Axon Data Governance 6.0 Checklist to Automate Data Quality Rules
Abstract

In Axon Data Governance version 6.0, you can automate the process of running data quality rules on Enterprise Data Catalog fields instead of manually generating the mappings and producing data quality scores. You can use Enterprise Data Catalog resources to scan for data objects, send the values of the data sources to Informatica Data Quality, and run rules on thousands of fields instantly. Axon retrieves the scores of the data quality rules from Informatica Data Quality and displays the scores for each field.

This document describes the process to automate data quality rules. It contains a checklist of the steps that you must perform so that Axon can onboard resources and fields from Enterprise Data Catalog and display data quality scores from Informatica Data Quality.

For details and guidelines on automating data quality rules, see the Automate Data Quality Rules chapter in the Axon Data Governance 6.0 User Guide.

Supported Versions

- Informatica Axon Data Governance 6.0

Table of Contents

Automating Data Quality Rules Process. .............................................. 3
Checklist for Informatica Data Quality. .............................................. 5
Checklist for Enterprise Data Catalog. .............................................. 6
Automating Data Quality Rules Process

The following image shows the steps required to set up Axon with Enterprise Data Catalog and Informatica Data Quality so that Axon runs the rules automatically on data store fields:
The following steps describe how to automate data quality rules:

1. Set up Axon, Enterprise Data Catalog, and Informatica Data Quality.
2. Link objects as required between Axon and Enterprise Data Catalog. Link objects as required between Axon and Informatica Data Quality.
3. Onboard objects from Enterprise Data Catalog to Axon. Optionally, manually create links between Enterprise Data Catalog fields and Axon attributes.
4. Run rules on the objects onboarded from Enterprise Data Catalog in Informatica Data Quality.

Step 1. Set Up Axon, Enterprise Data Catalog, and Informatica Data Quality

The first step to automate rules is to set up Axon, Enterprise Data Catalog, and Informatica Data Quality by creating the required objects.

Perform the following steps to create the objects:

1. In Enterprise Data Catalog, create a resource to scan data sources, and configure the resource to discover data profiles and data domains.
2. In Axon, create a System object that maps to the resource you created in Enterprise Data Catalog. Link the System object to at least one Enterprise Data Catalog resource to enable onboarding objects from Enterprise Data Catalog to Axon.
3. In Axon, create a Glossary object.
4. Create an Axon resource type in Enterprise Data Catalog. Use the Axon resource type to scan the Axon glossaries.
5. In Axon, create a standard data quality rule for the glossary objects. The standard data quality rule is a textual definition of the rule that you want to run on the fields.
6. In Informatica Data Quality, create rule specifications and rule mapplets to run the rules on the fields. If the required rule specifications and rule mapplets already exist, modify them to conform to the rule automation guidelines. The rule specifications and rule mapplets contain the logic of the rule that you want to run on the fields.

Step 2. Link Objects Between Axon, Enterprise Data Catalog, and Informatica Data Quality

The second step is to link the objects in Axon and Enterprise Data Catalog to onboard the fields, and link the objects in Axon and Informatica Data Quality to connect the rules.

Perform the following steps to link the objects:

1. In Axon, associate the Glossary object with the Enterprise Data Catalog data domains. In Enterprise Data Catalog, you can associate data domains and fields with the Axon glossary.
2. In Axon, link the standard data quality rule with a rule specification or rule mapplet in Informatica Data Quality. This step associates a textual rule definition in Axon with a logical rule in Informatica Data Quality.
3. In Axon, choose the option to automatically create local data quality rules for the standard data quality rule. This step enables data quality rule automation.

Step 3. Onboard Objects from Enterprise Data Catalog to Axon

The third step is to display the objects from Enterprise Data Catalog in Axon. You can manually link the attributes in Axon with the fields in Enterprise Data Catalog. You can also choose to automatically onboard the objects from Enterprise Data Catalog to Axon. When you choose to onboard objects, the data sets and attributes are automatically created and onboarded to Axon.

