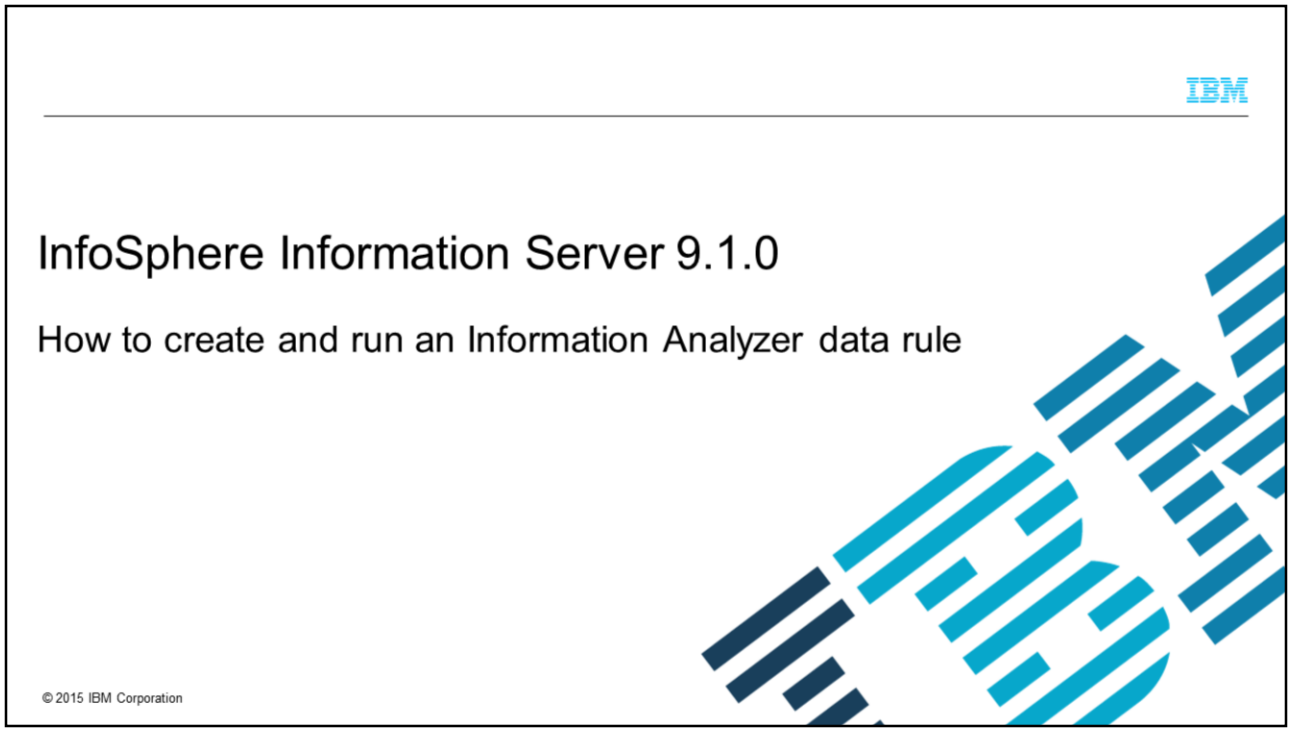


This presentation explains how to create an Information Analyzer data rule from a rule definition with the regular expression check.

Objectives

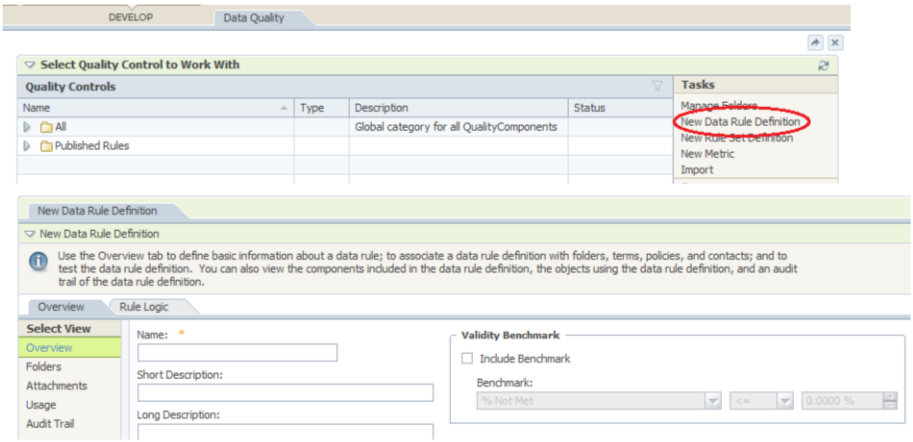
- Create a rule definition and define the rule logic
- Generate a data rule from a rule definition
- Run the data rule and access the output results

