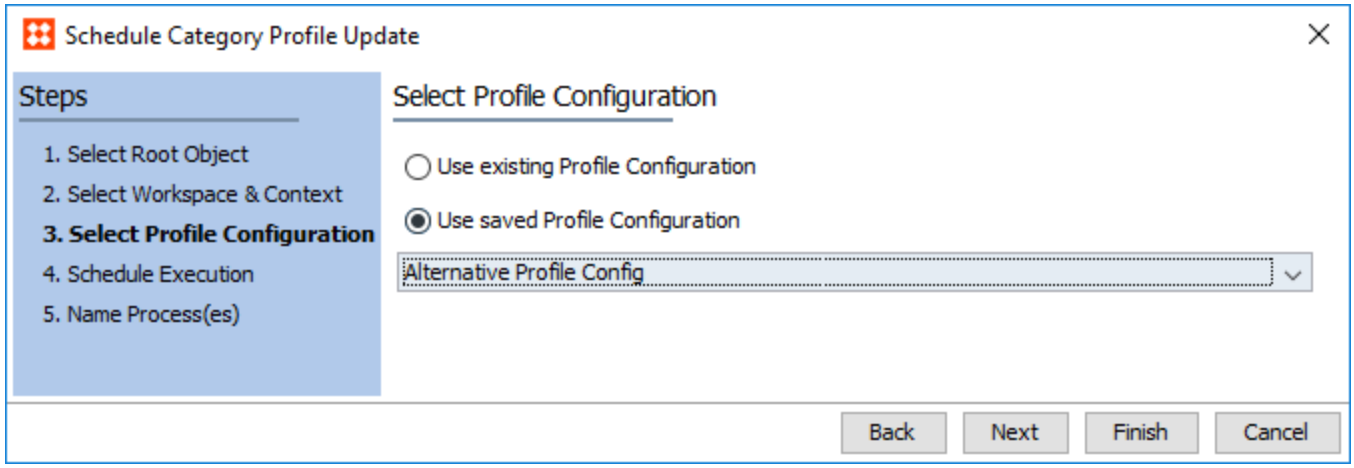
If the collection is search based, and you want the collection to be refreshed as part of the processing, select the **Update collection prior to profiling** checkbox.



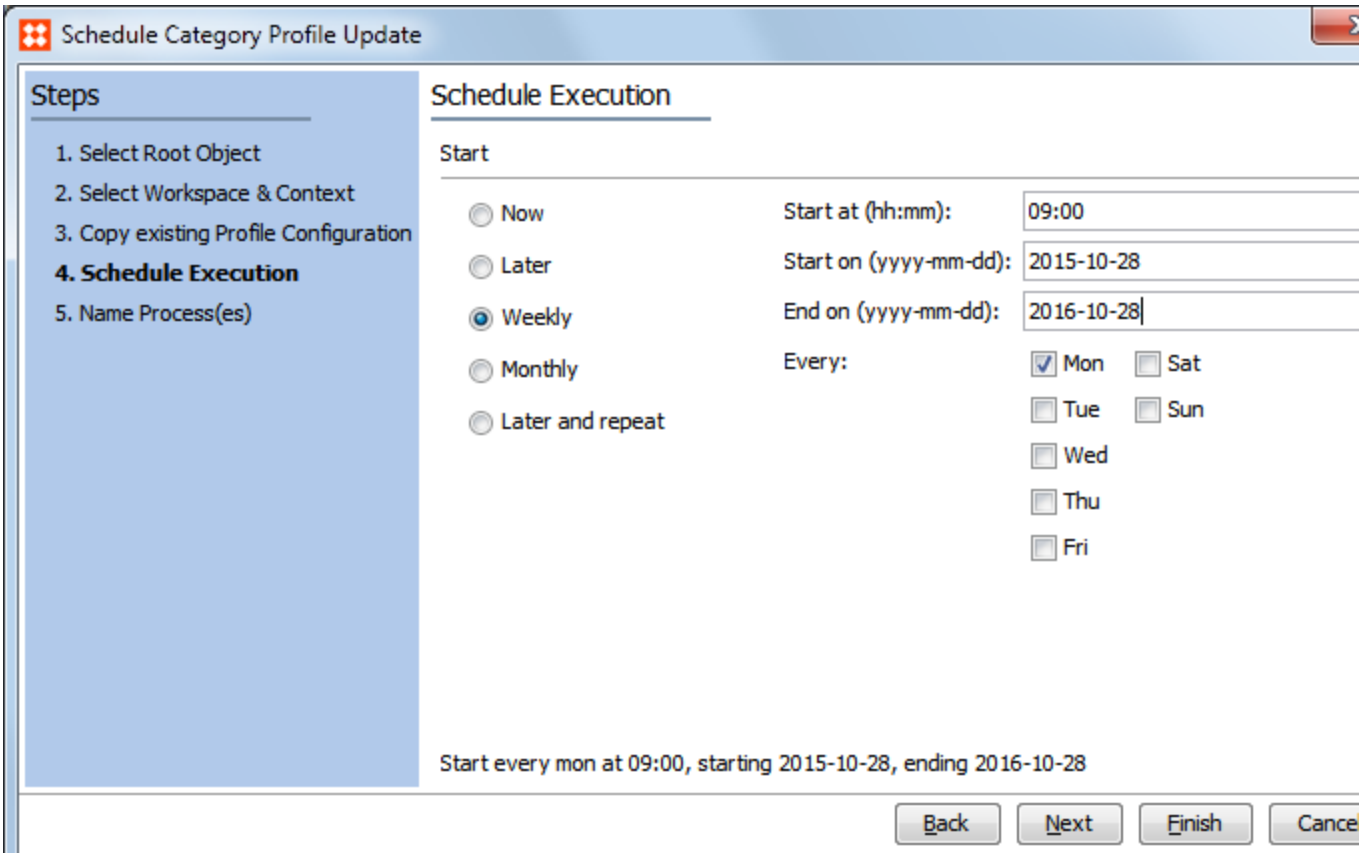
2. **Step 2 Select Workspace & Context** - Select the workspaces and contexts that the profile applies to. This allows the scheduled update to apply to all relevant context and workspace combinations.



3. **Step 3 Select Profile Configuration** - Specify which profile configuration to apply to the data profiles. If **Use existing Profile Configuration** is chosen, all data profiles will use whatever configuration is set on the node. If **Use saved Profile Configuration** is chosen, use the dropdown to specify which saved profile configuration to apply to all data profiles. The selected configuration overrides the configuration set on the node(s) in question.



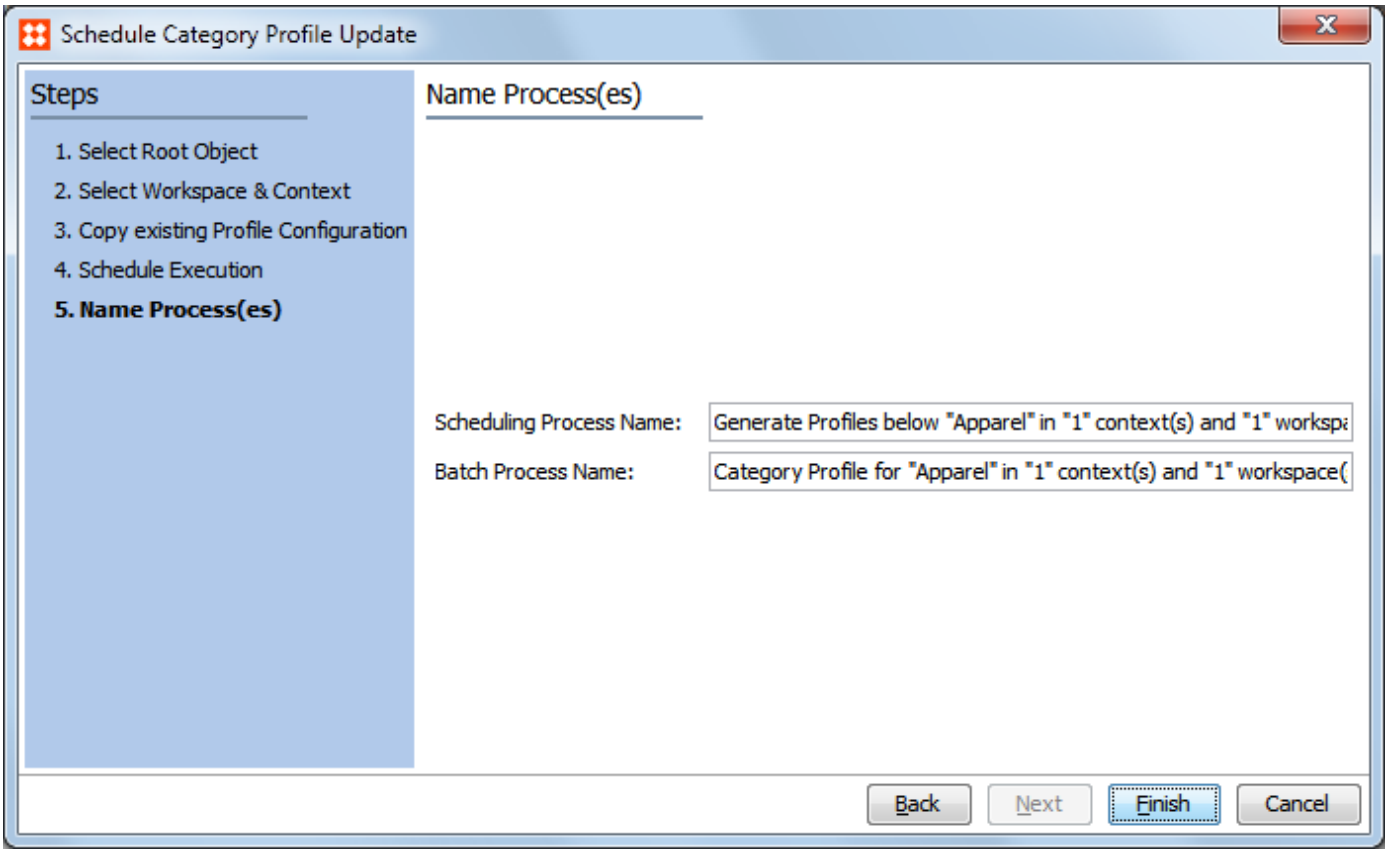
4. **Step 4 Schedule Execution** - Define the preferred schedule. You can specify whether you want the process to run immediately, at a later point in time, or as a recurring weekly or monthly process.



Important: Consider the time zone of the application server compared to that of the workbench (the client) where the schedule is created or viewed. When scheduling a job, the local time zone is displayed in the workbench, but the time zone of the server is used to run the background process. Although displayed, the time zone of the client is not included in the instruction to the server to run the job. This can cause confusion about when the job will run since the scheduled time is not automatically converted to accommodate potential differences in time zones.

Note: If profiling a collection, selecting 'Now' will execute the profile immediately and a profile will be generated for the collection itself alongside all the objects in the collection.

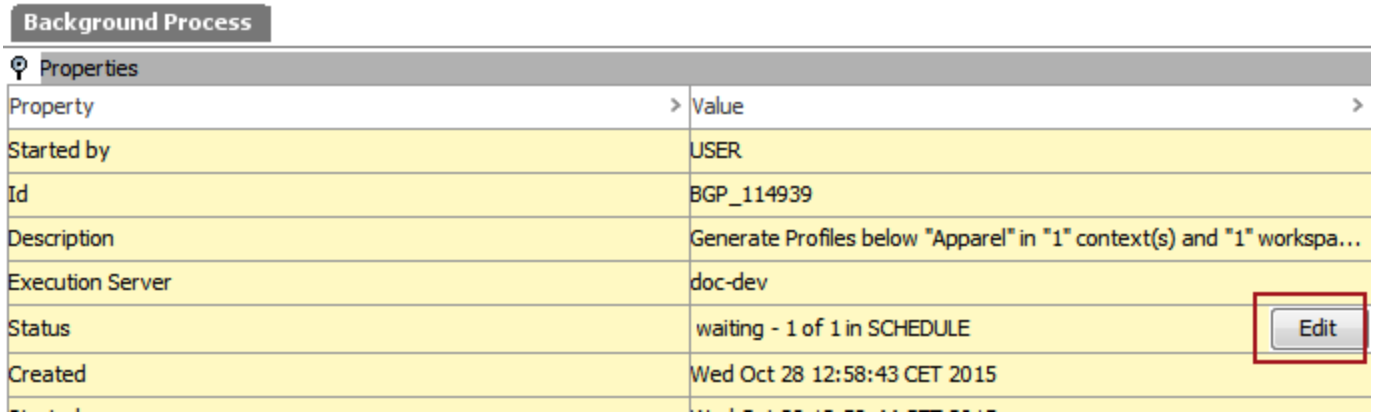
5. **Step 5 Name Process(es)** - Give the process a name, and then click **Finish**. A background process is started.



Access the Schedule Category Profile Update Wizard from the Background Process

It is possible to edit the profiling schedule from the Background Process.

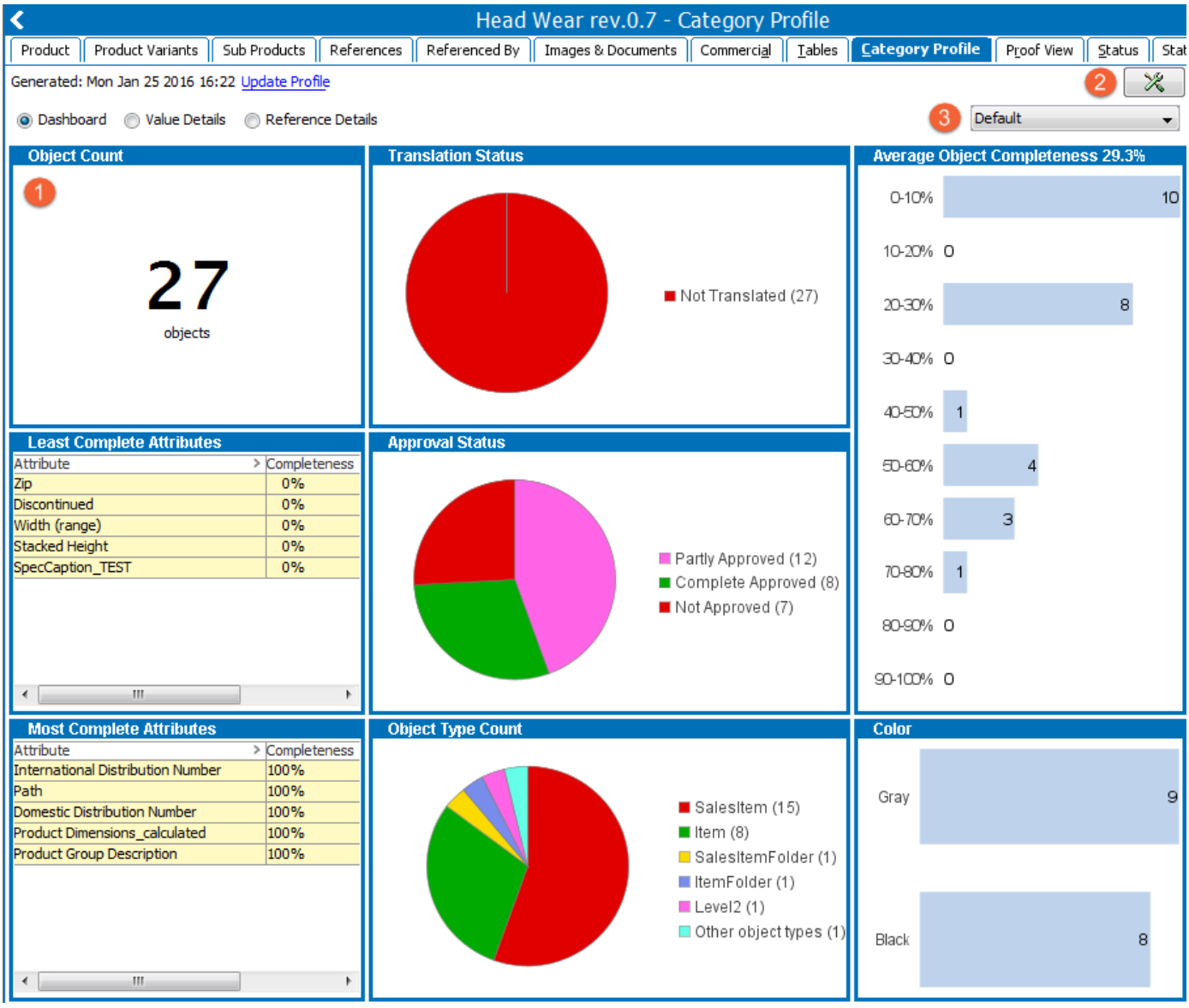
1. On the **Background Process** tab, locate and select the relevant background process.
2. In the **Status** property **Value** field, click **Edit**. The **Schedule Category Profile Update** wizard opens.



Data Profile Dashboard

The category / data profile dashboard is a highly customizable view that allows users to display profiled data using a collection of configurable widgets. Each profile can have its own set of dashboards, allowing users to group widgets together in a logical manner, and to create different views for specific user roles or tasks. Dashboard configurations are tied to the profile but are shared across all contexts and workspaces.

Pictured below is an example of a Category Profile Dashboard:



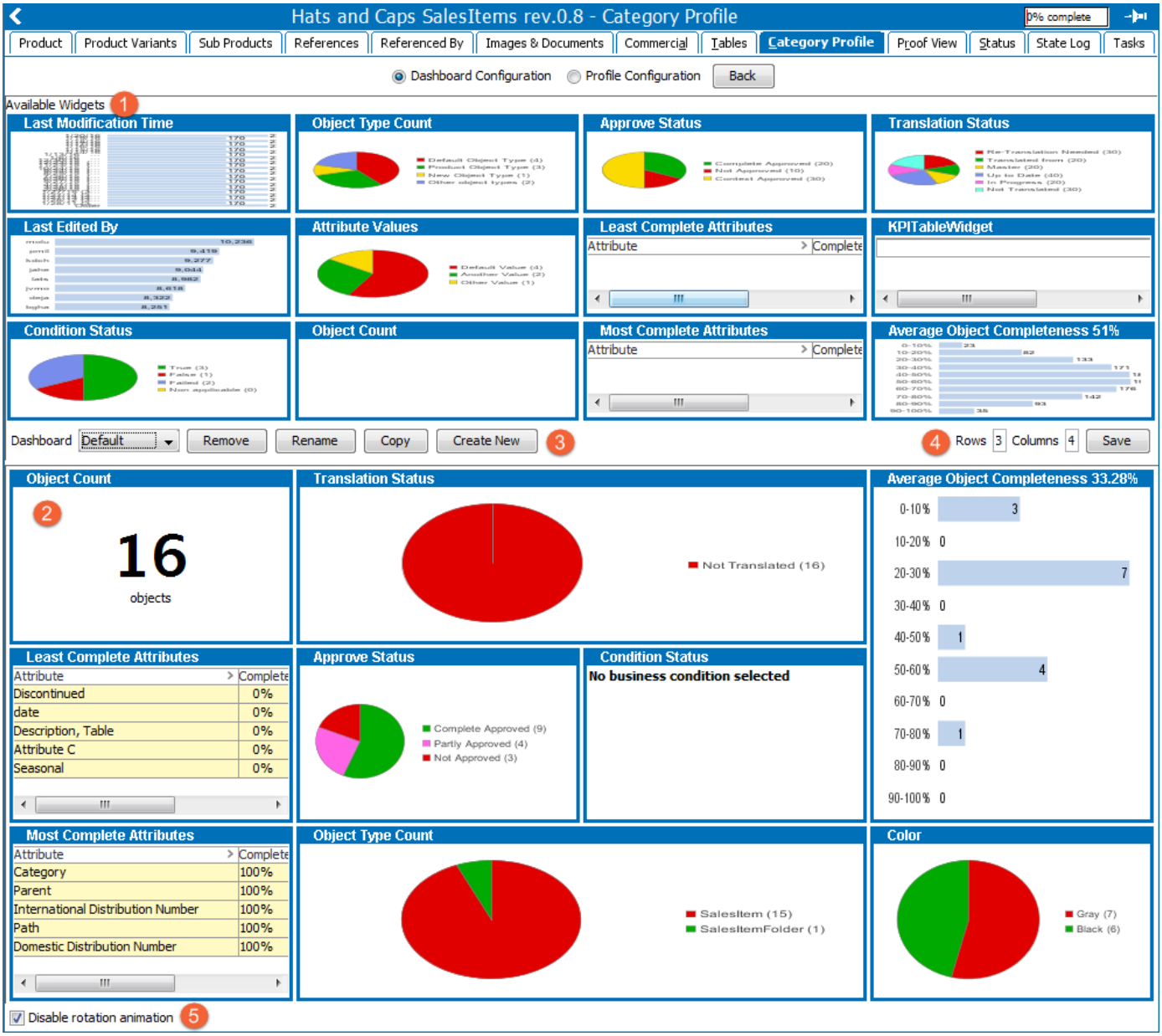
1. This area displays the currently selected dashboard.
2. Clicking the  **Tools** button allows users to access the **Dashboard Configuration** and **Profile Configuration** views.
3. Users can toggle between dashboards using this dropdown.