© 2015 IBM Corporation

The objectives of this presentation are to show how to create a data rule definition with rule logic and how to generate a data rule. The presentation also describes how to run the data rule and how to access output results from the IBM InfoSphere® Information Analyzer Data Quality workspace. This presentation assumes that you already created an Information Analyzer project and imported your metadata.

Create a rule definition and define the rule logic (1 of 3)

- Create data rule definition
- Click Develop => Data Quality

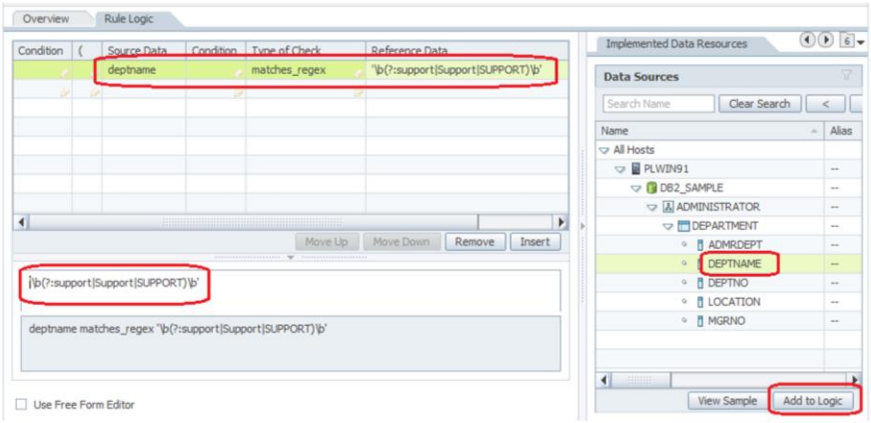


© 2015 IBM Corporation

In order to work with IBM InfoSphere Information Analyzer rules, go to the Develop icon on the Navigator menu, and select Data Quality from the menu. On the Data Quality workspace, click New Data Rule Definition in the Tasks menu on the right side of the workspace. Creating a rule definition requires two components: a name for the rule definition and a logical statement or rule logic. Create a rule definition by defining the name on the Overview tab. You can optionally add a short and long description of the rule. After you create a name and a description for your rule, switch to the Rule Logic tab. Define the specific condition that you want to analyze or validate with the IA rule.

Create a rule definition and define the rule logic (2 of 3)