Changes to a category / data profile dashboard are made via the **Dashboard Configuration** view, as detailed below.

Accessing the Dashboard Configuration

1. In the **Tree**, select the relevant object hierarchy or object.
2. On the **Category / Data Profile** tab, click the  **Tools** icon in the upper right corner, and then select **Dashboard Configuration**.

Navigating the Dashboard Configuration



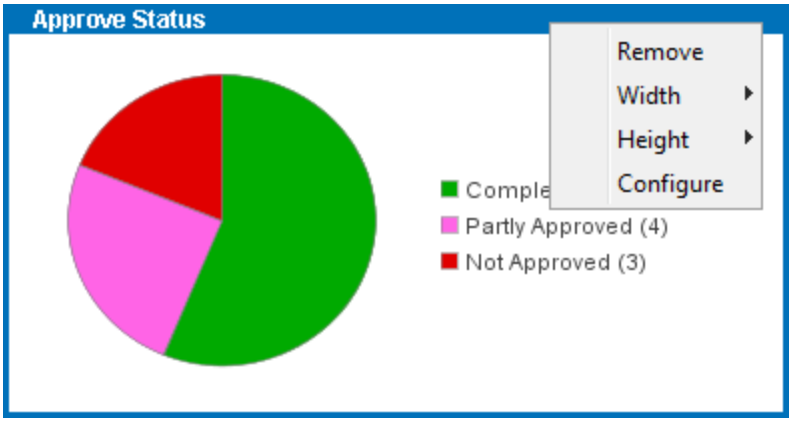
1. The top area of this view contains all **Available Widgets** for the dashboard. These widgets can be added to the currently selected dashboard by left-clicking the title bar and dragging the widget to an empty space in the dashboard preview area (the lower half of the screen).
2. The bottom area of this view contains the **Dashboard Preview**. Once a widget is added to this area, they can be configured by right-clicking the title bar and selecting the desired configuration option. Further details on configuring the widgets are detailed below.

Note: The data shown in the **Available Widgets** area is sample data. The **Dashboard Preview**, however, displays correct data if the category has been profiled.

3. Different dashboards can be configured using the buttons found in-between the **Available Widgets** and **Dashboard Preview** areas on the left:
   - The **Dashboard** dropdown list allows users to select whichever dashboard they want to display / edit. Note that this dropdown selector is also available in the Category / Data Profile Dashboard view, underneath the **Tools** icon, and is used to toggle between dashboard displays.
   - The **Remove** button allows users to delete the currently selected dashboard.
   - The **Rename** button allows users to rename the currently selected dashboard.
   - The **Copy** button allows users to copy the currently select dashboard, including all of its widget configurations. If clicked, a naming prompt will appear. Once the copied dashboard is named, it will automatically appear as an option in the **Dashboard** dropdown list.
   - The **Create New** button allows users to create a new dashboard for the currently selected profile.
4. In the **Rows** and **Columns** fields, found in-between the **Available Widgets** and the **Dashboard Preview** on the right-side of the view, the user can specify how many rows and columns the currently selected dashboard should have.
5. Check the **Disable rotation animation** checkbox to disable the animation that occurs when toggling between the Category / Data Profile Dashboard and the Dashboard Configuration views.

Configuring Dashboard Widgets

Once added to the Dashboard Preview, users can configure the size and shape of their widgets, as well as filter what data appears. To access these configuration options, right-click the title bar of the desired widget and a list of options will appear.



Clicking **Remove** will delete the widget from the currently selected dashboard. Clicking **Width** or **Height** will display a list of values - the higher the value selected the taller / wider the widget becomes. Note that increasing the size of a widget may push other widgets out of view if there is not enough space. As mentioned above, the **Rows** and **Columns** fields dictate how large the dashboard can be.

Clicking **Configure** displays a configuration window with additional options. These options vary between widgets, but common configuration options include: widget title, what context to display data from, what workspace to display data from, filtering data via object type, sorting order, and chart type. For more details on which options are available to each individual widget, see the **Available Widgets** section in the **Data Profiling** documentation.

Saving the Configuration

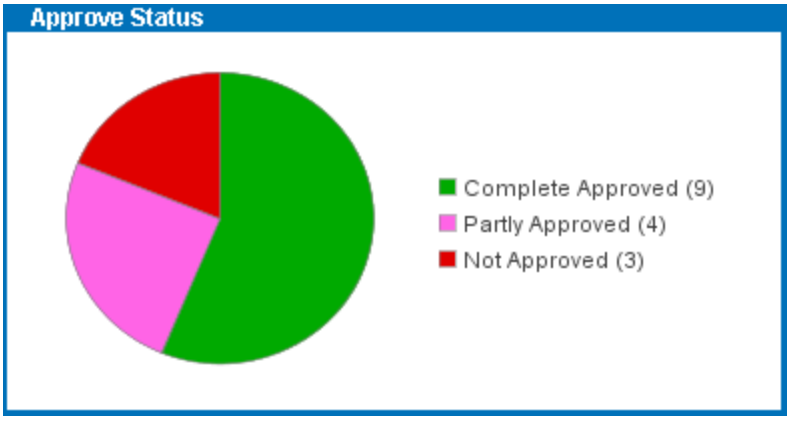
1. Once the dashboard has been configured, click **Save**, located next to the **Rows** and **Columns** fields, then click the **Back** button at the top of the screen to return to the category / data profile.
2. Click **Update Profile** for the changes to take effect.

Available Widgets

All available category / data profile dashboard widgets are detailed below. These widgets are also available in Web UI unless noted otherwise. For more information about configuring dashboard widgets in Web UI, see the **Data Profile Widgets in Web UI** section of the **Web User Interfaces** documentation.

Approve Status

This widget displays the approval status of objects in the data profile.

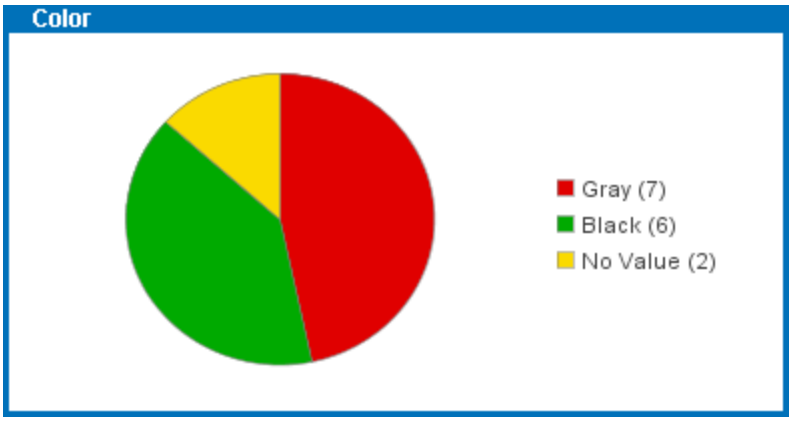


Configuration options include:

- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.

Attribute Values

This widget shows the distribution of values for the specified attribute across objects in the data profile.



Configuration options include:

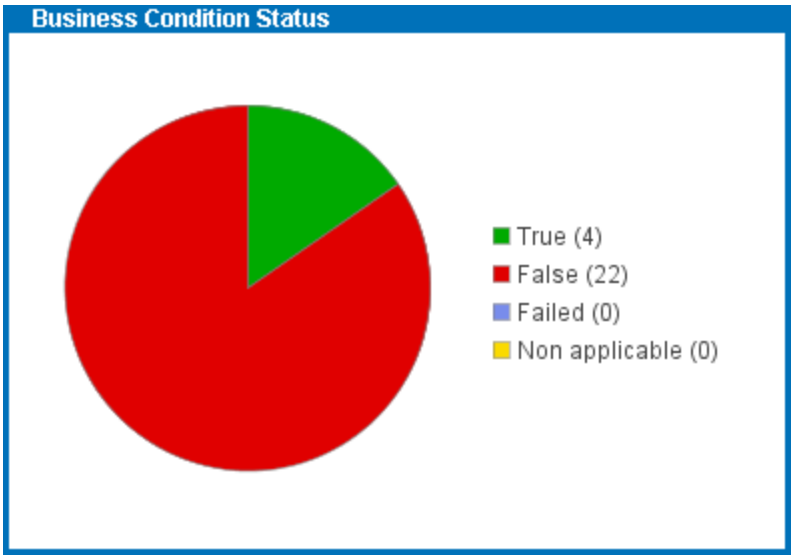
- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.
- **Attribute ID** - Specifies which attribute to pull values from. Defining this configuration is mandatory.
- **Chart / Dashboard Type** - Determines whether to display the data as a pie chart or bar chart.
- **Max Number of Values / Max Values** - Specifies the maximum number of distinct values to display. By default this field is set to -1, which means 'all values.' If the specified maximum is exceeded, the remaining (smaller) values will be placed in the "Other values" group.
- **Show No Values** - Displays the number of products that have no value for the specified attribute when the box is checked.

Important: Any attribute that is full text indexed, multi-valued, calculated, or has values with inline references is not compatible with this widget.

Business Condition Status

This widget displays the result of testing a business condition during profiling.

Note: If the test of the business condition results in an error, the test is reported as failed.



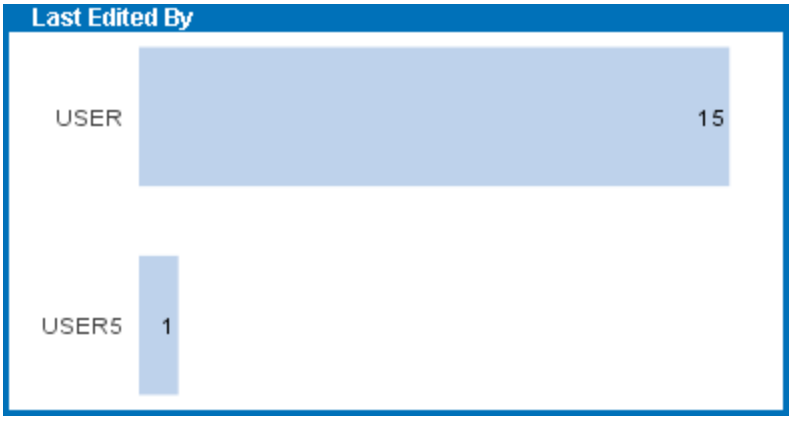
Configuration options include:

- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.
- **Condition** - Determines which business condition to test when profiling. Note that the condition must be specified in the Profile Configuration and data must have been profiled using that configuration before the widget can show any data. For more information, see the **Profile Configuration** section of the **Data Profiling** documentation.

Last Edited By

This widget shows which users last edited the object in the data profile.

Note: Not available in Web UI.



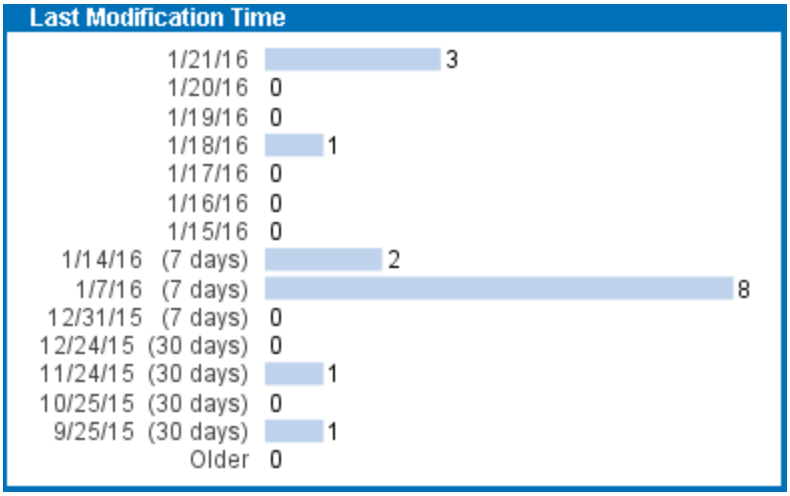
Configuration options include:

- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.
- **Sort Order** - Sorts the data in descending order by users with the most frequent edits or by users with the rarest.
- **Chart Type** - Determines whether to display the data as a pie chart or bar chart.
- **Max Number of Values** - Specifies the maximum number of distinct values to display. By default this field is set to -1, which means 'all values.' If the specified maximum is exceeded, the remaining (smaller) values will be placed in the "Other values" group.

Last Modified Time

This widget shows the last time each object in the profile was edited.

Note: Not available in Web UI.



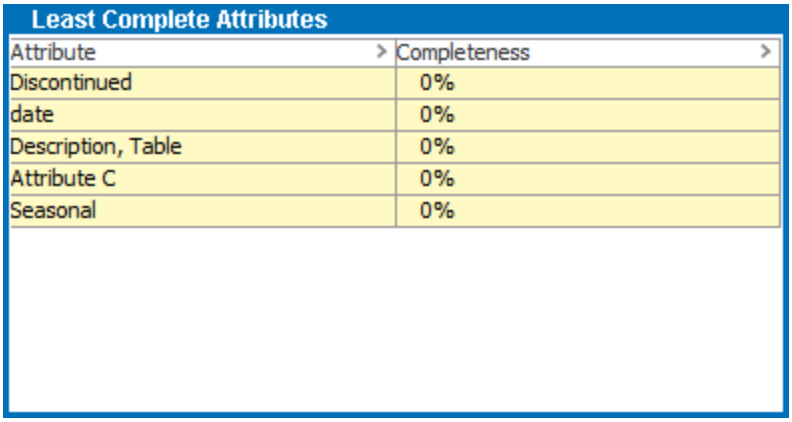
Configuration options include:

- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.

Least Complete Attributes

This widget displays the five least complete attributes in the data profile. Completeness is rounded to the nearest whole percentage and attributes with the same completeness are presented in the order in which they are retrieved from the database. Inherited values are included in the calculation.

Note: This widget has no relation to the Completeness Metric defined in the Profile Configuration and will evaluate all attributes in the data profile.

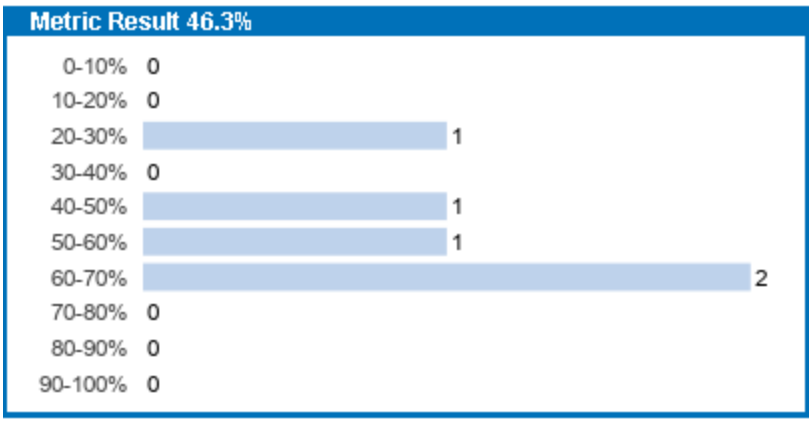


Configuration options include:

- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.

Metric Result

This widget displays the results of a specific metric applied to the data profile.



Configuration options include:

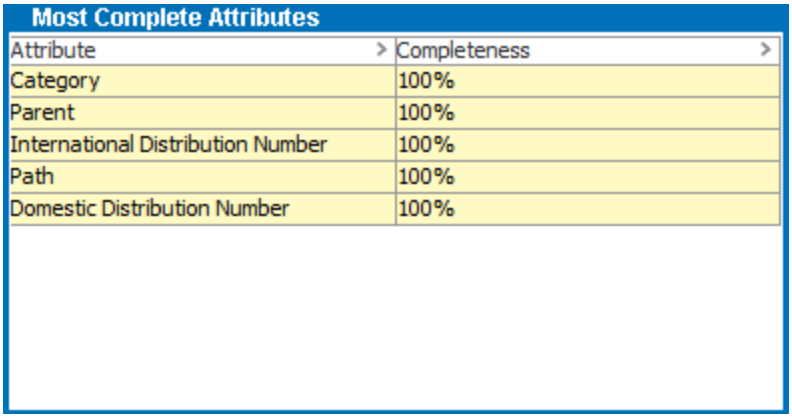
- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.

- **Metric** - Determines which metric to evaluate the data with. Note that the metric must be specified in the Profile Configuration and data must have been profiled using that configuration before the widget can show any data. For more information, see the **Profile Configuration** section of the **Data Profiling** documentation.
- **Intervals** - Displays the bars of the graph in intervals of 2, 4, 5, 10, or 20.

Most Complete Attributes

This widget lists the five most complete attributes in the data profile. Completeness is rounded to the nearest whole percentage and attributes with the same completeness are presented in the order in which they are retrieved from the database. Inherited values are included in the calculation.

Note: This widget has no relation to the Completeness Metric defined in the Profile Configuration and will evaluate all attributes in the data profile.



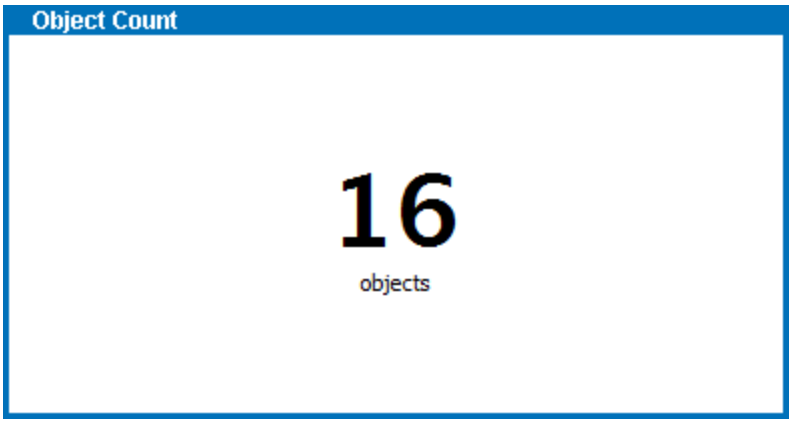
Configuration options include:

- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.

Object Count

This widget displays the number of objects in the data profile.

Note: If any objects being profiled are linked multiple times in the sub-tree, they are also counted multiple times. This only applies to sub-trees under classifications and entities, as products can only link to other products once in a parent / child relationship.



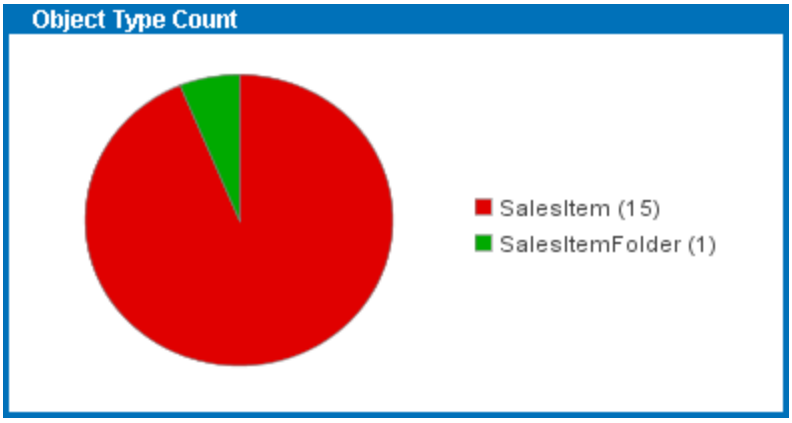
Configuration options include:

- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.

Object Type Count

This widget displays the distribution of object types in the data profile.

Note: Only the five most common types are displayed as separate entries. The remaining types are grouped in "Other object types."

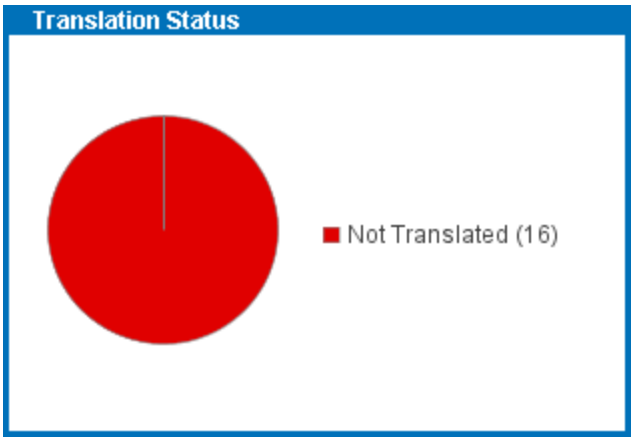


Configuration options include:

- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Chart Type** - Determines whether to display the data as a pie chart or list.

Translation Status

This widget displays the translation status of the objects in the data profile.



Configuration options include:

- **Title** - Changes the name that displays on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Chart Type** - Determines whether to display the data as a pie chart or list.

Note: The KPI Table Widget has been deprecated.

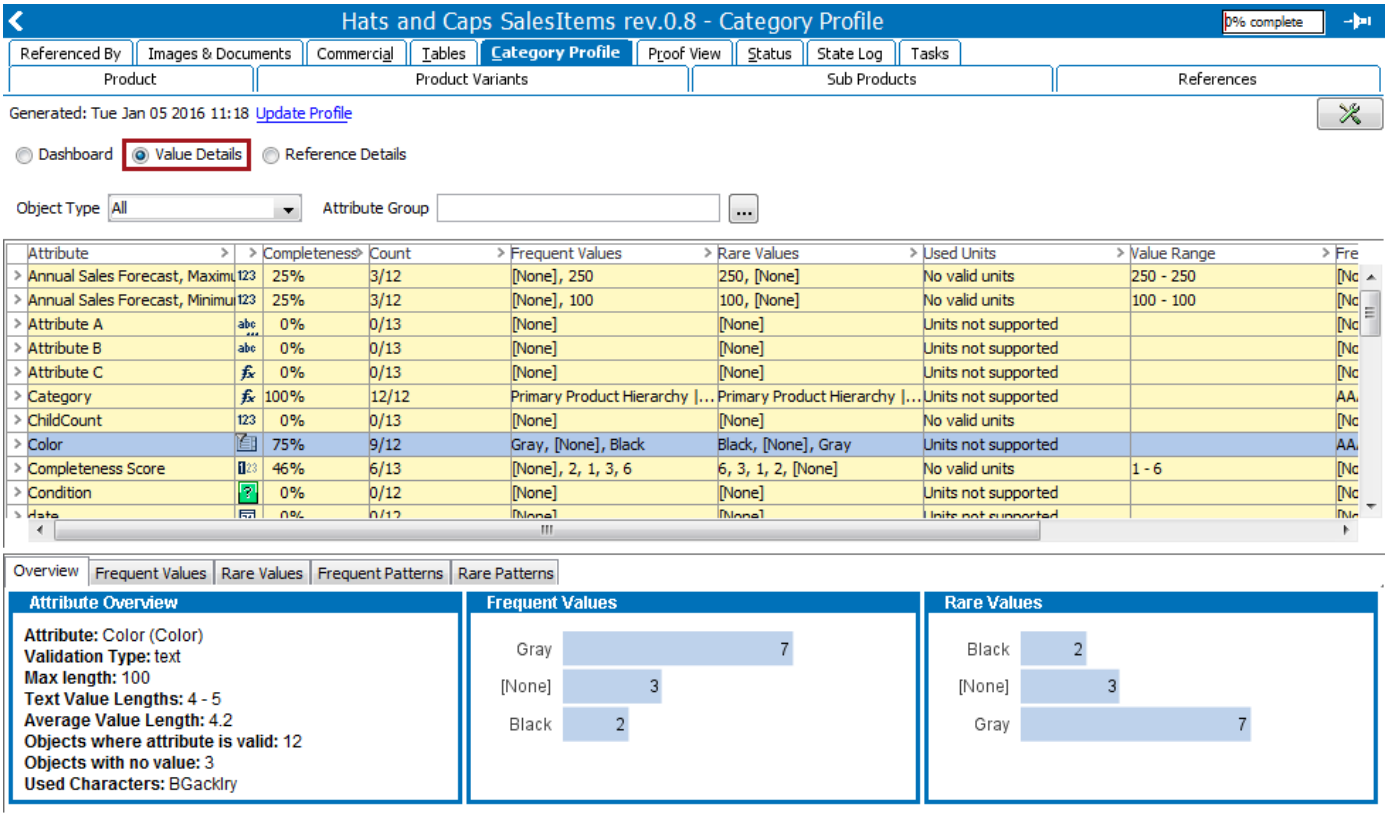
Value Details View

The **Value Details** view provides information about all attributes found within a category / data profile. As described below, this view provides users with multiple ways to display attribute data, including tables, bar graphs, and lists.

Accessing and Navigating Value Details View

1. In the **Tree**, select the relevant object hierarchy or object.
2. On the **Category Profile / Data Profile** tab, click **Generate Profile** or **Update Profile**, and then select **Value Details**.

   The **Value Details** page is divided into two areas, though at first only the upper area displays. The upper area lists all the attributes that are valid for the objects in the data profile. When selecting a row, detailed information about the selected attribute appears in the lower area.



3. From the **Type** list, select a single object type or all object types. The number in parenthesis after the object type name shows how many instances of each object type are in the profile.
4. The user may filter their results so that only those attributes that belong to the specified attribute group will appear on the list. This can be done by entering the name of the specific attribute group into the **Attribute**

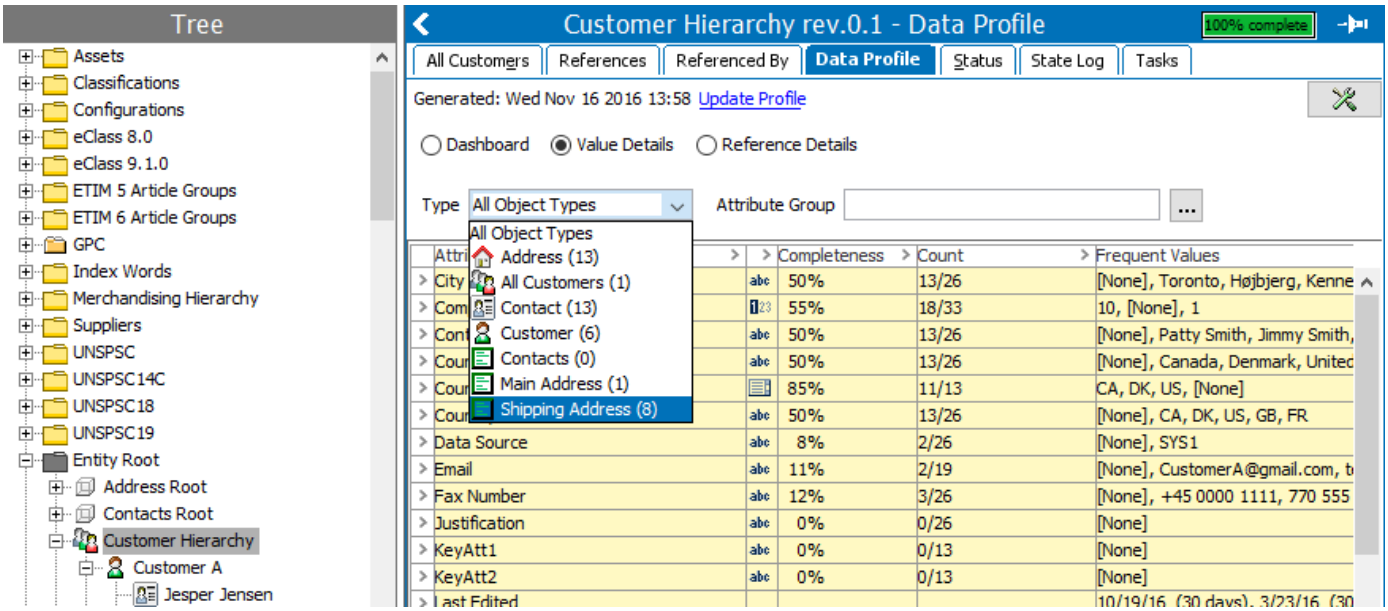
Group field. Alternatively, you may also click the ellipsis button (...) and select the attribute group from the list that appears.

Note: Valid attributes are attributes that are valid on the object types used in the data profile. This includes specification attributes that are linked into the hierarchy in or above the profiled nodes. Specification attributes that are only made valid through the classification hierarchy are not profiled. Attributes and values on references and links, as well as orphan attributes and values, are not profiled.

For information on the tabs in the lower area of this view, see the Value Details Tabs section of the Data Profiling documentation.

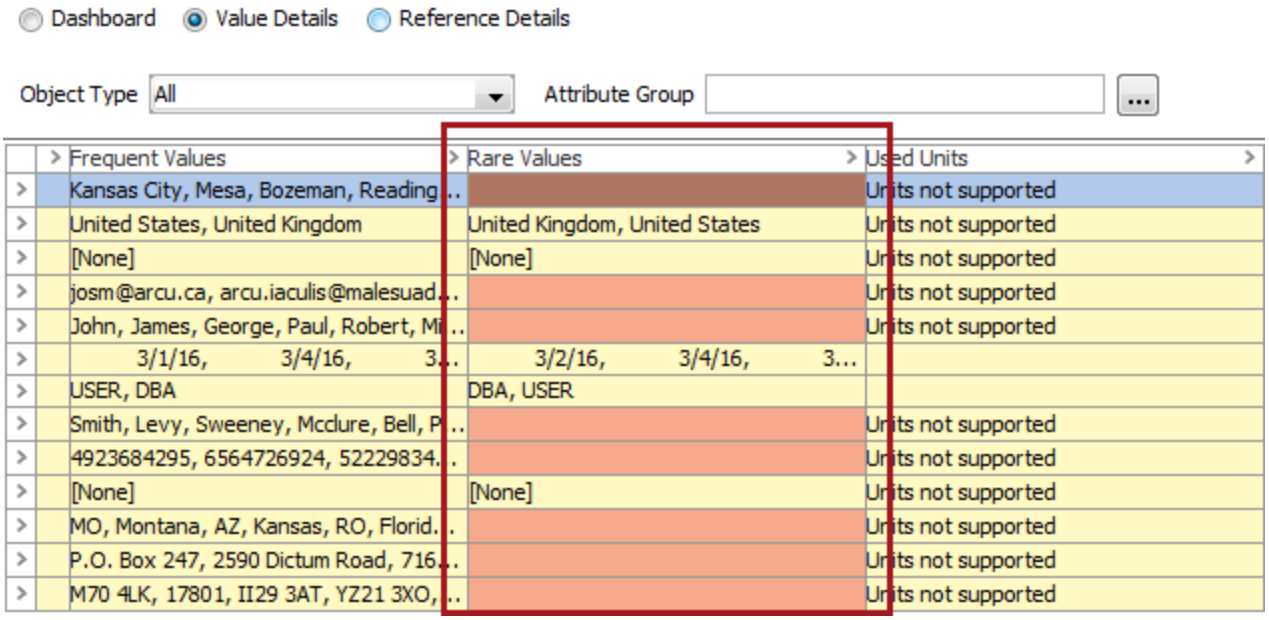
Profiling Entities with Data Containers

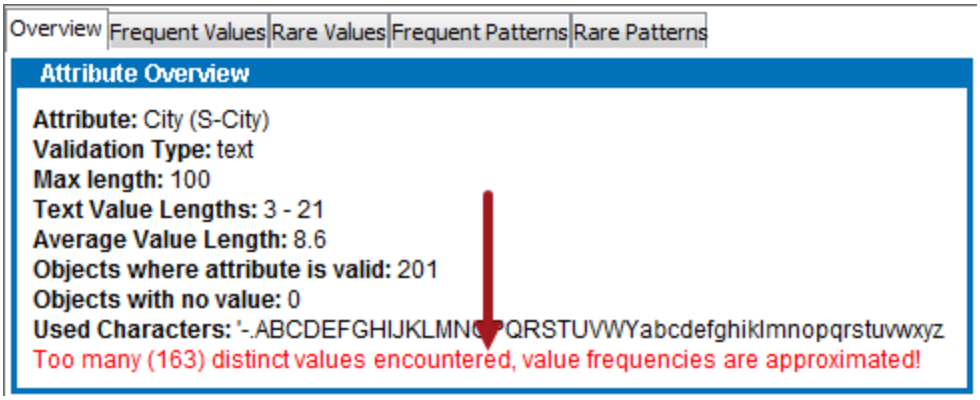
When profiling entities, it is possible to filter the attribute values based on Data Container types.



Maximum Distinct Values

It is possible for an attribute to have too many values to display. In these cases, the applicable rows will turn red and a message will appear in the Attribute Overview section.





By default, the system limits the amount of values the functionality will consider to 100. This limit can be changed by changing these properties via the sharedconfig.properties file:

- DataProfile.MaxDistinctAttributeValuesConsideredDuringProfileGeneration

Additionally, altering this property changes the maximum number of different reference targets that will be considered:

- DataProfile.MaxDistinctTargetsConsideredDuringProfileGeneration

Attribute Table Details

The following list details the columns found in the attribute list table. All columns can be filtered and sorted.

Column	Description
Attribute	Displays the name of the attribute. When clicking the name, the attribute definition in System Setup appears.
Column 2	Shows the Attribute Validation Base Type icon.
Completeness	Shows the degree to which the attribute is populated in the profile rounded to the nearest whole percent. Inherited values are included in the calculation.
Count	Lists the number of objects in the profile where the attribute has a value (including inherited values) and the number of objects where the attribute could possibly have a value.
Frequent Values	Displays a comma-separated list of the most frequently appearing values for the attribute in the profile. The list is sorted in ascending order based on the frequency of that value's appearance. The list includes inherited values. Tags are stripped.
Rare Values	Comma-separated list of the least frequently appearing values for the Attribute in the profile. The list is sorted descending on frequency. Includes inherited values. Tags are stripped.
Used Units	Displays which units are used for the attribute. Contains the following information: • List of units used in values for attributes in the profile. • "Units not supported" - This text is displayed for attributes of a validation base type that does not support units. • "No valid units" - This text is displayed for attributes of a validation base type that supports units but where no units are configured. • "No units used" - This text is displayed for attributes of a validation base type that supports units but where no attribute values in the profile have a unit.
Value Range	Shows the value range for attributes with the validation base type "Integer" or "Number." If units that share the same base validation are used, the value ranges are shown with Units. For example: 123 cm - 2 m."
Frequent Patterns	Displays the most frequent patterns of the values of a given attribute in addition to the number values that match the pattern. More information can be found in the **Value Details Tabs** section of the **Data Profiling** documentation.
Rare Patterns	Displays the most uncommon patterns of the values of a given attribute in addition to the number values that match the pattern. More information can be found in the **Value Details Tabs** section of the **Data Profiling** documentation.

Value Details Tabs

The lower area of Value Details is divided into a number of tabs that contain detailed attribute information.

- Overview Tab
- Frequent Values Tab
- Rare Values Tab
- Frequent Patterns Tab
- Rare Patterns Tab

Actions

With the exception of the Overview tab, all tabs contain a table where three different actions are possible. From left to right, the icons represent the following actions: Bulk Update, Search, and Save as Collection  .

- **Bulk Update** When the Bulk Update icon is clicked, a search is run within the profile for objects that have the selected attribute value. The result is used as input data set for the Bulk Update wizard.

  The Bulk Update wizard opens on the **Operations** step with the attribute selected in the **Set Value** operation.

  This functionality can be used to fix data errors. Since the input data set is the result of a search, you will not accidentally update the wrong data due to running the operation from an outdated profile.

  If viewing value details for a specific object type, the search is filtered so that only objects of the selected type are used as input data set for the Bulk Update operation. Likewise, if you have selected "Only show values entered as local values", the search only includes local values.

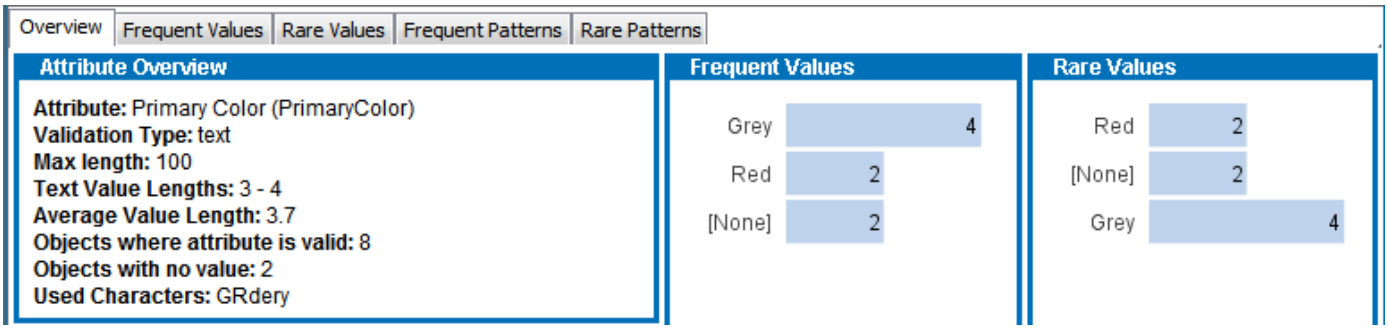
  For more information about the Bulk Update wizard, see the **Creating a Bulk Update** section of the **Bulk Updates** documentation.

  Important: If "Only show values entered as local values" is not checked before you run the bulk update, you run the risk of creating local instances of data from before the operation was inherited through the hierarchy.

- **Search** - When the Search icon is clicked, a search is run within the profile for objects that have the selected attribute value. When the search is complete, the result is displayed on the **Search** navigation tab. Searches are filtered on the object type selection in the profile and inherited values are included or excluded depending on whether you select "Only show values entered as local values."
- **Save as Collection** When the Save as Collection icon is clicked, the Save as Collection in Background Process dialog is displayed. Select a location and enter a Collection Name and a Collection ID, and then click **OK**. A search is run within the profile for objects that have the selected attribute value, and the result is saved in the collection.

Overview Tab

The **Overview** tab contains a number of widgets that each present different information about the selected attribute. Users cannot customize this view.



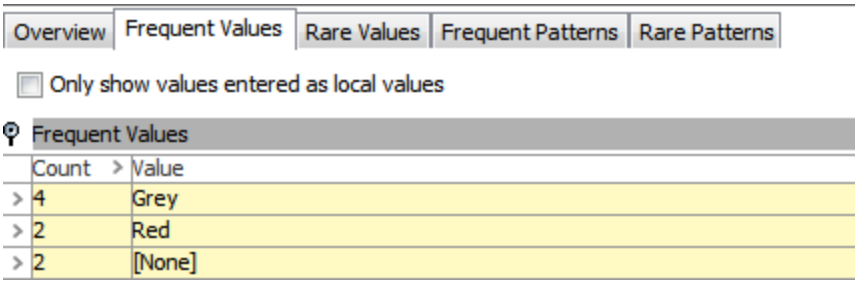
Widgets on the Overview tab.

Widget	Description
Attribute Overview	Displays information retrieved from the attribute definition in System Setup and information generated in the profiling. You can, for example, see which characters are used, the value length span, and the numerical value span of number and integer attributes.
Frequent Values	Bar chart that displays the ten most frequently appearing values or words. Includes inherited values. Tags are stripped. Values with an equal count are presented in the order they are retrieved from the database.
Rare Values	Bar chart that displays the ten least frequently appearing values or words. Includes inherited values. Tags are stripped. Values with an equal count are presented in the order they are retrieved from the database.
Used Units	This widget is only displayed for attributes of validation base types that support units. The widget can contain the following content: • Pie chart of units used in values for attributes in the profile. • "No valid units" - This text is displayed for attributes of a validation base type that supports units but where no units are configured. • "No units used" - This text is displayed for attributes of a validation base type that supports units but where no attribute values in the profile have a unit. Inherited values are included in the pie chart calculation.

Frequent Values Tab

On this tab, the 100 most common values in the profile of the selected attribute are listed. Tags are resolved.

The default view includes inherited values, however, if you check "Only show values entered as local values", inherited values are excluded. Values inherited from other contexts are still displayed. Values with an equal count are presented in the order they are retrieved from the database.



Rare Values Tab

On this tab, the 100 least common values in the profile of the selected attribute are listed. Tags are resolved.

The default view includes inherited values, however, if you check "Only show values entered as local values", inherited values are excluded. Values inherited from other contexts are still displayed. Values with an equal count are presented in the order they are retrieved from the database.

Used Unites Tab

This tab lists all units used in the values of the selected attribute, such as length in centimeters (cm), weight in pounds (lbs), price in dollars ($), etc. This tab only appears for those attributes with values containing units.

Frequent Patterns Tab

This tab lists the most frequent patterns of the values of a given attribute and the number values that match the pattern.

A value pattern is a way of describing the structure of a value in terms of characters, numbers, and symbols. When the patterns of the values in a set are compared, certain types of errors can be detected by finding values whose patterns are different from the rest. For example, phone numbers with missing digits or numbers that are mixed with letters, and so on.

A pattern for a given value is created by replacing any letter with A, any number with 9, and any unrecognized character with X. Other symbols are left as they are.

The following is an example of a value pattern:

Value	Pattern
123-ABB	999-AAA
234.23	999.99
mail@mail.com	AAAA@AAAA.AA

The pattern [None] represents an empty value.

Rare Patterns Tab

This tab resembles the Frequent Patterns but lists the least frequently used patterns rather than the most frequent. This information is useful when searching for faulty values.

Special Attribute and Value Handling

Full text indexed attributes and multi-valued attributes are handled specially.

- **Full text indexed attributes**

For attributes configured as full text indexable, the most frequent or rare words are listed with the exception of words on Oracle's English Default Stoplist. However, if attribute values have units, the attribute is not profiled as full text indexable.