- Define the rule logic
- Click Add to Logic button



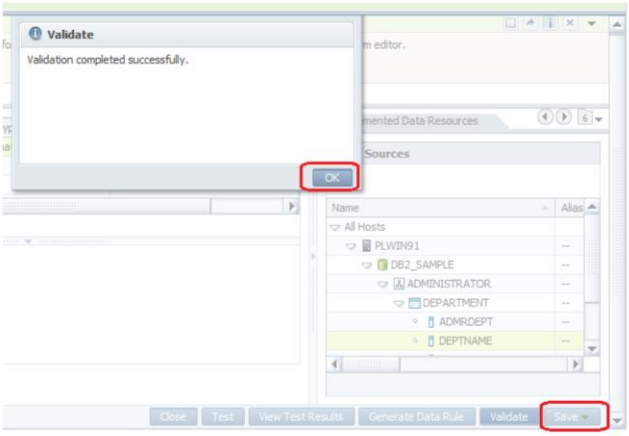
© 2015 IBM Corporation

Next, click the Rule Logic tab to define the specific condition you want to analyze or validate with the IA rule. This presentation will use the rule definition logic builder to define the data for the data rule. The rule logic builder is on the Rule Logic tab when you create a new rule definition or when you work with an existing data rule. Each line of the rule logic builder must contain at least the source data and a type of check. Most of the checks also require reference data to be set. This presentation will create rule logic with the regular expression check to search for a word in the text. You can also use the other regular expressions to find a particular variable, a range of characters or specific kinds of characters. To create the rule logic with the regular expression check, add the source data, type of the check and the reference data into the rule logic builder. Usually source data is a column reference and represents the value that you want to test in the rule definition. While in the IA rule logic builder, click in the source data box and type your source data in the text box that opens below the rule builder workspace. Alternatively, to define the source data you can use the Tabbed Pallet that is located on the right side of the screen. Click the source data box. Next, select your source data from the Data Sources tab, click on the column reference, and click the Add to Logic button. Next click the Type of Check field and from the drop-down list select the type of check that you want the rule definition logic builder to execute. Next, click the reference data box and type the expression that represents the reference data set for the data rule definition. It is important to note that the reference data can be a single value, a list of values, or a set of values against which the data rule definition compares the source data. To enter reference data, click in the reference data box and type your

reference data in the text box that opens below the rule builder workspace.

Create a rule definition and define the rule logic (3 of 3)

- Validate the rule logic
- Click Save, Validate, and Close

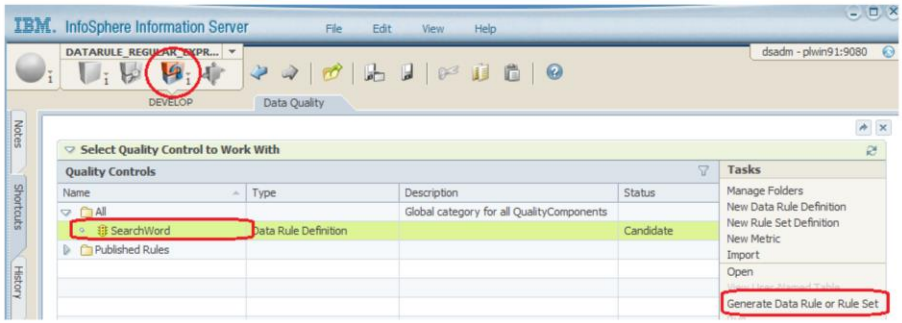


5 How to create and run an Information Analyzer Data Rule © 2015 IBM Corporation

Next, you can optionally validate the rule logic settings to verify that the rule logic definition you are creating is correct. Click the Validate button under the rule logic builder. Once the validation completes successfully, a new pop-up window will open. Click OK to close it. Save your changes by clicking the Save button in the lower right corner of the screen.

Generate a data rule from a rule definition (1 of 3)

- Generate Data Rule
- Click Develop => Data Quality
- Click Generate Data Rule or Rule Set



© 2015 IBM Corporation

Once the rule logic settings are validated, you can generate a data rule to analyze the real data in your project. From the Develop icon on the Navigator menu, select Data Quality. In the Data Quality workspace, you can find all the created data rule definitions, data rule set definitions, and data rules in this project.

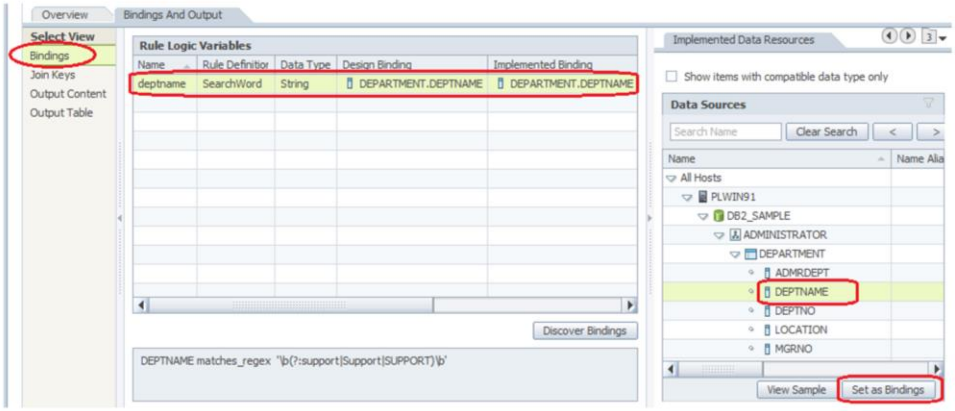
Highlight the rule definition that you want to generate a data rule for. Next, on the Data Quality workspace, click Generate Data Rule or Rule Set in the Tasks menu on the right side of the screen. In this presentation, the name of the data rule definition is SearchWord.

When the new workspace opens, you need to specify a name for the data rule on the Overview tab. The name must contain at least one character, it cannot contain the backslash (\) character, and it must be unique in your project.