- **Multivalued attributes**

For multivalued attributes, the entries in each value are split out. In the following example, only three products in the profile have a value for the multivalued attribute color:

- Product A: Red, Green, Blue
- Product B: Yellow, Red
- Product C: Blue, Red

The order of the frequent values information is:

Red, Blue, Green, Yellow

- **Values with inline references**

If a value consists of an inline reference only, and the reference does not resolve to any value, the value appears as an empty string. In overview charts, however, the text "EMPTY" is displayed.

- **Empty values and references**

Empty attribute values and empty reference values are handled as ordinary values. Empty values show up in detail views and graphs as [None].

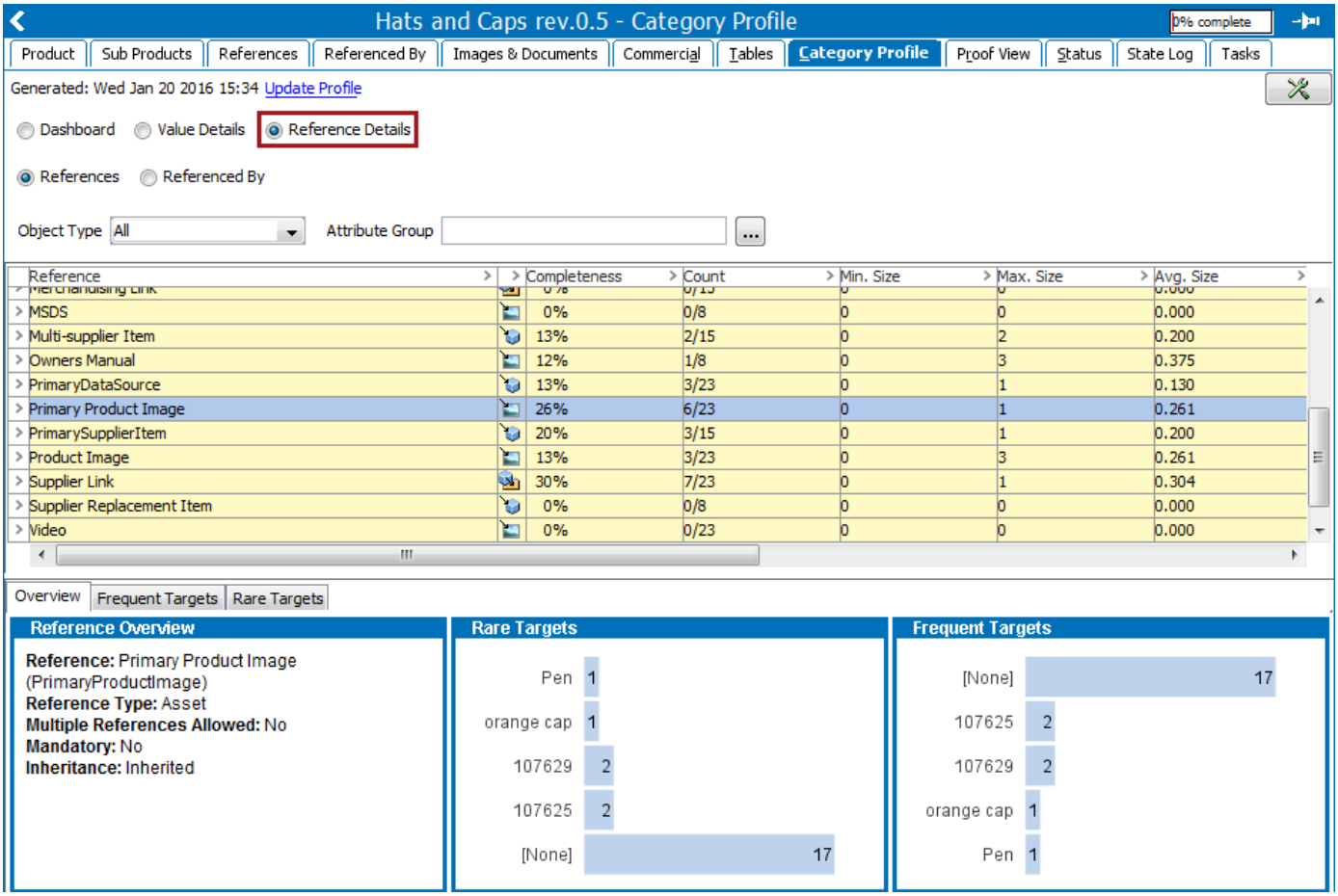
Reference Details View

The **Reference Details** view provides information about all references, image and document references, and product to classification links for which objects within the selected profile are valid sources.

Accessing and Navigating Reference Details View

1. In the **Tree**, select the relevant hierarchy or object.
2. On the **Category / Data Profile** tab, click **Generate Profile** or **Update Profile**, and then select **Reference Details**.

   The **Reference Details** page is divided into two areas, though at first only the upper area displays. The upper area lists all the references and links that are valid for objects in the profile. When you select a row, detailed information about the selected reference or link will then display in the lower area.



3. From the **Object Type** list, select a single object type or all object types. The number in parenthesis after the object type name shows how many instances of that object type are in the category.

4. To view references available to the objects, select **Referenced By**.
5. The user may filter their results so that only those attributes that belong to the specified attribute group will appear on the list. This can be done by entering the name of the specific attribute group into the **Attribute Group** field. Alternatively, you may also click the ellipsis button (...) and select the attribute group from the list that appears.

For information on the tabs in the lower area of this view, see the Reference Details Tabs section of the Data Profiling documentation.

Reference Table Details

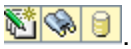
The following list details the columns found in the reference list table. All columns can be filtered and sorted.

Column	Description
Reference	Displays the reference and link type titles. When clicking the name, the reference or link type definition in System Setup appears.
Column 2	Shows the Attribute Validation Base Type icon.
Completeness	Shows the degree to which the reference or link type is populated in the profile rounded to the nearest whole percent. Inherited references and links are included in the calculation. If one reference or link of the specified type exists on an object, it is counted as populated regardless of whether or not the type is configured to allow multiple references or links.
Count	Displays two numbers as a ratio. The first is the number of objects in the profile where the reference or link type is populated (including inherited references and links), and the second is the number of objects in the profile that are valid sources for the reference or link type.
Min. Size	Minimum number of references of this type on an object in the profiled data set.
Max. Size	Maximum number of references of this type on an object in the profiled data set.
Avg. Size	Average number of references of this type on objects in profiled data set.
Frequent Targets	Displays a comma-separated list of the most frequently referenced or linked targets for the reference or link type in the profile. The list is sorted in ascending order by frequency and includes inherited references and links.
Rare Targets	Displays a comma-separated list of the least frequently referenced or linked targets for the reference or link type in the profile. The list is sorted in ascending order by frequency and includes inherited references and links.

Reference Details Tabs

The lower area of Reference Details is divided into a number of tabs that contain detailed attribute information.

- Overview Tab
- Frequent Targets Tab
- Rare Targets Tab

The Frequent Targets and the Rare Targets tabs contain a table where three different actions are possible. From left to right, the icons represent the following actions: Bulk Update, Search, and Save as Collection  .

- **Bulk Update** When the Bulk Update icon is clicked, a search is run within the profile for the selected reference or link type. The result is used as input data set for the Bulk Update wizard.

  The Bulk Update wizard opens on the **Operations** step with the reference or link type selected in the **Add Reference** operation. Choose a target that is valid for the selected type.

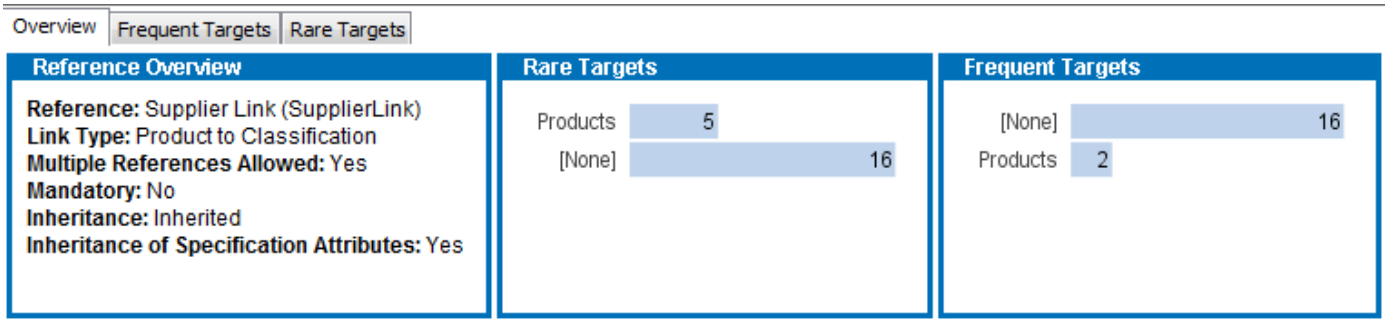
  This functionality can be used to fix data errors. Since the input data set is the result of a search, you will not accidentally update the wrong data due to running the operation from an outdated profile.

  For more information about the Bulk Update wizard, see the **Creating a Bulk Update** section of the **Getting Started / User Guide** documentation.

- **Search** - When the Search icon is clicked, a search is run within the profile for objects that have the selected reference or link type. When the search is complete, the result is displayed on the **Search** navigation tab. Searches are filtered on the object type selection in the profile.
- **Save as Collection** When the Save as Collection icon is clicked, the Save as Collection in Background Process dialog is displayed. Select a location and enter a Collection Name and a Collection ID, and then click **OK**. A search is run within the profile for objects that have the selected reference or link type, and the result is saved in the collection.

Overview Tab

The **Overview** tab contains a number of widgets that each present different information about the selected reference or link type. You cannot customize this view.



Widgets on the Overview tab.

Widget	Description
Overview	Displays information retrieved from the reference or link type definition.
Frequent Targets	Bar chart that displays the ten most frequently referenced or linked targets for the reference or link type in the profile. Includes inherited references and links. Targets with an equal count are presented in the order they are retrieved from the database.
Rare Targets	Bar chart that displays the ten least frequently referenced or linked targets for the reference or link type in the profile. Includes inherited references and links. Targets with an equal count are presented in the order they are retrieved from the database.

The Frequent Targets Tab

On this tab, the 100 most common referenced / linked targets for the Reference / Link Type in the Category are listed. Inherited References/Links are included in the count. Note that targets with an equal count are presented in the order they are retrieved from the database.

The Rare Targets Tab

On this tab, the 100 least common referenced / linked targets for the Reference / Link Type in the Category are listed. Functionality-wise, the tab works exactly like the detail **Frequent Targets** tab described above.

Profile Configuration

Category Profiles can be configured with strict profiling parameters via the **Profile Configuration** screen. From this view, users can select a preexisting (saved) profile configuration to use or create one specifically for the selected object / node. If you choose to create a local configuration for the selected object / node, additional configuration options are made available.

Note: If using a saved profile configuration, all relevant parameters will already be defined.

These configuration options allow users to:

- Use a business condition to specify which objects to profile (allows you to exclude discontinued products, for example)
- Include or exclude data from specified attribute groups when profiling the data
- Designate business conditions to test when a profile is run
- Split up the value of a specified attribute so that the individual words in the attribute's value are profiled as separate entries
- Define which Metrics to use when profiling the data

Note: The relevant profile(s) must be updated manually every time changes are made to any of the parameters.

Using Saved Profile Configurations

Some setup is required to make profile configurations available for use in multiple objects / nodes.

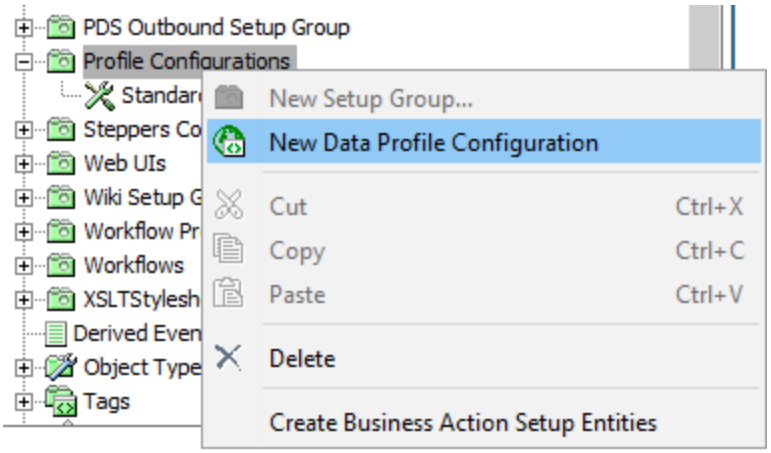
Note: It is possible to skip this setup if you intend to use a local profile configuration for each individual data profile.

Prerequisites

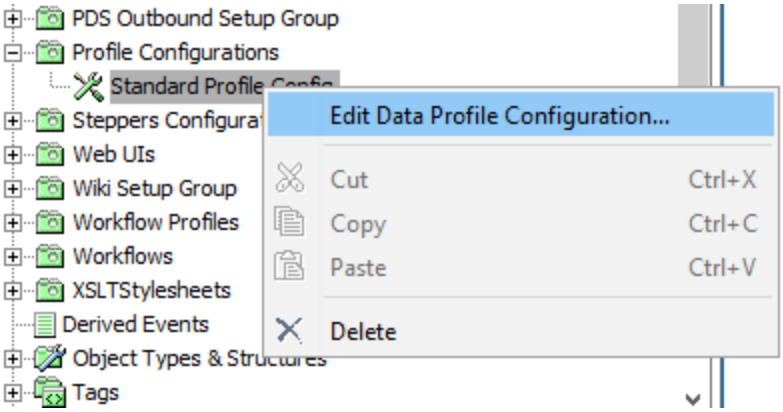
Before you can create and use **Profile Configuration** setup entities, a setup group must first be created to house the configurations. For more information, see the **Creating Setup Groups in System Setup** section of the **System Setup / Super User Guide** documentation.

Creating a Profile Configuration Setup Entity

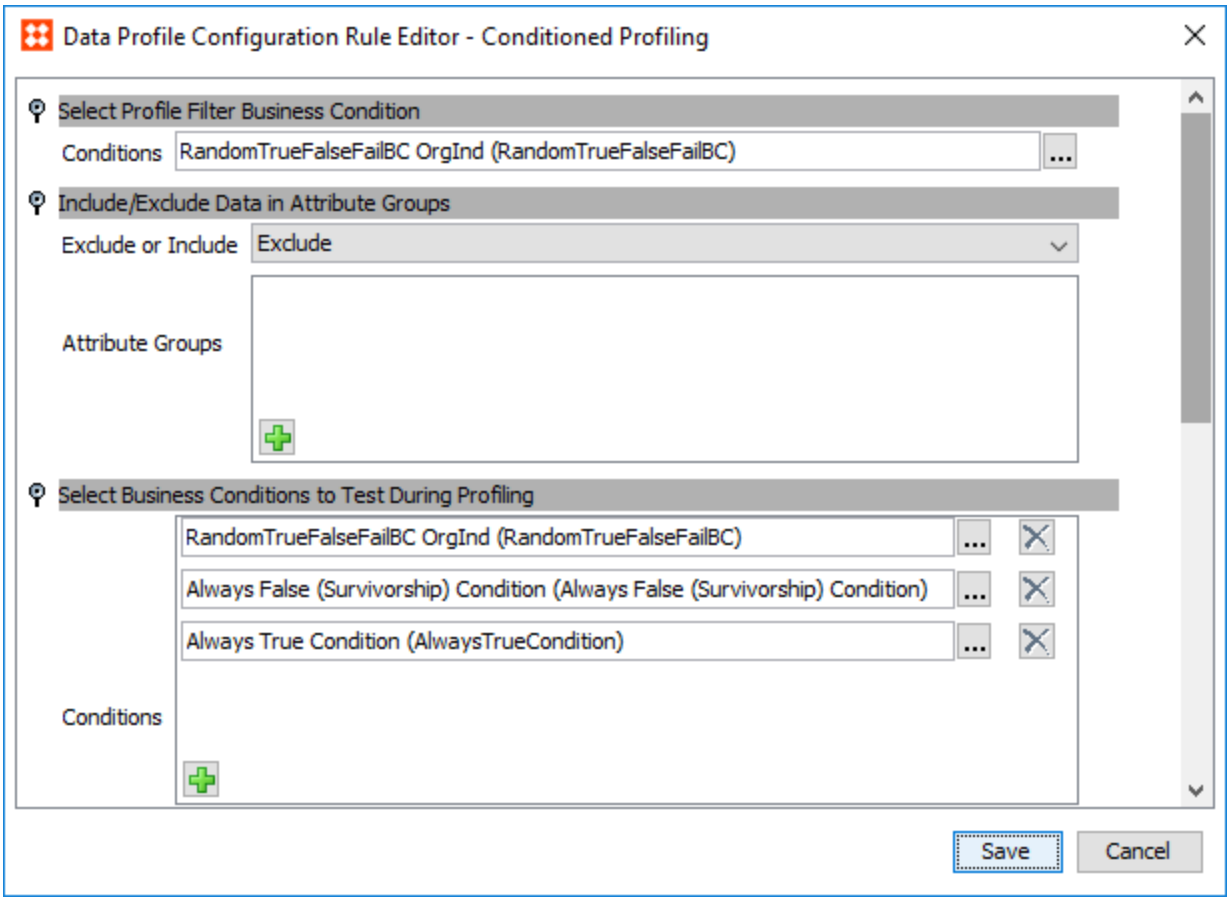
In order to create a new profile configuration for use in multiple different profiles, navigate to **System Setup** in workbench, right-click the **Profile Configurations** root node, and select **New Data Profile Configuration**.



To edit the newly created profile configuration, right-click the node and select **Edit Data Profile Configuration...**.



In the window that appears, make any desired changes to the configuration and then click **Save**.

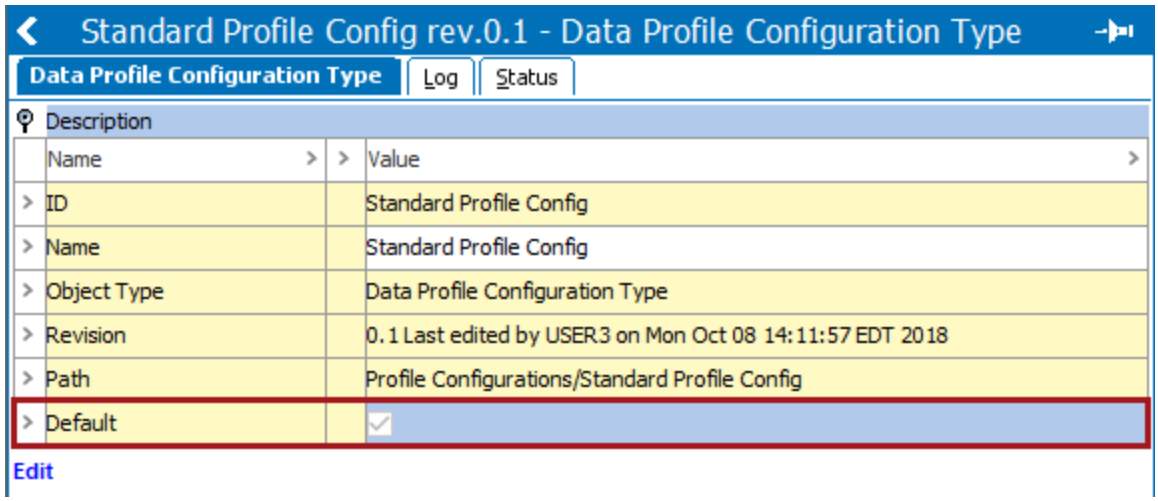


For more information on the available configuration options, see the **Editing a Profile Configuration** section below.

Setting a Default Configuration

If you have several different profile configurations in your system it may be necessary to set one as the default. This is useful whenever a configuration is *not* specified on a node / object, as this default configuration will automatically be chosen when a profile is run.

To set a configuration as the default, navigate the configuration node in **System Setup**, and on the **Data Profile Configuration Type** tab, check the box labeled **Default**.



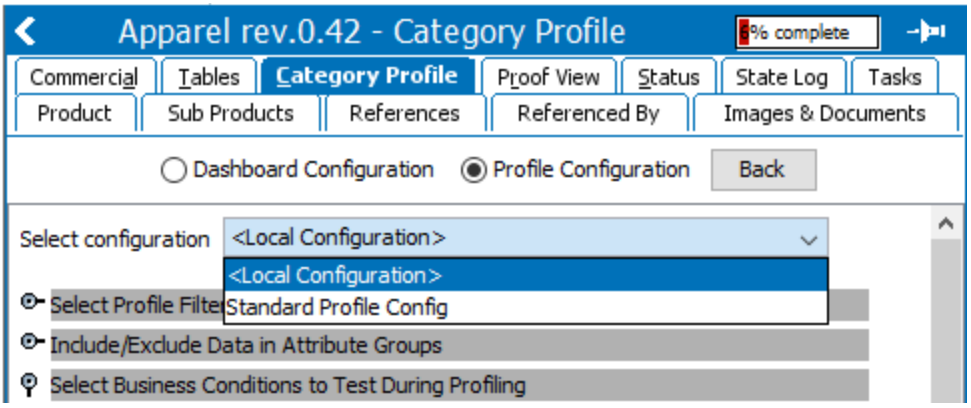
Accessing the Profile Configuration on an Object / Node

1. In the **Tree**, select the relevant hierarchy or object.
2. On the **Category Profile** tab, click the **Tools** icon in the upper right corner, and then select **Profile Configuration**.