Optionally you can type a short description and long description for your data rule. The Created By, Created On, and Last Modified fields are automatically populated after you create and save your data rule. You can also provide information for the Owner and Data Steward fields. You will need to decide whether you would like to set a validity benchmark for your data rule. The validity benchmark quantifies the quality of your data as well as monitors your data. The validity benchmarks are not covered in this presentation.

Generate a data rule from a rule definition (2 of 3)

- Click the Bindings and Output tab
- Check bindings for a data rule



© 2015 IBM Corporation

Next, at the top of the workspace, switch from the Overview tab to the Bindings and Output tab. Click Bindings in the Select View menu on the left side of the screen.

When you set bindings for a data rule, you connect Rule Logic variables to either Implemented Data Resources, Global variables, or Literal Values. This connection assigns values to the logical elements of your definition. In this example, the Rule Logic variable “deptname” is bound to the Implemented Data Resource “DEPTNAME”, which is a column in the table “DEPARTMENT”.

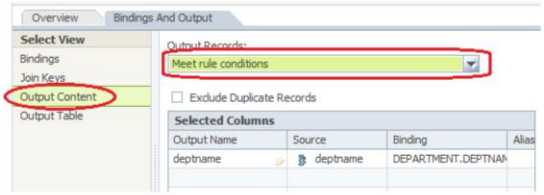
Under Rule Logic Variables, click the box for the data rule element that you are working with. With the box you are working with still highlighted, move your cursor over to the Tabbed Palette in the right side of the screen.

Select the Implemented Data Resources tab. Click the data that you want to bind to your data rule element. Next, click the Set as Binding button in order to bind your data rule element to the actual column in one of your data source tables. In this example, the Design Binding is DEPARTMENT.DEPTNAME. The Implemented Binding includes implementation models that implement the Design Binding DEPARTMENT.DEPTNAME.

In this example, Join Keys is not going to be set so this step will be skipped.

Generate a data rule from a rule definition (3 of 3)

- Specify the output content
- Select the output records that you want to display
  - Default: **Do not meet rule conditions**

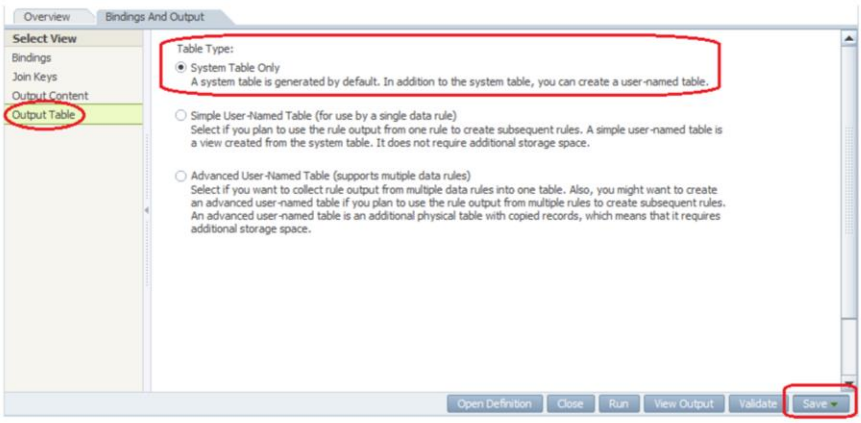


© 2015 IBM Corporation

For each source record, the data can meet or does not meet the logical conditions that are expressed in your data rule. You can specify the output that you want to see from particular columns in the output tables produced when you executed a data rule. Click Output Content in the Select View menu on the left side of the screen. Next, select the output records that you want to display in your output from the Output Records menu. The default output type is Do not meet rule conditions. Also, if you select Statistics Only, then no output details records are created.

Generate a data rule from a rule definition (3 of 3)