Editing a Profile Configuration

Whether editing a saved profile configuration or editing a local profile configuration on an individual object / node, the configuration options detailed in this section apply to both.

When editing a local configuration on an object / node, however, one additional parameter called **Select Configuration** is available. This parameter allows you to select between a local configuration and any saved profile configurations from a dropdown menu.



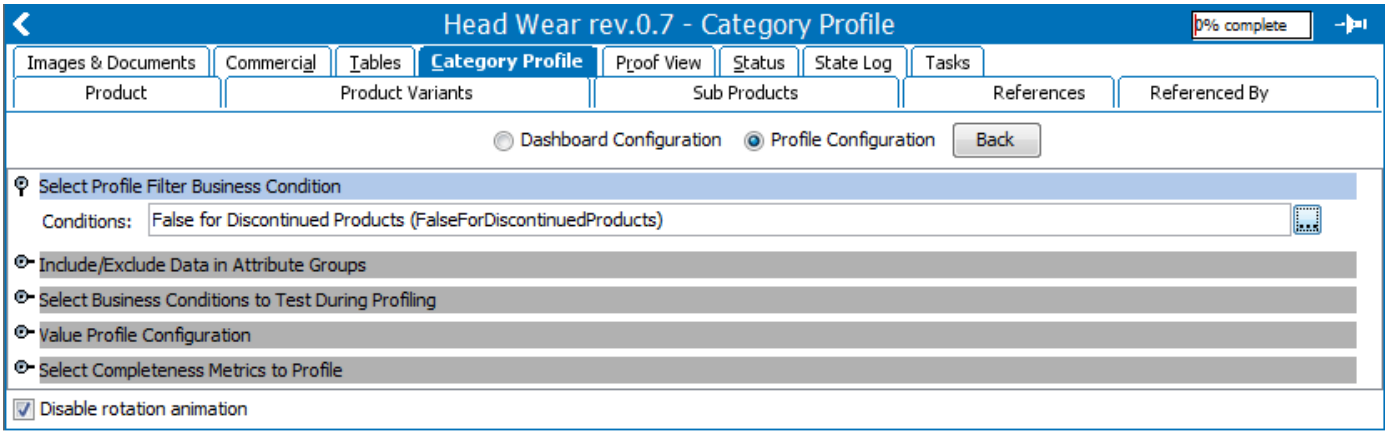
Selecting **<Local Configuration>** opens the rest of the **Profile Configuration** parameters for editing, allowing you to create a unique profile configuration for the currently selected object / node.

If you instead wish to apply a saved profile configuration to the selected object / node, select the desired configuration from the dropdown.

Note: Any examples detailed below assume you are editing a local configuration and have relevant dashboard widgets configured. The actual configuration steps of each parameter also apply to editing profile configuration setup entities (saved profile configurations).

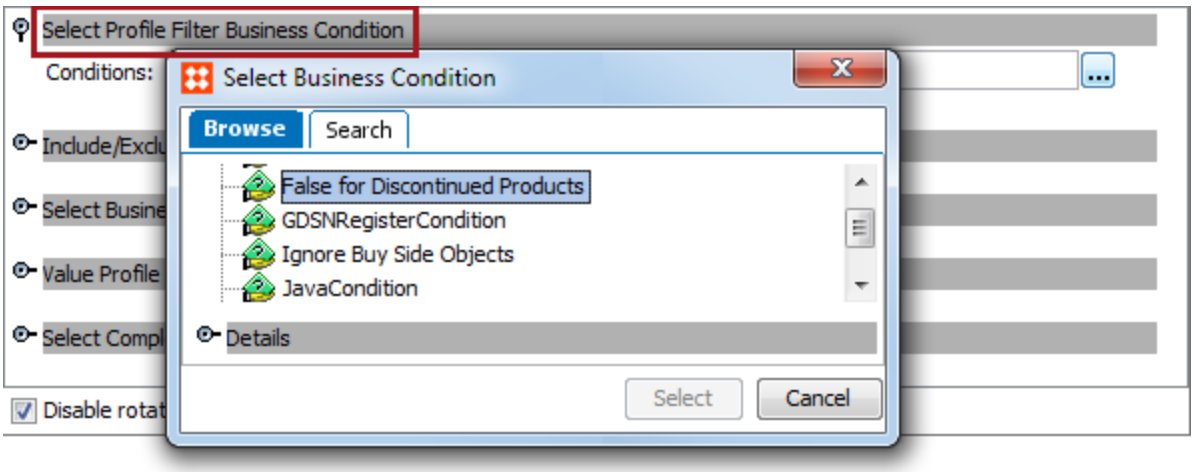
Select Profile Filter Business Condition

The **Select Profile Filter Business Condition** parameter allows users to filter the data being profiled with the specified business rule. This can be particularly useful if only objects of a specific type should be profiled or objects with specific attribute values should be left out.

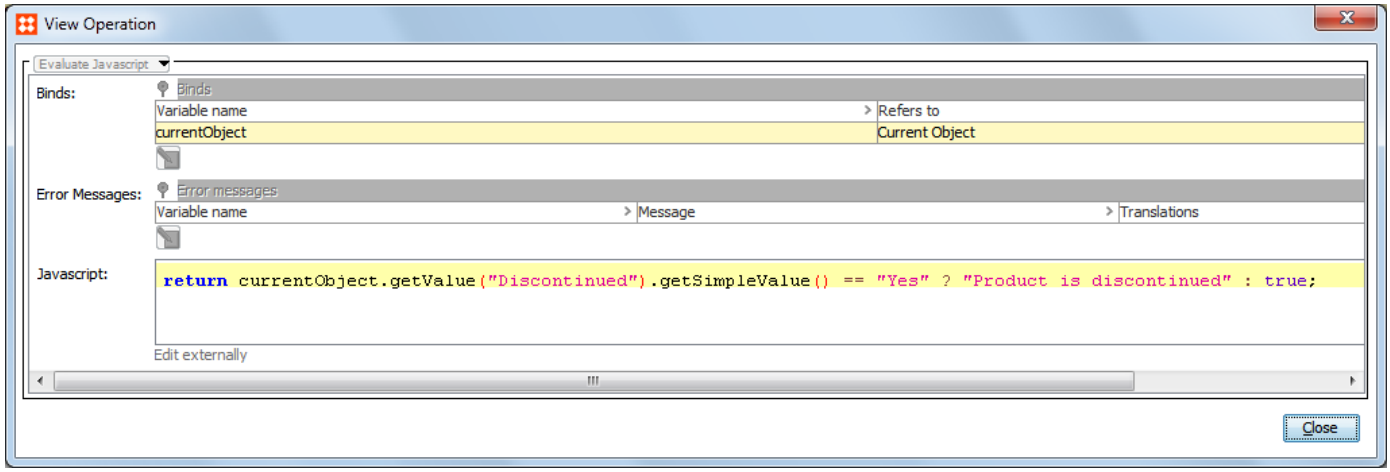


1. In the **Select Profile Filter Business Condition** section, click the ellipsis button (...), then browse or search for the relevant business condition.

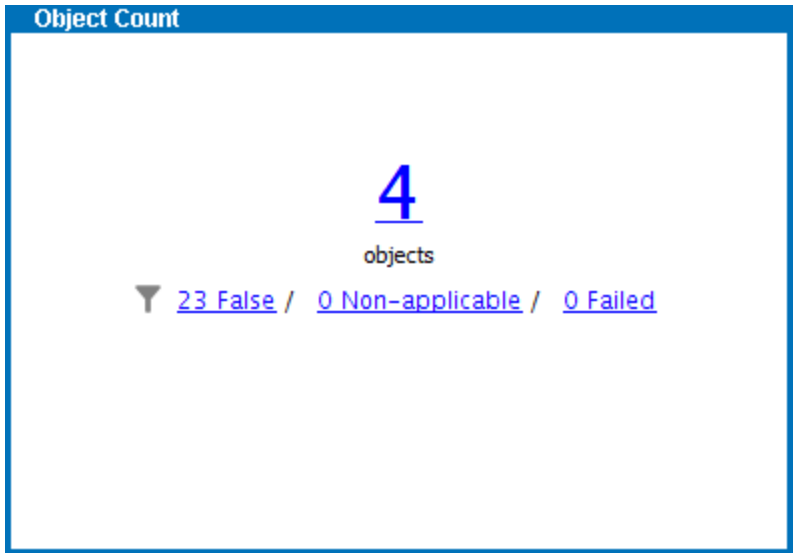
   As an example, in the following screenshot a business condition that excludes discontinued products is selected:



The selected condition specifies that if the attribute **Discontinued** is set to **Yes**, the object is excluded from the profiling. The selected business condition is pictured below:



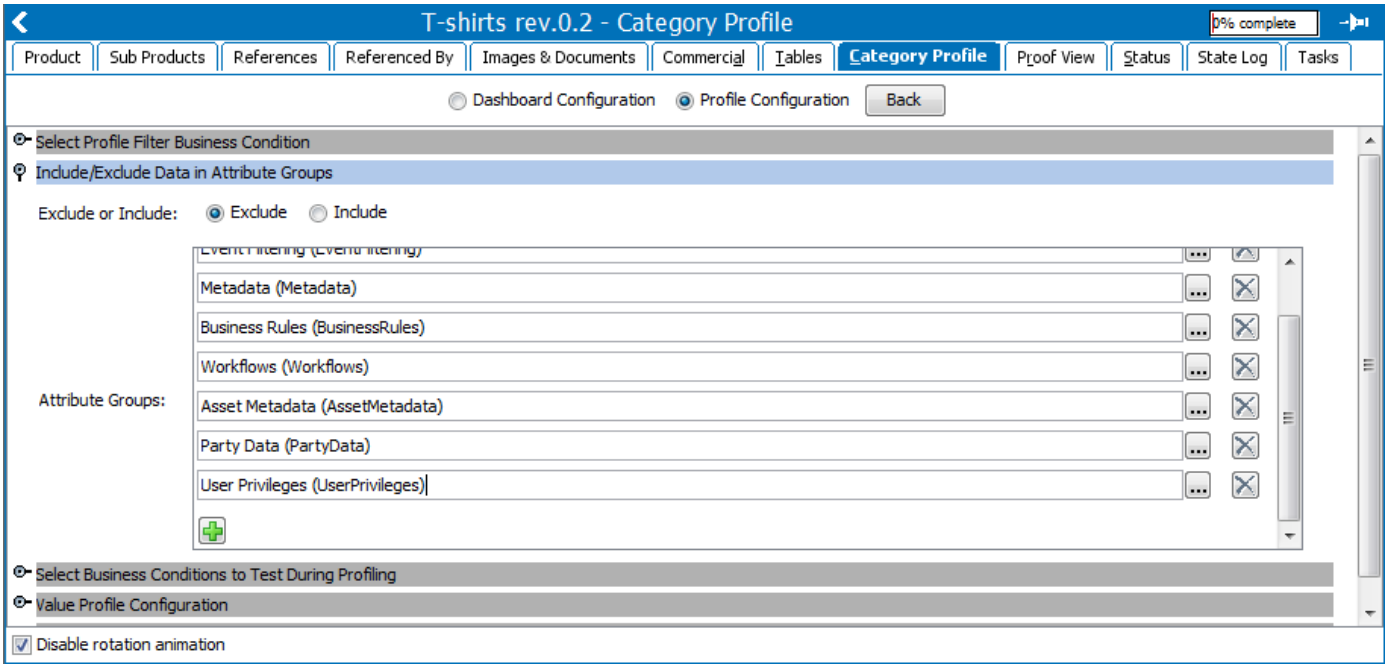
2. Click **Select** to apply the business condition to the profile.
3. If editing a profile configuration setup entity, click **Save**. If editing a local profile, click the **Back** button to return to the **Category Profile Dashboard**, then click the **Update / Generate profile** link. If an **Object Count** widget is configured on the dashboard it will display the number of objects that were filtered out by the condition. In the example below, twenty-three discontinued objects were excluded as these were deemed False during profiling:



For more information on how to configure the Object Count widget, see the **Category Profile Dashboard** section of the **Data Profiling** documentation.

Include / Exclude Data in Attribute Groups

The **Include / Exclude Data in Attribute Groups** parameter allows users to filter the data being profiled via attribute group.



1. Navigate to the **Include / Exclude Data in Attribute Groups** section and click either the **Exclude** or **Include** radio button. If **Include** is selected, the data profile will only pull data from the attribute groups specified by the user in the Attribute Groups field. If **Exclude** is selected, the data profile will pull data from all attribute groups except those specified in the Attribute Groups field. Typically, the choice between **Include** and **Exclude** will largely depend on which requires the shorter list.

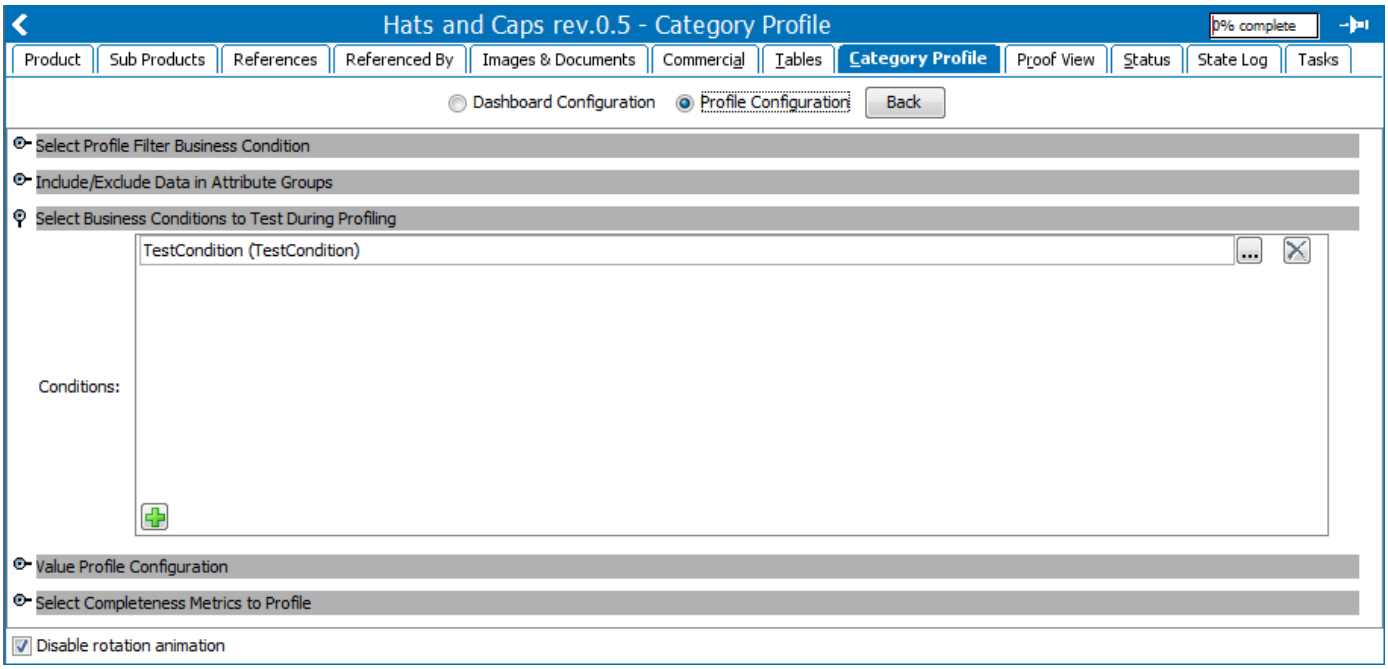
Note: To include all attribute groups in the system, the user can select **Exclude** and leave the Attribute Groups field blank.

---

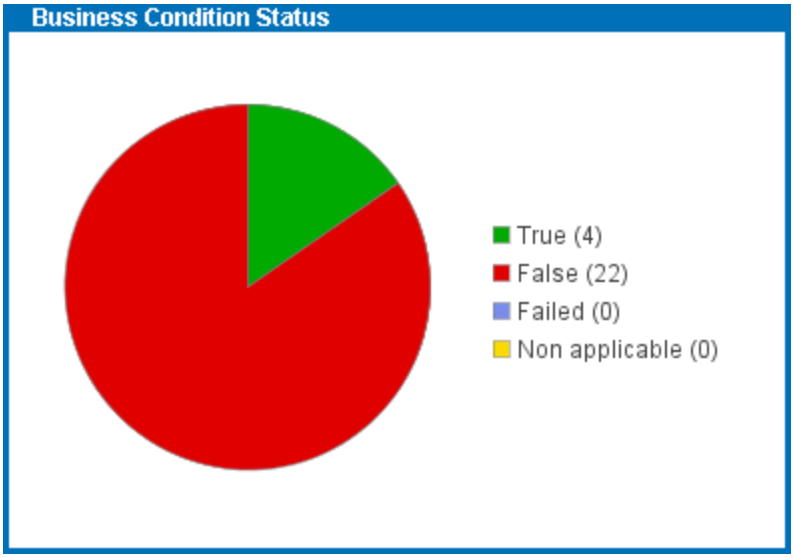
2. Click the green plus sign to add a new attribute group selector field.
3. Click the ellipsis button (...), then browse or search for the relevant attribute group.
4. Click **Select**. If no other attribute groups need to be specified, click the **Back** button when editing a local configuration or **Save** if editing a profile configurations setup entity.

Select Business Conditions to Test During Profiling

The **Select Business Conditions to Test During Profiling** parameter allows users to test business conditions for all products, entities, classifications, and assets when profiling a selected hierarchy. The result of the tests can be displayed in Business Condition Status widgets on the dashboard.



1. Navigate to the **Select Business Conditions to Test During Profiling** section and click the green plus sign to add a new business condition to the selector field.
2. Click the ellipsis button (...), then browse or search for the relevant business condition.
3. Click **Select**. If no other business conditions need to be specified, click the **Back** button when editing a local configuration or **Save** when editing a profile configurations setup entity. If the Business Condition Status widget is configured on the dashboard, it will display the results of the test.

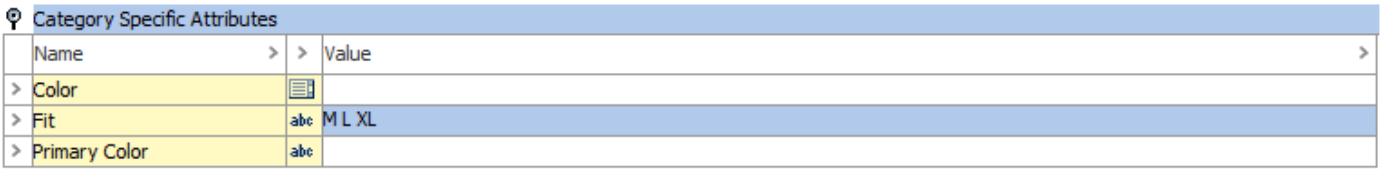


For more information on how to configure the Business Condition Status widget, see the **Category Profile Dashboard** section of the **Data Profiling** documentation.