- Specify the output table for data rule
- Click Save to create the data rule

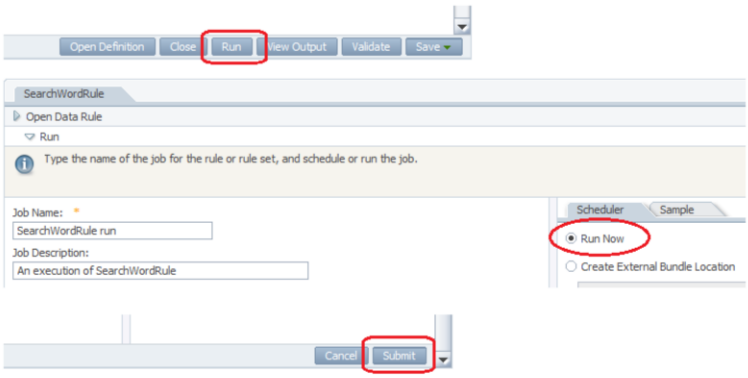


© 2015 IBM Corporation

This example uses the default table type of System Table only. If you want to use output tables from data rules as data sources for other tasks, you can select user-named tables. When the default option is selected, no user-named rule output table is associated with the rule. Once you completed all the steps for generating the data rule, click the Save button in the lower right corner of the screen.

Run the data rule and access the output results (1 of 3)

- Click the Run button
- Click Submit button

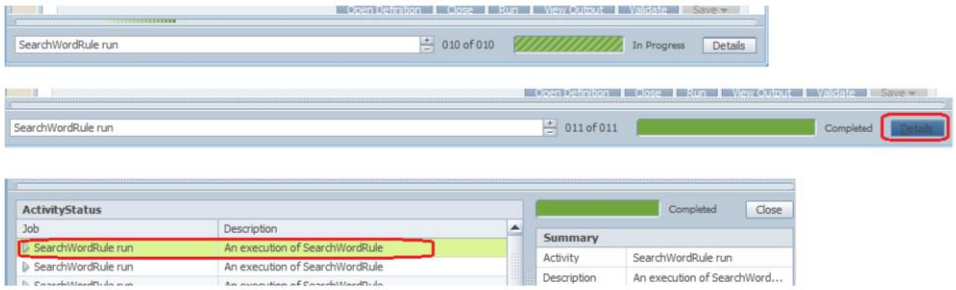


© 2015 IBM Corporation

Next click the Run button and select the option "Run Now" and Submit the data rule for execution.

Run the data rule and access the output results (2 of 3)

- Hover over the job Activity Status
- Click Details button
- Expand the job for more information

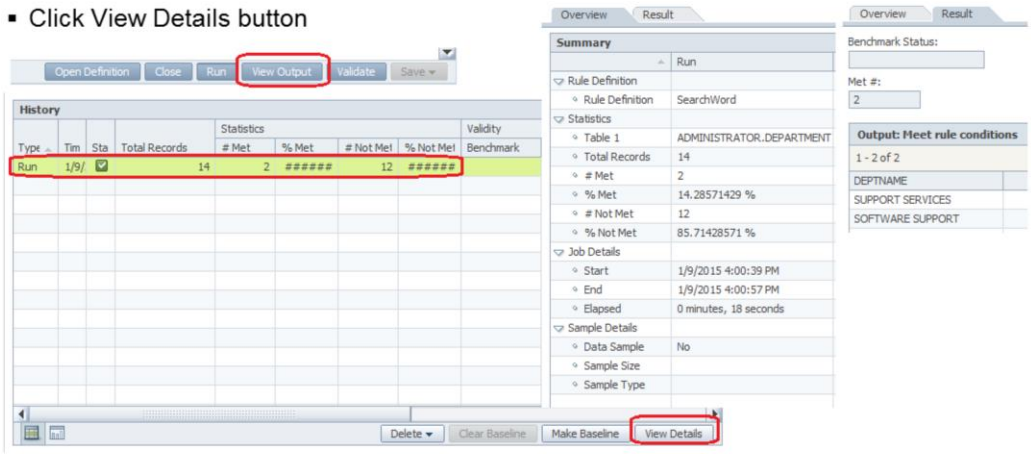


© 2015 IBM Corporation

On the Activity Status panel, hover the cursor over the job Activity Status to monitor the data rule execution. Click the Details button in order to review if it completed successfully and for more details.

Run the data rule and access the output results (3 of 3)

- Click View Output button
- Select history run
- Click View Details button



© 2015 IBM Corporation

From the details screen, click View Output, select the last run, and click View Details. This will bring you to the Data Quality workspace where you can view the individual output results by clicking on the Result tab.

Trademarks, disclaimer, and copyright information

IBM, the IBM logo, ibm.com, IA, and InfoSphere are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

IA, is a trademark or registered trademark of Cast Iron Systems, Inc. (or its affiliates), an IBM Company, in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.

© Copyright International Business Machines Corporation 2015. All rights reserved.