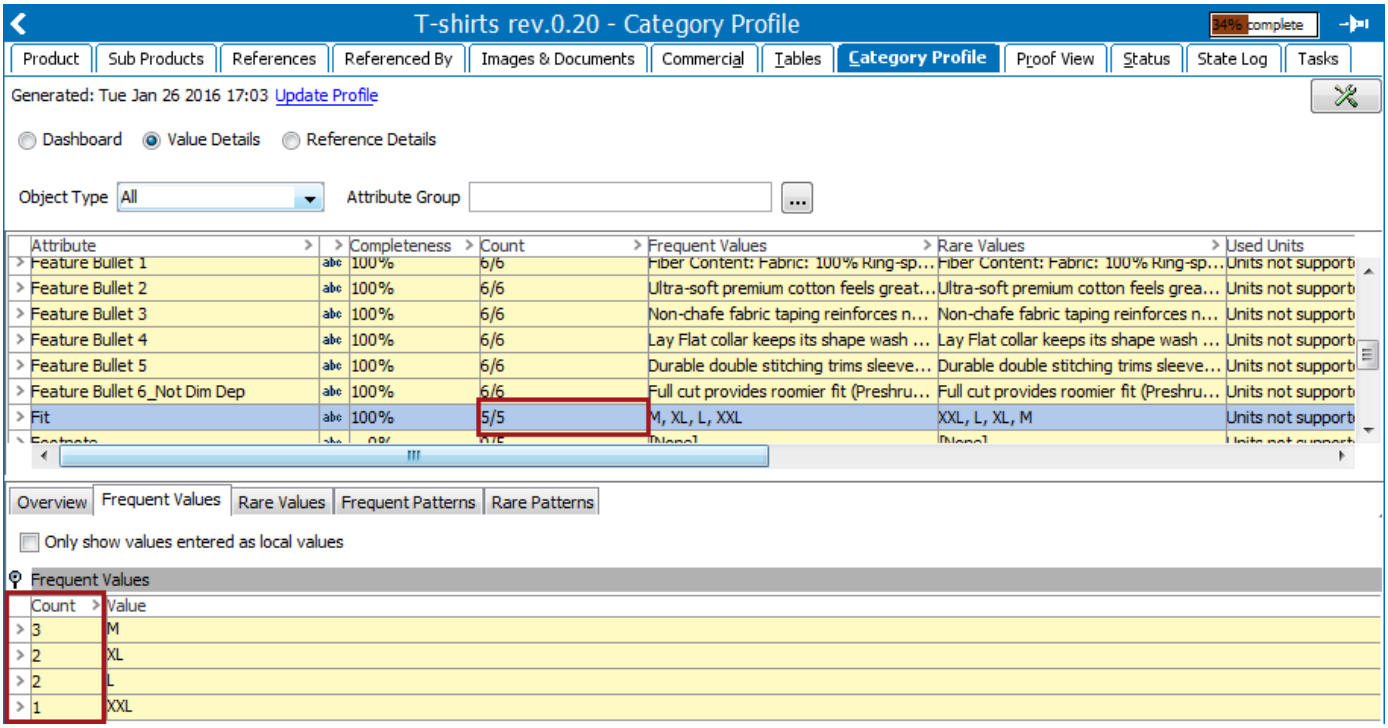
Value Profile Configuration

The **Value Profile Configuration** parameter allows users to profile the individual words of an attribute's value. As an example, an attribute 'Fit' with validation base type 'Text' lists all applicable shirt sizes for a product. These individual sizes can be profiled separately so that the attribute value 'M L XL' is profiled as three different values, 'M', 'L', and 'XL.'

Attribute value as seen on the product node:



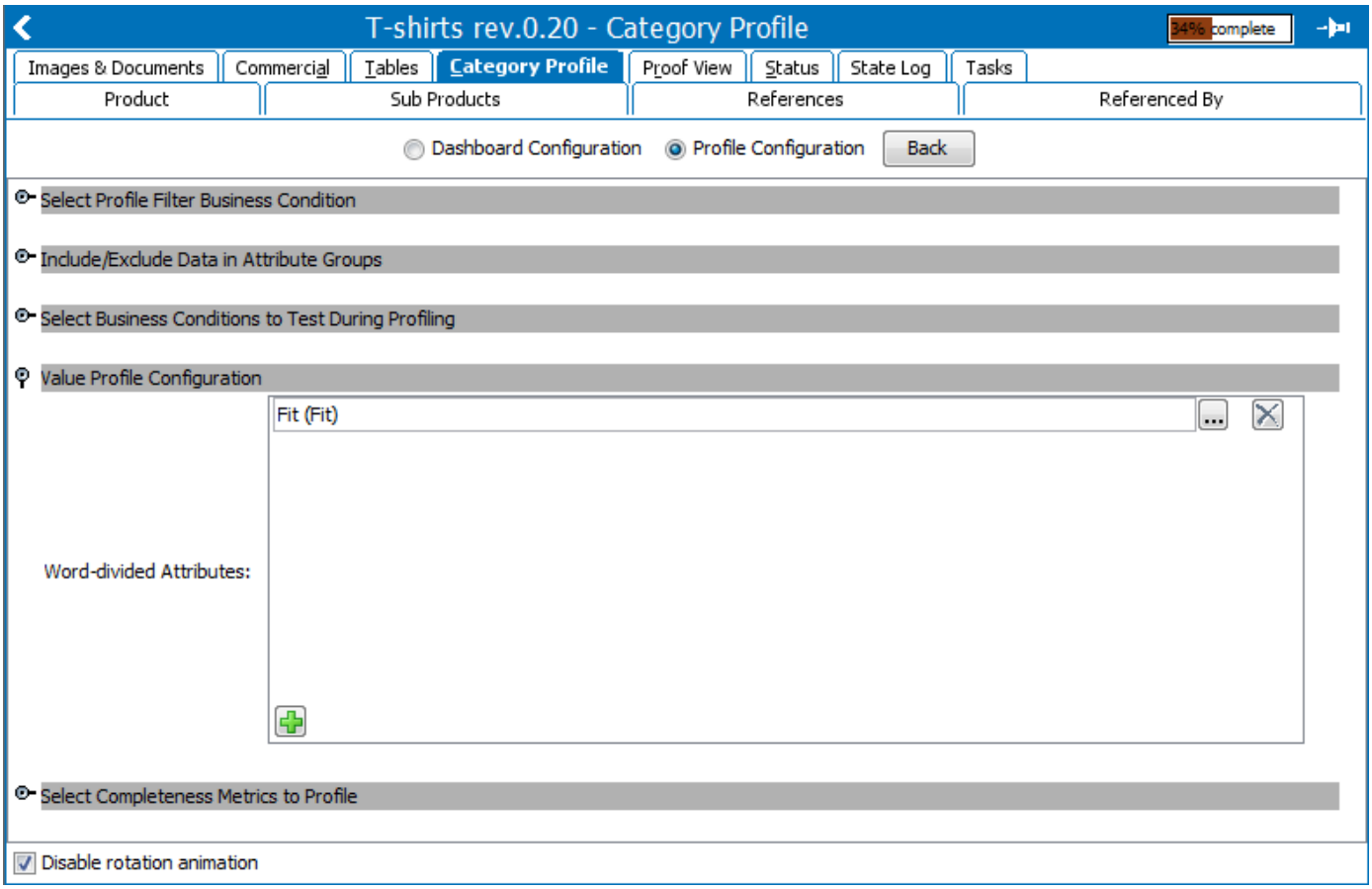
Profiled values as seen in the Value Detail view:



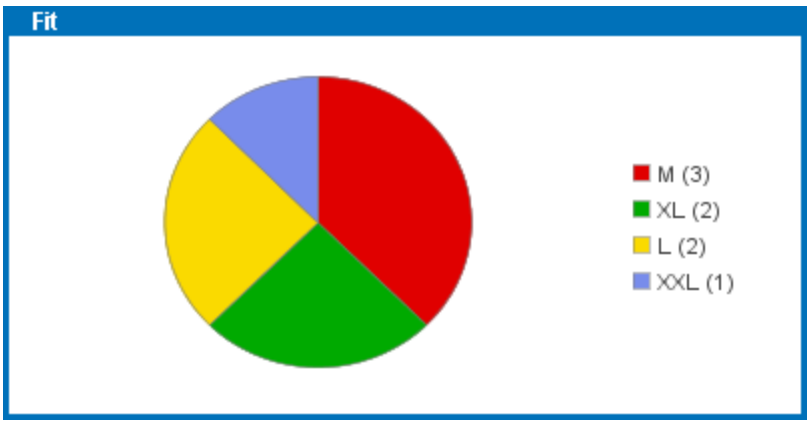
In addition to the **Value Details** view, this separation of values is also reflected on all applicable dashboard widgets. For more information on the Value Details view, see the **Value Details** section of the **Category Profile** documentation.

Parameter Configuration Steps

Select the attributes whose values should be profiled separately.

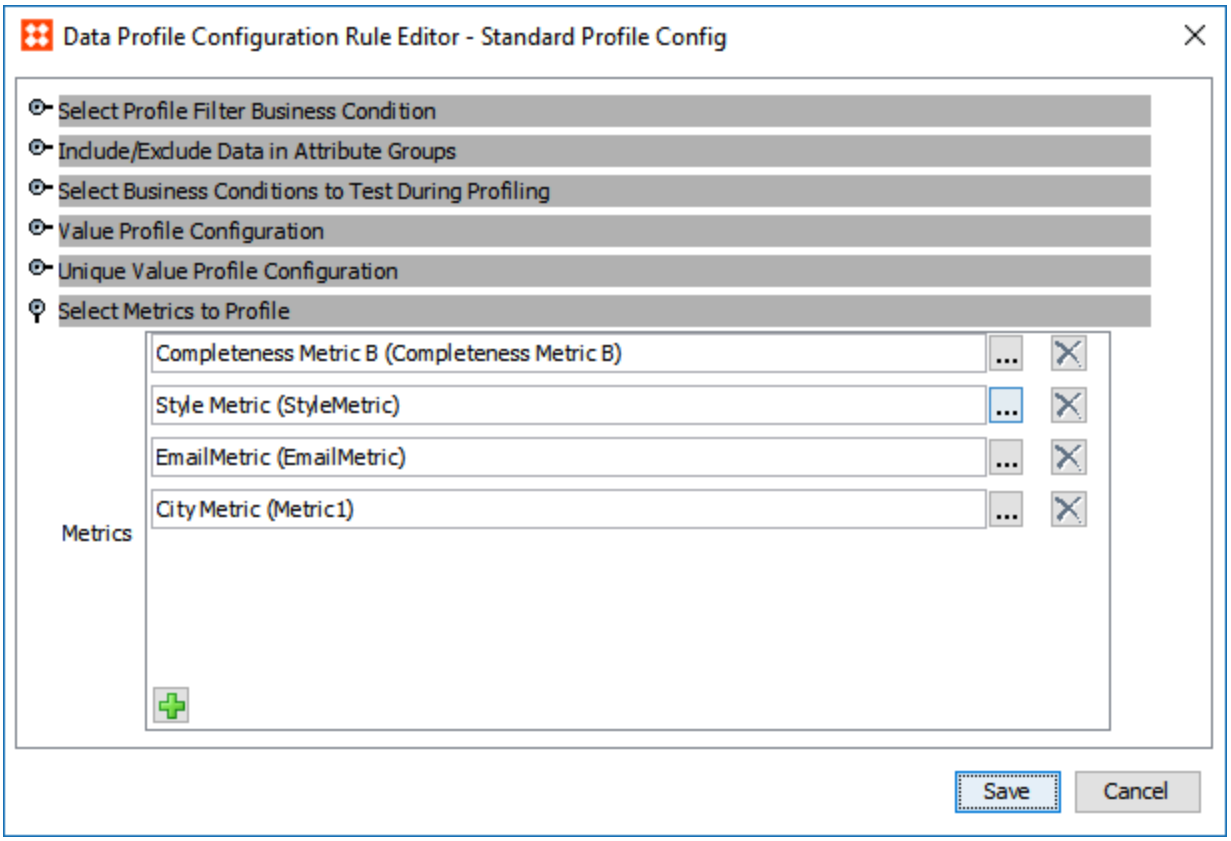


1. Navigate to the **Value Profile Configuration** section and click the green plus sign to add a new word-divided attribute to the selector field.
2. Click the ellipsis button (...), then browse or search for the relevant attribute.
3. Click **Select**. If no other attributes need to be specified, click the **Back** button when editing a local configuration or **Save** when editing a profile configurations setup entity.



Select Metrics to Profile

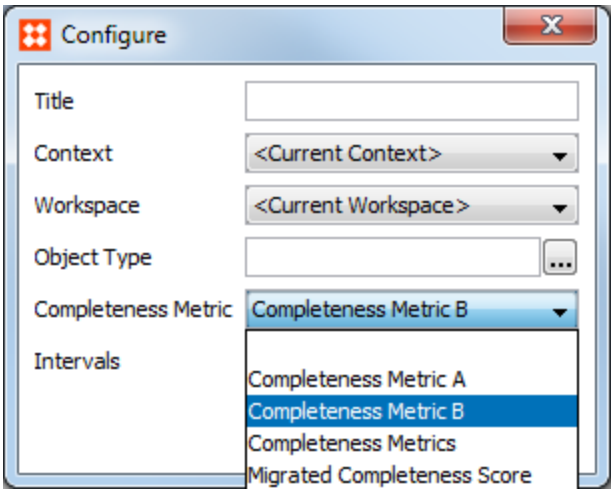
The **Select Metrics to Profile** parameter allows users to specify which metrics to use when profiling the data.



1. Navigate to the **Select Metrics to Profile** section and click the green plus sign to add a new completeness metric to the selector field.
2. Click the ellipsis button (...), then browse or search for the relevant metric.
3. Click **Select**. If no other metrics need to be specified, click the **Back** button when editing a local configuration or **Save** when editing a profile configurations setup entity.

If using the Metric Result widget, ensure that this field includes whichever metric is configured on the widget. If they do not match, a 'no data available' message will appear in place of the data on the widget.

Important: Even if the metric specified by the widget is the same as the system default, it will still display the 'no data available' message. To use this widget, the proper metric must be specified in the Profile Configuration.



For more information see the **Metrics** section of the **System Setup** documentation.

Data Profiling in Web UI

Data profiles can be generated from within the Web UI via the **Update / Refresh Profile** actions and displayed via the **Data Profile** component screen.

Note: The **Data Profile** screen is available for product, entity, classification and collection profiles.

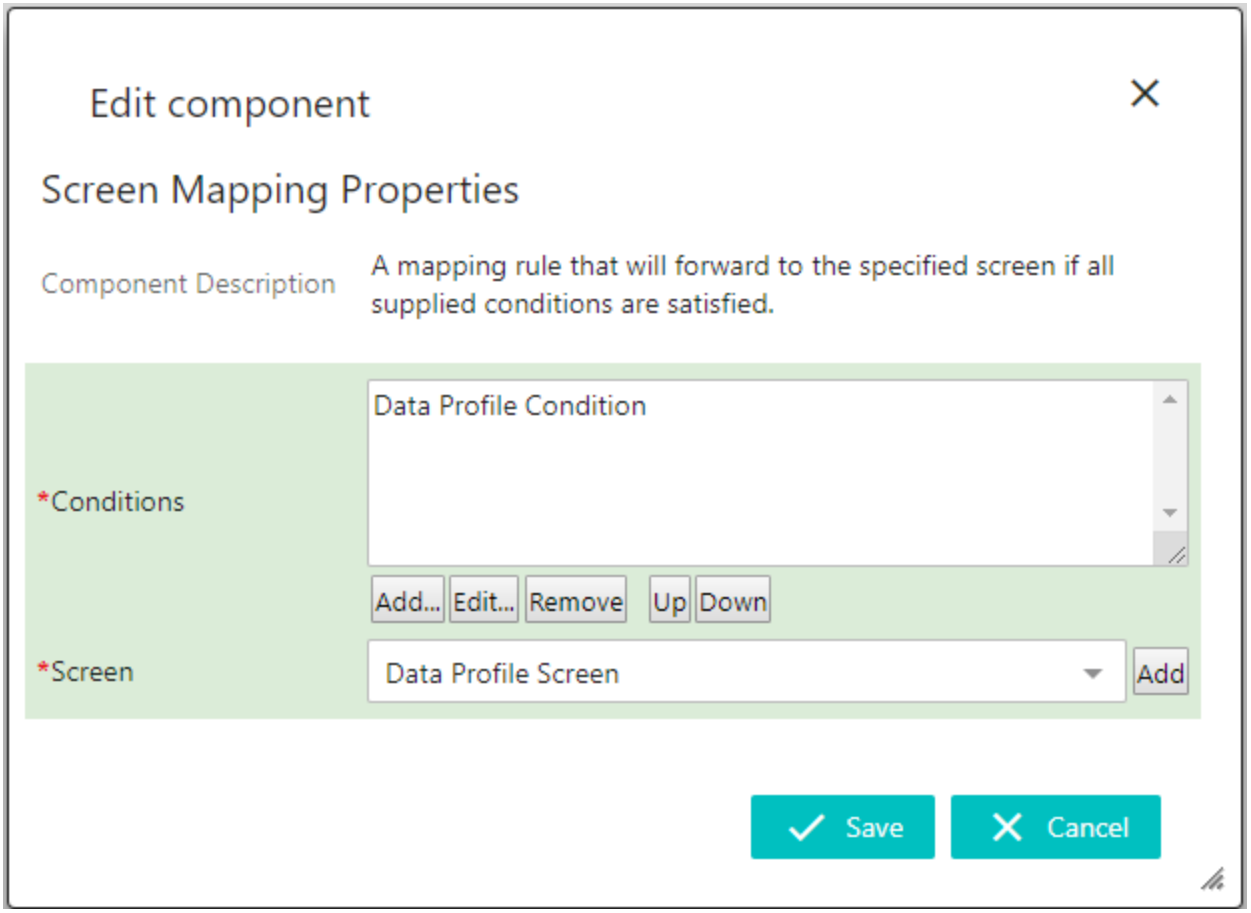
Accessing the Data Profile Screen

The Data Profile screen Web UI component is not configurable like profile dashboards in the workbench and has a uniform format for displaying profile information. However, a number of initial setup steps are required to access profiles in Web UI.

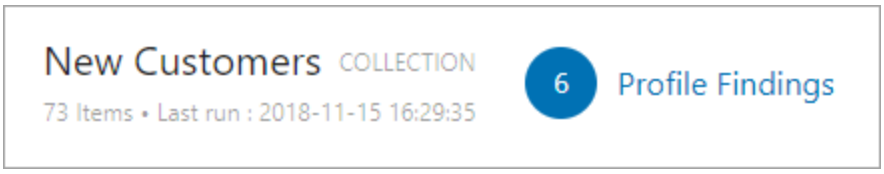
Note: Data profile dashboard widgets are available in the Web UI via KPI widgets. For more information, see the **Data Profile Widgets in Web UI** section of the **Web User Interfaces / Web UI Setup and User Guide** documentation.

---

1. The 'reactive-webui', 'graphql', and 'profiling' components must be activated. These are all included in baseline installations. All three must be present on your system to access and use the Data Profile screen.
2. Create a new screen and select the **Data Profile** component. The component itself does not require any configuration once the screen has been created. For more information on creating screens, see the **Design Mode Basics** section of the **Web User Interfaces / Web UI Setup and User Guide** documentation.
3. Map the new **Data Profile** screen to the **Data Profile Condition**. For more information on mapping screens, see the **Mappings** section of the **Web User Interfaces / Web UI Setup and User Guide** documentation.



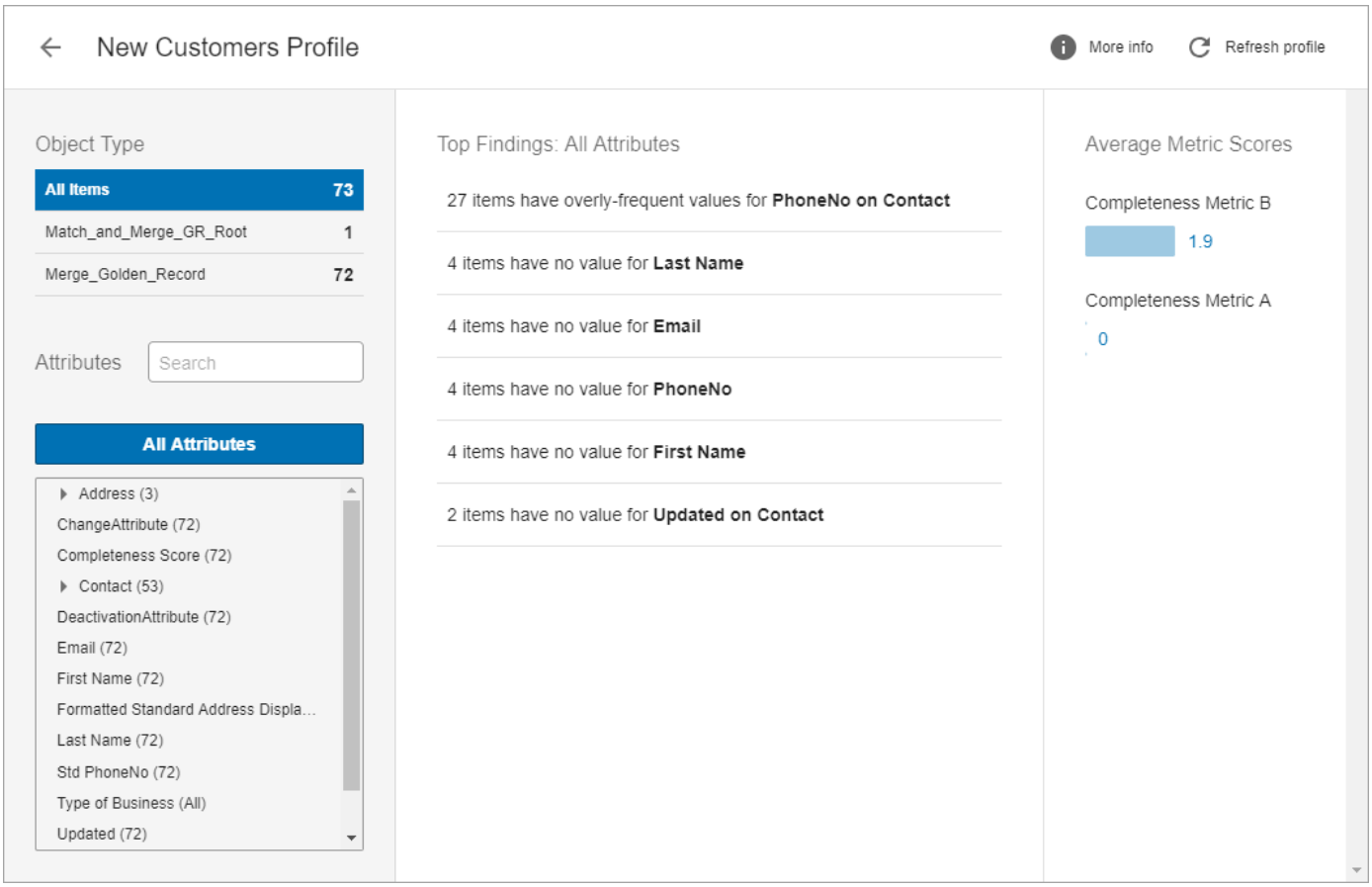
4. Enable **Include Profile** in the Collection / Entity Summary component of the applicable Entity and/or Collection screen. When enabled, a link to the data profile of the currently selected node is provided in the summary component. This link also displays the number of **Profile Findings** associated with the data profile.



For more information on Profile Findings and the Data Profile screen, see the **Data Profile Screen** section of the **Web User Interfaces / Web UI** documentation. For more information on configuring the Entity Summary and Collection Summary components, see the **Below Title Components** section of the **Web User Interfaces / Web UI** documentation.

Note: The **Profile Findings** link will only appear for nodes that have already been profiled.

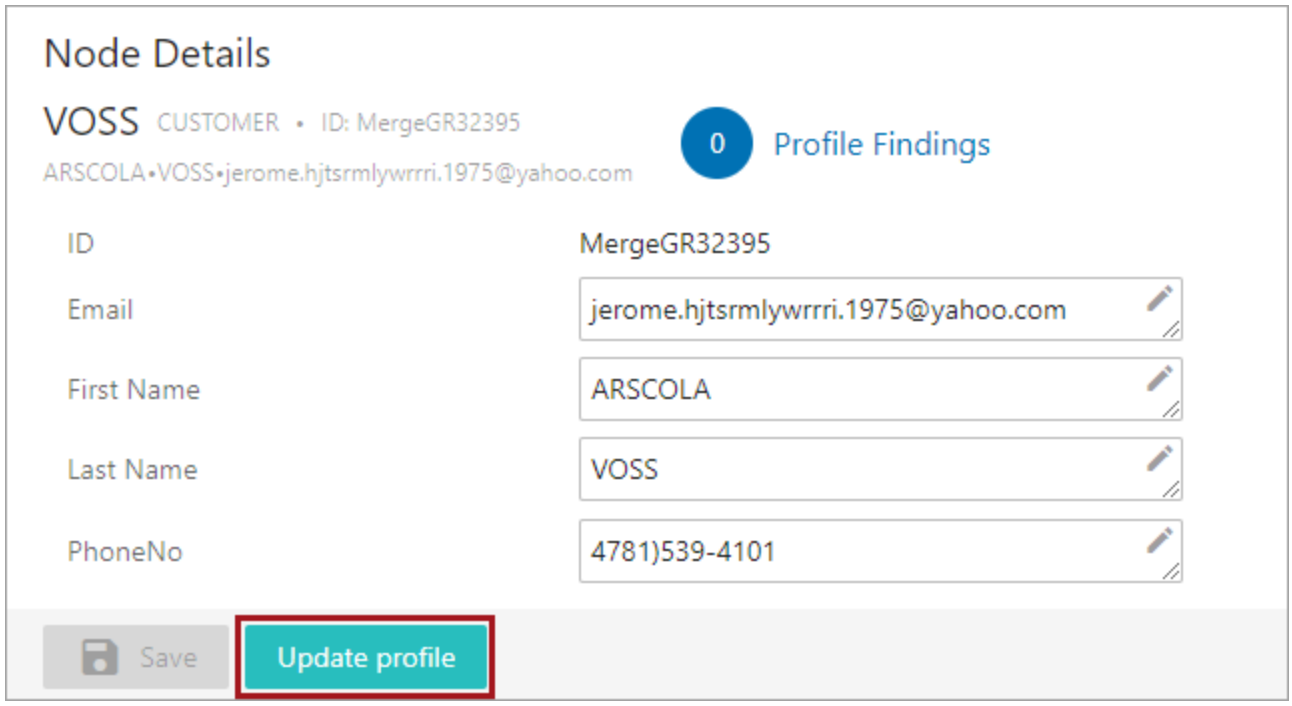
Once the above setup has been completed, the user can access an entity's or collection's data profile by navigating to the node in question and clicking the **Profile Findings** link.



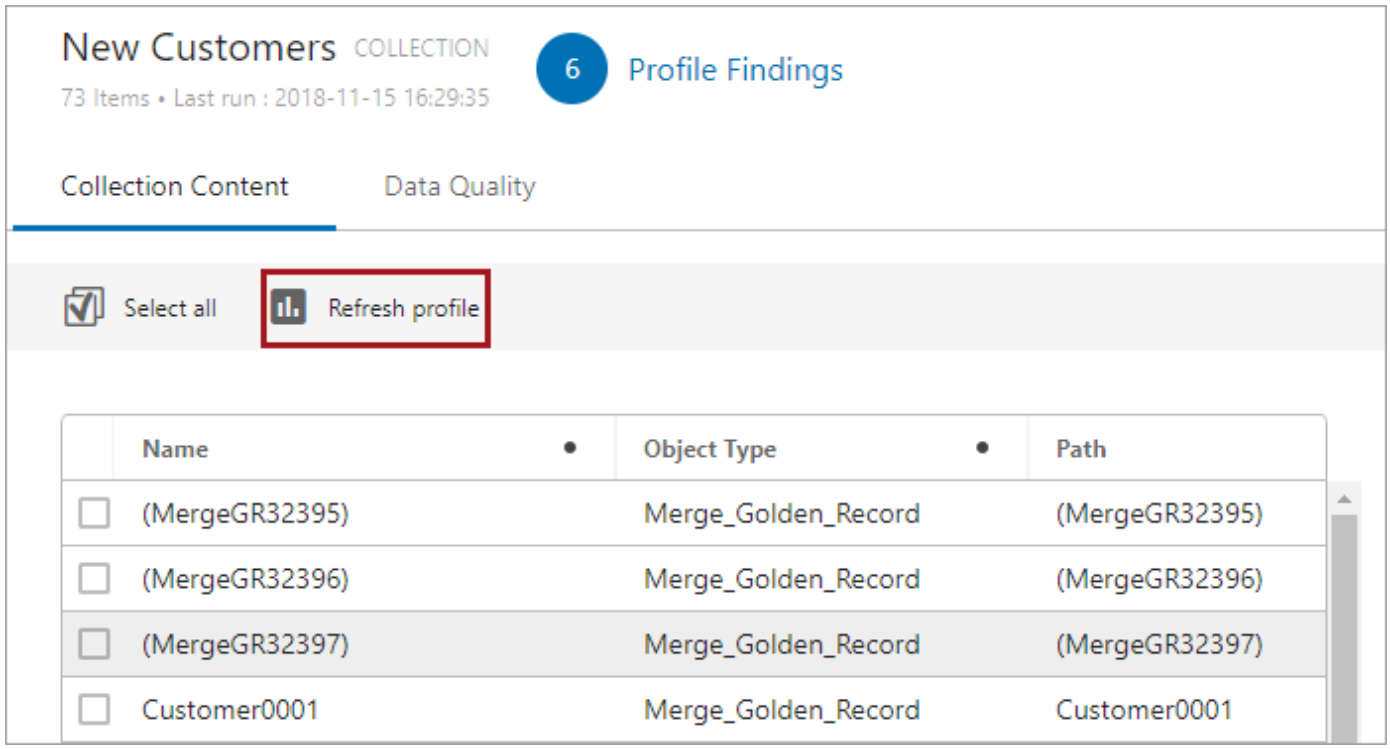
Creating and Updating Data Profiles in Web UI

Data profiles can be created / updated via two different action buttons in Web UI: the **Update Profile Action** and the **Refresh Profile Action**.

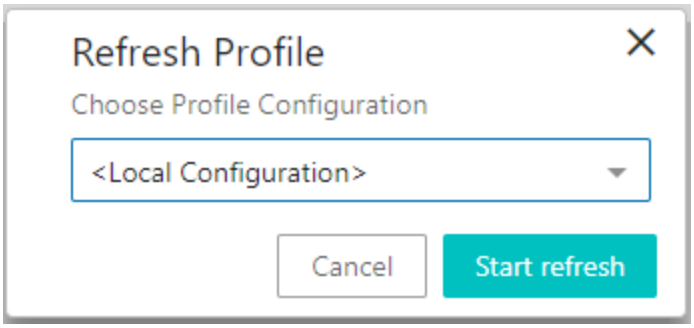
The **Update Profile Action** can be configured on a Node Details screen and can be used for any object with profiling enabled.



The **Refresh Profile Action** can only be configured on a Node List component and only for collection objects.



When either action button is clicked, a **Create / Refresh Profile** window will appear and prompt the user to select a profile configuration via a dropdown menu. Once a configuration is selected, click **Create profile / Start refresh** to create / update the profile.



For more information on profile configurations, see the **Profile Configuration** section of this documentation. For more information on the action buttons, see the **Action Buttons** section of the **Web User Interfaces / Web UI** documentation.

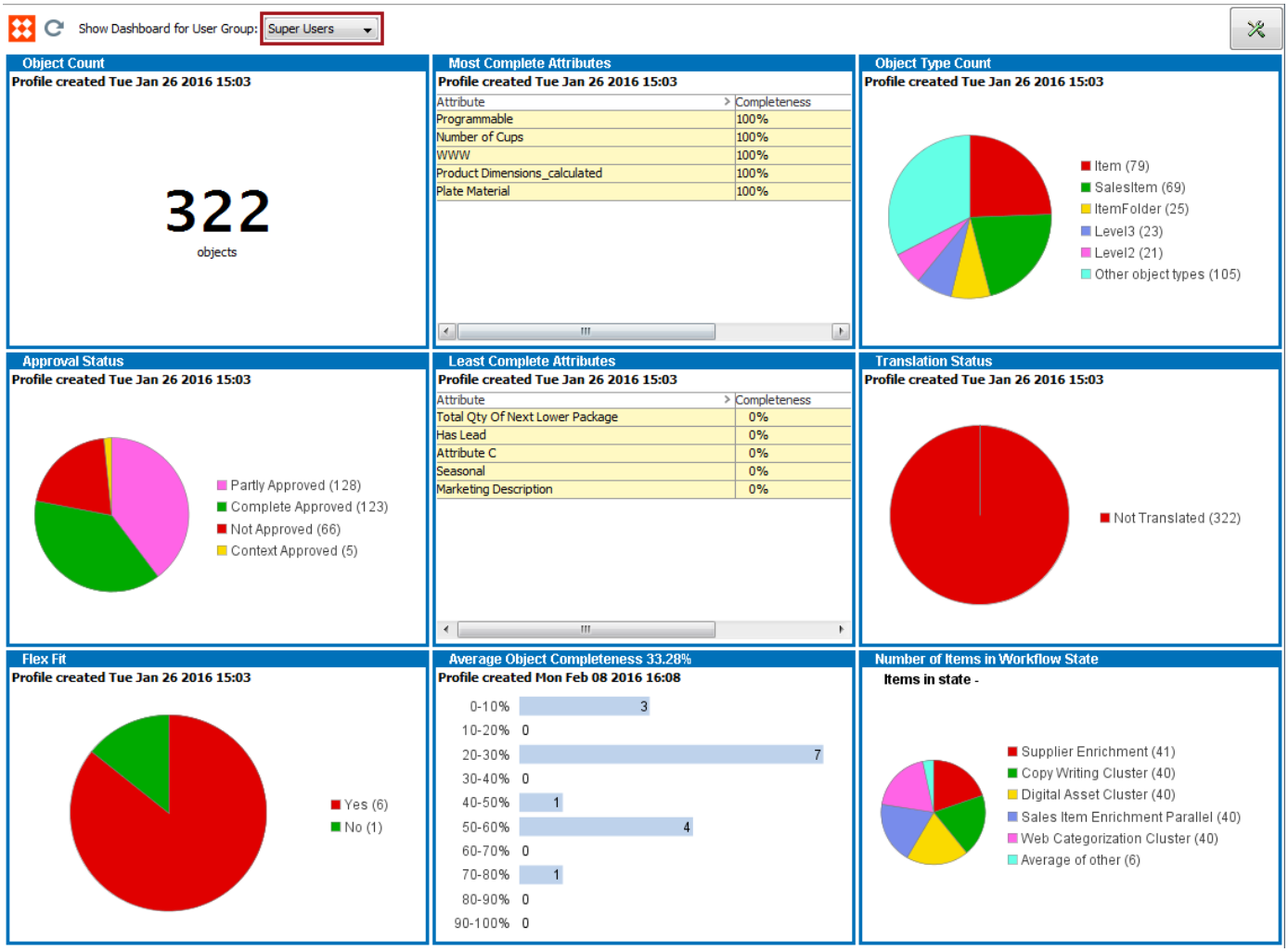
Global Dashboard

The global dashboard is a highly customizable view that allows users to display profiled data using a collection of configurable widgets. Multiple dashboards can be configured to display the most relevant data for specific user groups. Additionally, widgets can be configured to display data from specific category / data profiles. This allows users to view profile information without having to navigate through the **Tree** to locate the relevant profile.

Displaying the Global Dashboard

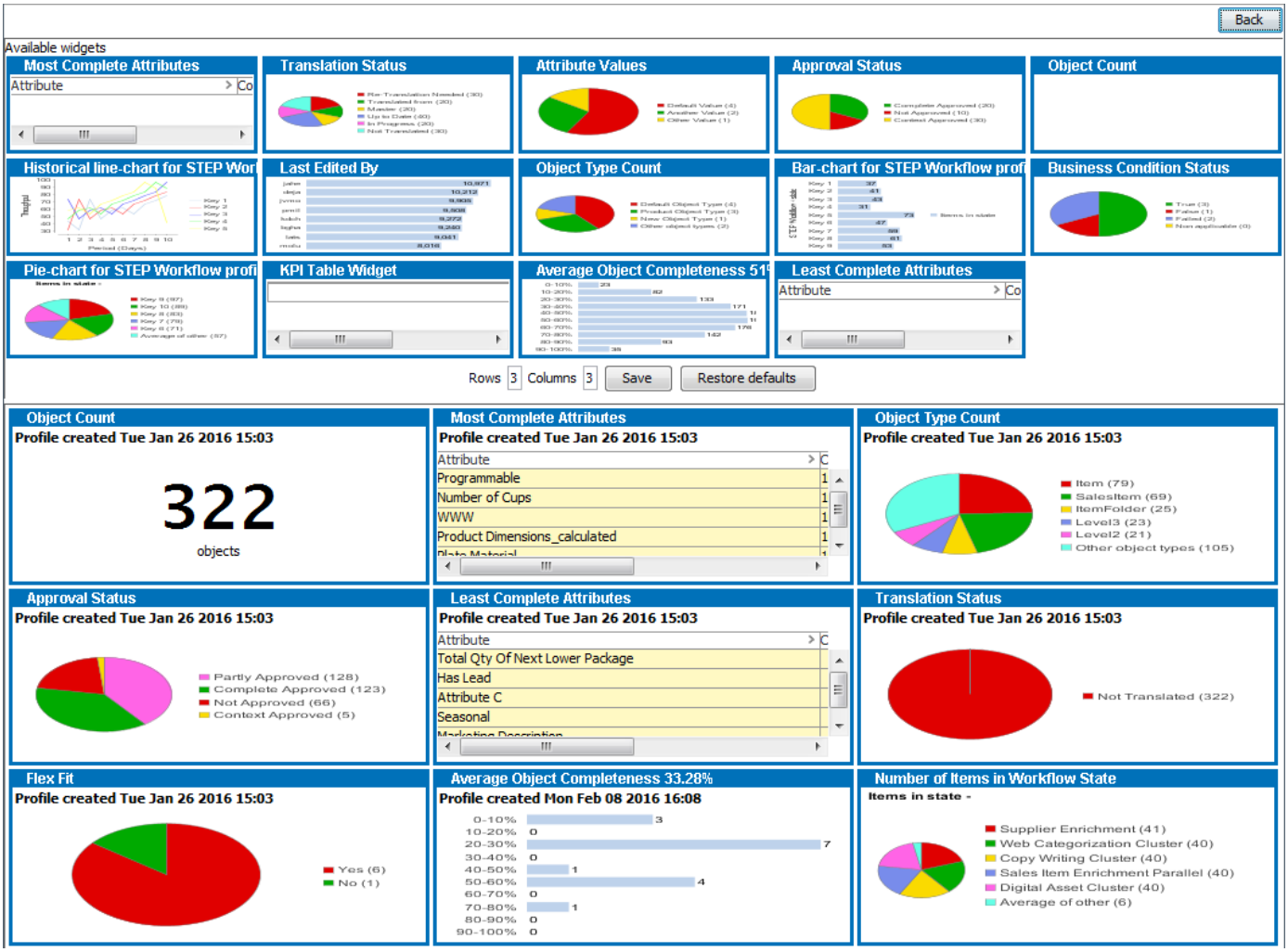
1. To view the dashboard, click the STEP logo icon in the upper left corner of the workbench.

2. From the **Show Dashboard for User Group** list, select a user group you belong to view their dashboard.

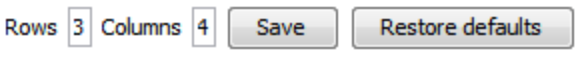


Customizing the Global Dashboard

1. On the **Global Dashboard**, select the preferred User Group, and then click the **Tools** icon  . The configuration view of the dashboard is displayed. Note that the upper half of the view displays the available widgets, and the lower half is a preview of the dashboard. Adding these widgets to the dashboard is detailed in Step 3 below.



2. In the **Rows** and **Columns** fields, located between the available widgets and dashboard preview areas, specify how many rows and columns you want on the dashboard.



3. Configure the widgets for the currently selected dashboard:
   - To add a widget, left-click the header bar of the preferred widget and drag the widget to an empty area in the preview.

- To remove a widget, in the preview, right-click the header bar of the widget you want to remove, and then select **Remove**.
- To resize a widget, in the preview, right-click the header bar of the widget you want to re-size, mouse over **Width** or **Height**, and select the preferred value.
- To configure what data is displayed, in the preview, right-click the header bar of the applicable widget, and then select **Setup**.

For more information about configuring specific widgets, see the **Available Widgets** section of the **Data Profiling** documentation.

The main difference between configuring widgets for the global dashboard and for category /data profiles is that for each widget the user must specify which category / data profile to draw data from. The Workflow Profile widgets are specific to the global dashboard and are detailed below.

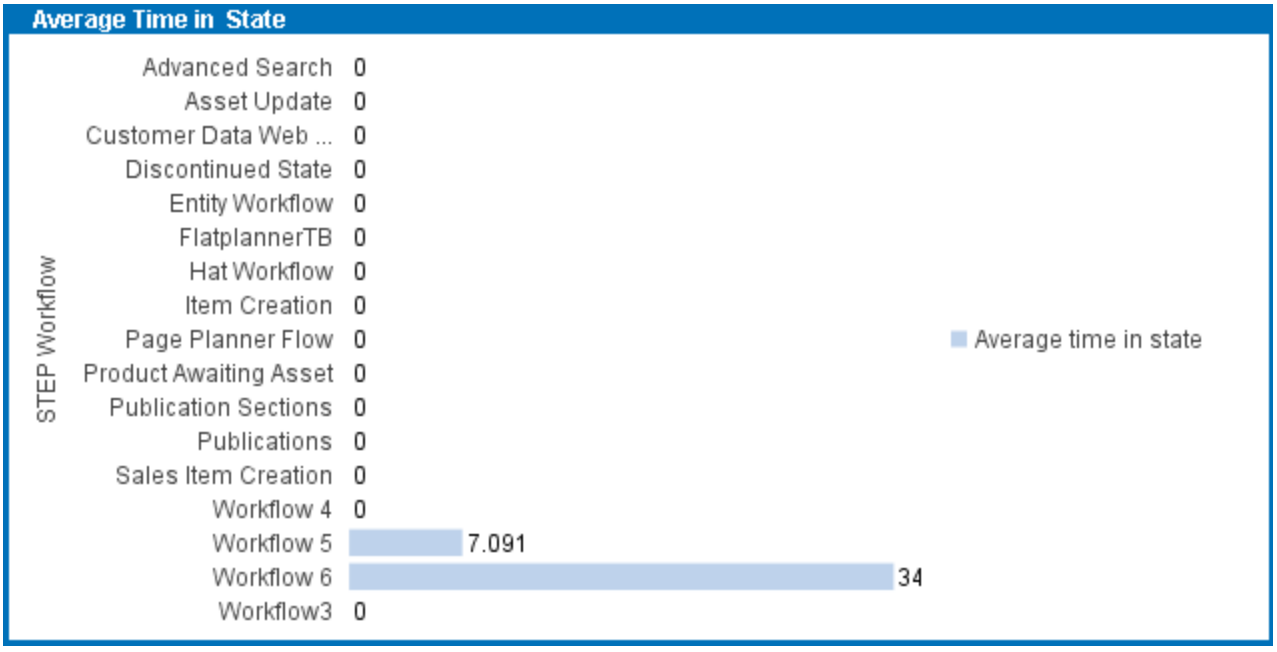
4. Once the dashboard has been configured, click **Save** and then click the **Back** button at the top to return to the main global dashboard view.
5. Click the **Refresh** icon for the changes to take effect.

Workflow Profile Widgets

Like the other widgets, the Workflow Profile widgets need to be configured to display the desired data.

Bar-chart for STEP Workflow Profile

This widget displays data for any available Workflow Profile in a bar graph.

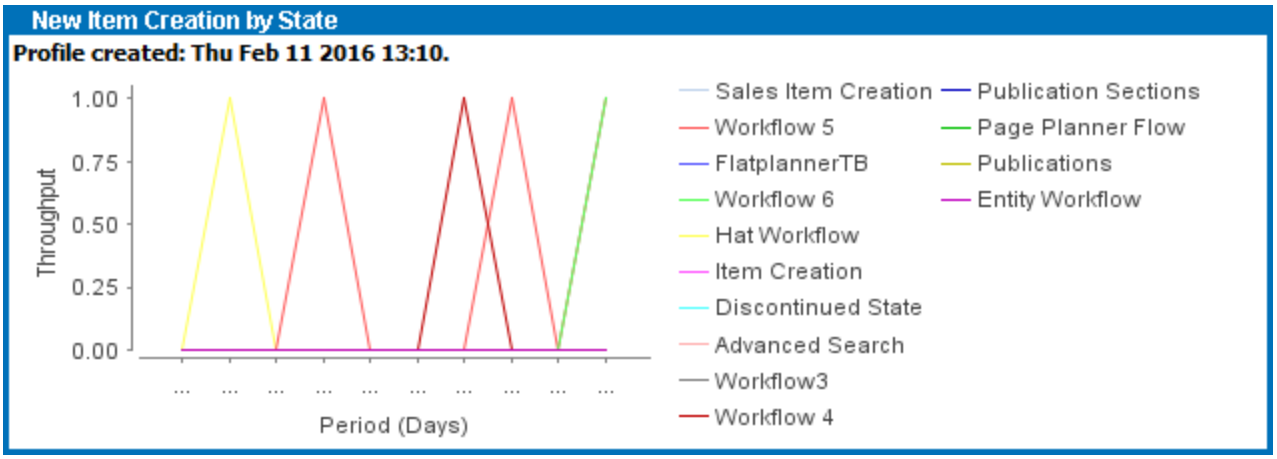


Configuration options include:

- **Chart Title** - Changes the name that appears on the widget's title bar.
- **Available Profiles** - Pulls data from the specified Workflow Profile.
- **Legend** - Determines which aspect of the Workflow Profile to model the data around (such as the workflows themselves or the states of the workflows).
- **Value** - Determines what values of the Workflow Profile to measure (such as 'Max Time in State', 'Items in State', etc.).
- **Orientation** - Determines if the bar graph displays vertically or horizontally.

Historical line-chart for STEP Workflow Profile

This widget displays historical data for any available Workflow Profile in a line graph. To ensure this widget is displaying up-to-date data navigate to the applicable STEP Workflow Profile node and click the **Update on-demand data** button.

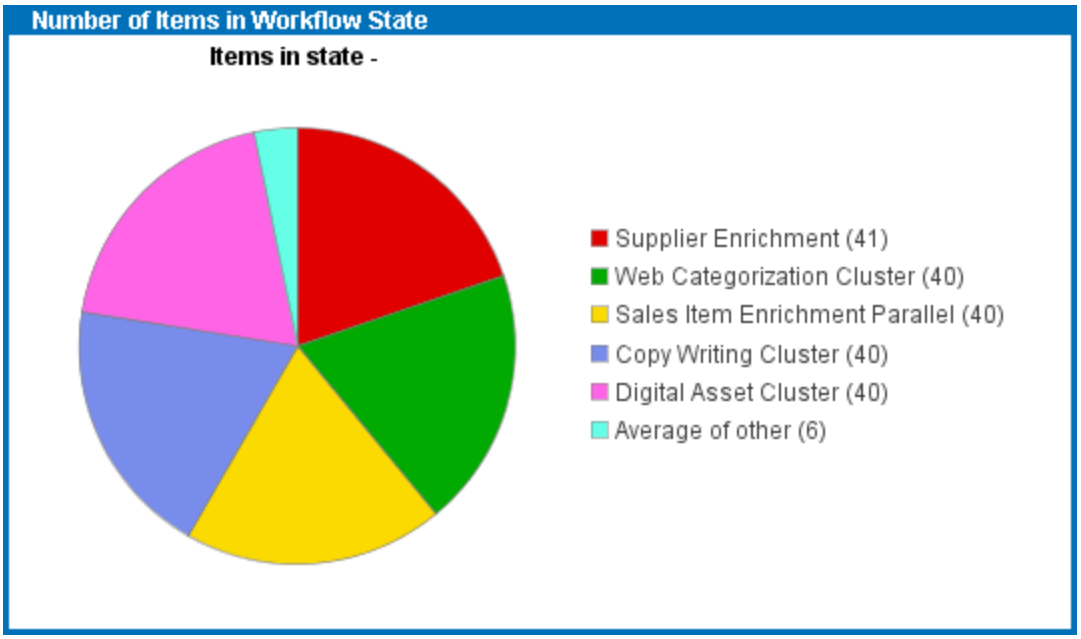


Configuration options include:

- **Chart Title** - Changes the name that appears on the widget's title bar.
- **Available Profiles** - Pulls data from the specified Workflow Profile.
- **Legend** - Determines which aspect of the Workflow Profile to model the data around (such as the workflows themselves or the states of the workflows).
- **Historical Column** - Determines which column of the Workflow Profile to display data from. The options may very based on how the individual profiles are configured in the system.

Pie-chart for STEP Workflow Profile

This widget displays data for any available Workflow Profile in a pie graph.



Configuration options include:

- **Chart Title** - Changes the name that appears on the widget's title bar.
- **Available Profiles** - Pulls data from the specified Workflow Profile.
- **Legend** - Determines which aspect of the Workflow Profile to model the data around (such as the workflows themselves or the states of the workflows).
- **Value** - Determines what values of the Workflow Profile to measure (such as 'Max Time in State', 'Items in State', etc.).
- **No. of pie-slices** - Determines the maximum number of pie slices that can be displayed in the chart.
- **Show average of the other** - Displays the average value for those workflows and/or states who's values were too low to be represented by themselves.

Limiting Memory Usage During Profiling

When large categories are profiled, the system uses a lot of memory. However, the following two config properties can be used to limit memory usage.

Note: Config properties can be set on the server by advanced users or by Stibo Systems technical support.

Config properties

Use one of the following properties to limit memory usage.

`DataProfile.MaxDistinctAttributeValuesConsideredDuringProfileGeneration`

Sets the maximum number of distinct values to profile for attributes. The default setting is 100. When the limit is reached the following happens:

- Frequent value counts might become inaccurate. (STEP uses a counting implementation dedicated for counting in big data collections with a limited memory usage from Clearspring Analytics)
- The rare value count is disabled because only a frequent count can be maintained. In the profile, the frequent and rare values cells for attributes with too many distinct values are displayed with a light red background color. The attribute completeness and count and the value instance counts for profiled attributes are correct.

`DataProfile.MaxDistinctTargetsConsideredDuringProfileGeneration`

Sets the maximum number of distinct targets for the reference or link type that is profiled. The default setting is 100.

Attribute Value Profiles

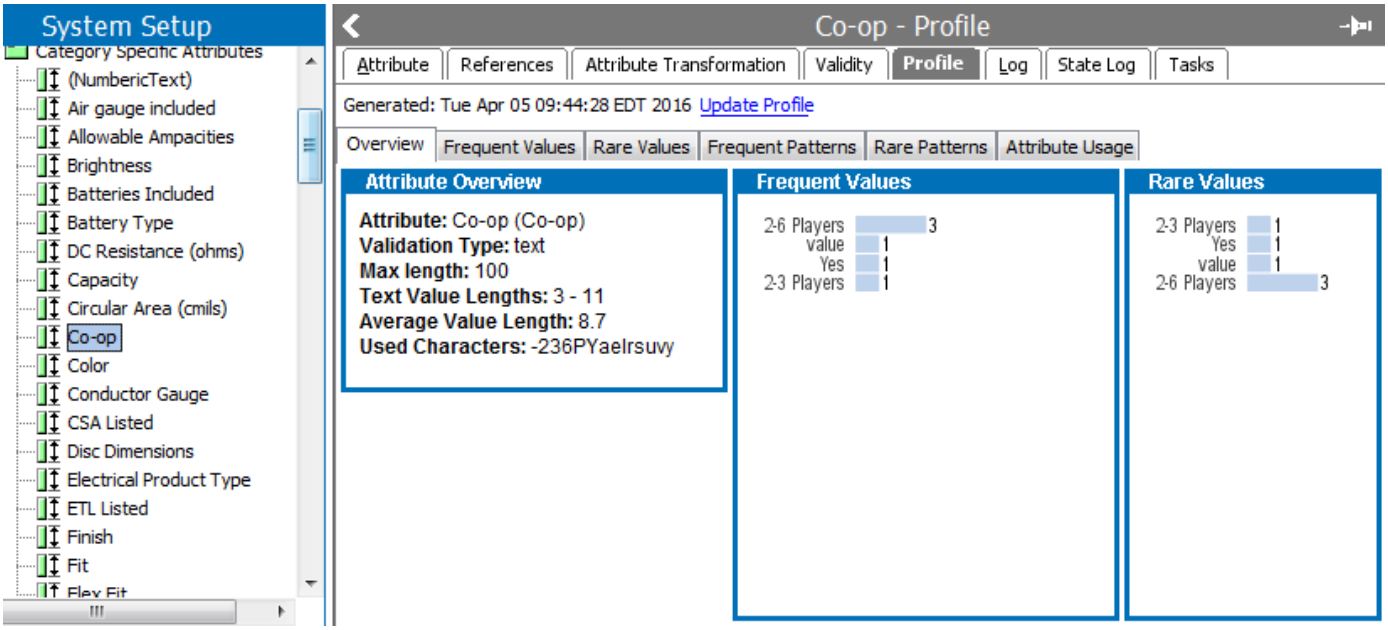
Attribute value profiles enable you to see where and how an attribute is used. The profile provides an overview of how attribute values are distributed across the current context and workspace, and enables users to easily clean up data.

Attribute profiling consists of two parts: Value Profiling and Attribute Usage.

- Value Profiling inspects all values for the selected attribute and collects information about these values.
- Attribute Usage uncovers how an attribute is used across business rules, calculated attributes, attribute transformations, integration endpoints, export and import configurations, key definitions, tables, and so on.

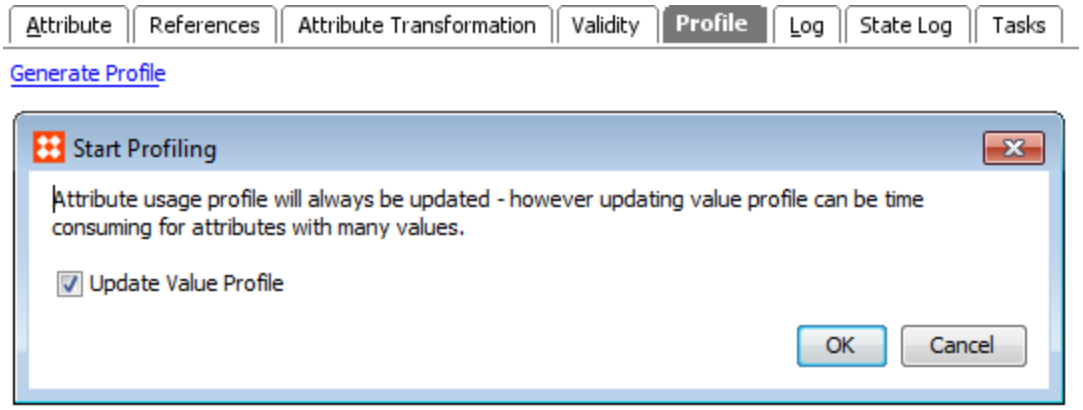
The value profile enables you to identify the most used values and the least used values. If the attribute is full text indexed, then all words are profiled as opposed to the entire value. Other aspects of the values are profiled as well.

Note: Value Profiles do not include inherited values, but do include values on objects in Recycle Bin.



View an Attribute Value Profile

1. In **System Setup**, expand **Attribute Groups**, and then select the preferred attribute.
2. On the **Profile** tab, click **Generate Profile** if no profile exists, otherwise click **Update Profile**. The **Start Profiling** message appears.



3. If you only need information about attribute usage, clear the **Update Value Profile** box, and then click **OK**.

Note: Collecting attribute values can be time-consuming depending on how prolific the attribute is in the system.

Pre-existing value profiles are not deleted if you choose only to update attribute usage. The profile is ready once the background process completes.

Config Property

To limit the system memory usage when profiling attributes, you can use the following config property:

`AttributeProfile.MaxDistinctValuesConsiduredDuringAttributeProfileGeneration`

The property limits the number of distinct values to be profiled. The property defaults to 1.000.

Attribute Value Profile Tabs

The information provided by the Attribute Value Profile is divided into a number of tabs:

- Overview
- Frequent Values
- Rare Values
- Frequent Patterns
- Rare Patterns
- Attribute Usage

Actions

All tabs, except the Overview and Usage tab, contain a table where different actions are possible. From left to right the icons represent the following functionality: Bulk Update, Search, Save as Collection, and Export Frequencies  .

- **Bulk Update** When you click the Bulk Update icon, a search is run for objects that have the selected attribute value. The result is used as input data set for the Bulk Update wizard.

  The Bulk Update wizard opens on the **Operations** step with the attribute selected in the **Set Value** operation.

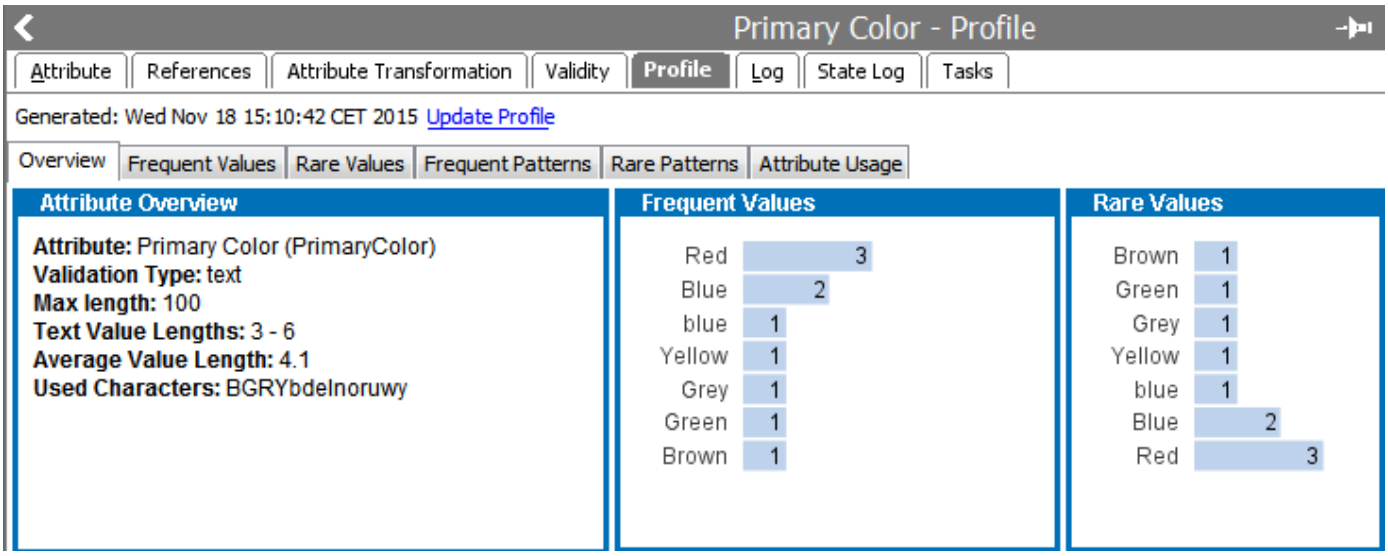
  This functionality makes it easy for you to correct data errors, and because the input data set is the result of a search, you cannot accidentally update the wrong data.

  For more information about the Bulk Update wizard, see the **Creating a Bulk Update** section of the **Getting Started / User Guide** documentation.

- **Search** - When you click the Search icon, a search is run for products that have the selected attribute value. When the search is complete, the result is displayed on the **Search** navigation tab. Searches are filtered on the object type selection in the profile.
- **Save as Collection** When you click the Save as Collection icon, the Save as Collection in Background Process dialog is displayed. Select a location and enter a Collection Name and a Collection ID, and then click OK. A search is run within for products that have the selected attribute value, and the result is saved in the collection.
- **Export Frequencies** When you click the Export Frequencies button, the Export Frequencies dialog is displayed. Select output file and if you want absolute or relative frequency. A CSV file with semi colon separated frequencies, will be create on the local system.

Overview Tab

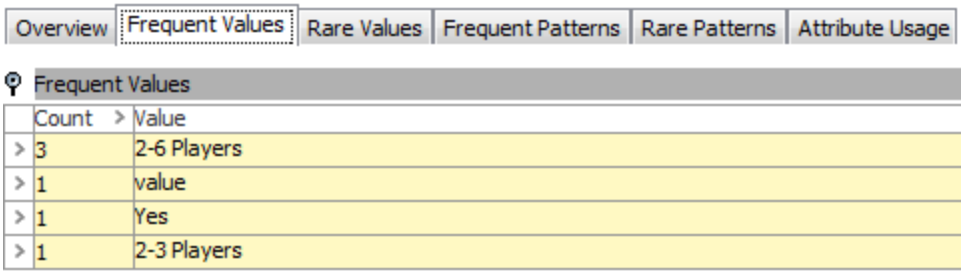
The **Overview** tab contains a number of widgets that each present different information about the selected attribute. Users cannot customize this view. Inherited values that are valid to a product hierarchy are included in this data profile.



Note: Orphaned attribute values are not included in this calculation nor are attributes that are linked via a classification hierarchy.

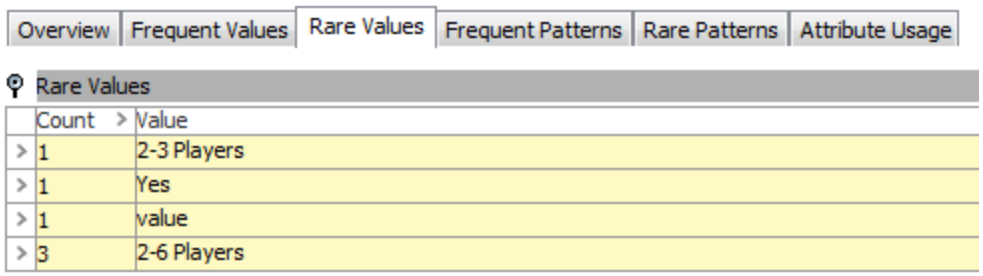
Frequent Values Tab

On this tab, the 100 most frequently appearing values of the selected attribute are listed. Tags are resolved.



Rare Values Tab

On this tab, the 100 least frequently appearing values of the selected attribute are listed.



Frequent Patterns Tab

This tab lists the most frequent patterns of the values of a given attribute and the number values that match the pattern.

A value pattern is a way of describing the structure of a value in terms of characters, numbers, and symbols. When the patterns of the values in a set are compared, certain types of errors can be detected by finding values whose patterns are different from the rest. For example, phone numbers with missing digits or numbers that are mixed with letters, and so on.

A pattern for a given value is created by replacing any letter with A, any number with 9, and any unrecognized character with X. Other symbols are left as they are.

The following is an example of a value pattern:

Value	Pattern
123-ABB	999-AAA
234.23	999.99
mail@mail.com	AAAA@AAAA.AA

The pattern [None] represents an empty value.

Rare Patterns Tab

This tab resembles the Frequent Patterns but lists the least frequently used patterns rather than the most frequent. This information is useful when searching for faulty values.

Attribute Usage Tab

On this tab users can view all the places where the attribute is used. The attribute usage can be qualified as either Used or Possibly Used. Possibly Used means there is a good indication that the attribute is used but the system cannot detect it with 100 percent certainty. This is, for example, the case in calculated formulas if the ID of the attribute is found in the formula text.