Perform the following steps to onboard the objects:

1. In Axon, choose the option to automatically onboard the objects from Enterprise Data Catalog.
2. After the key elements are identified, the data sets and attributes are automatically created and onboarded to Axon. Axon is now ready to send the fields to Informatica Data Quality to run the rules.

3. In Axon, you can choose to accept or reject the Attribute objects created automatically.

**Step 4. Run Informatica Data Quality Rules on Onboarded Objects**

The fourth step is to run the rules automatically. Axon automatically runs Informatica Data Quality rules on Enterprise Data Catalog fields and displays the scores.

Axon performs the following steps to display the scores:

1. Axon creates a local data quality rule for the attributes that apply to each data set. In this step, Axon creates an instance of a rule for each field scanned by Enterprise Data Catalog. Axon then sends the rule mapping requests for the attributes to Informatica Data Quality.

2. Informatica Data Quality receives the mapping requests and runs the rule specification or rule mapplet on the fields as per the schedule that you have defined.

3. Informatica Data Quality generates data quality scores for each field.

4. Axon retrieves the data quality scores for each table column and displays the scores for each local rule.

**Checklist for Informatica Data Quality**

Make sure that you set up Informatica Data Quality according to the following guidelines:

1. You have Informatica Data Quality version 10.2.0 HotFix 1.

2. The following services are running on Informatica Data Quality:
   - Model Repository Service with monitoring enabled
   - Data Integration Service
   - Content Management Service
   - Scheduling service configured to the Model Repository Service

3. The Informatica Data Quality administrator has defined the rule specifications and rule mapplets according to the following specifications:
   - The output value for the rule must be present in the first column of the output table.
   - The output data type must be a string value in one of the following formats:
     - 0 or 1
     - Yes or No
     - True or False
     - Valid or Invalid

4. Run the data quality rule on a test data set and verify that the output conforms to guidelines mentioned in the previous step.

5. Informatica Data Quality can connect to the sources scanned by Enterprise Data Catalog. Make sure that all critical connection parameters are configured in Informatica Data Quality, such as the password and Support Mixed-Case Identifier option for the resource.

6. Axon can connect to Informatica Data Quality. For more information, refer the Configure Access to Informatica Data Quality topic in the Informatica Axon Data Governance 6.0 Administrator Guide.

7. Axon connects to the Axon Agent using the HTTP protocol.

8. The database where Informatica Data Quality stores the data quality scores is a relational database.
Checklist for Enterprise Data Catalog

Verify the following configurations in Enterprise Data Catalog:

1. Enterprise Data Catalog runs on Informatica version 10.2.2.
2. Axon connects to Enterprise Data Catalog according to the following specifications:
   • You have entered all connection parameters correctly. For more information, refer the Configure Access to Enterprise Data Catalog topic in the Informatica Axon Data Governance 6.0 Administrator Guide.
   • In the Admin Panel > Customize & Configure > Configure Axon > Group > Enterprise Data Catalog > Edit page, select Enable Automatic Onboarding, and enter valid values for the Confidence Score Threshold and Axon Super Admin Email fields.
   • The source database that you want Enterprise Data Catalog to scan is Oracle, SQL Server, or IBM DB2.
   • The file sources for which you want to run rules are flat files in native file systems.
3. Link Enterprise Data Catalog resources to Axon systems. Perform the following steps to create the link:
   a. In Axon, go to Enterprise Catalog > Resources.
   b. Right-click a resource, and select Manage Linking > Create.
   c. Enter the details of the system that you want to create, and click Save.
   The Resources facet displays the linked system for the resource in the Linked Axon Systems column. You can also open the System object and see the linked resource in the Enterprise Catalog tab.
4. Create data sets and attributes for the system that is linked to the Enterprise Data Catalog resource:
   a. In Axon, open the system object, and go to the Enterprise Catalog tab.
   b. Click Edit > Fields to view the fields that Axon has retrieved from Enterprise Data Catalog.
   c. Right-click a field, and select Manage Linking.
   d. In the Link Axon Attributes dialog box, link to an existing attribute, or click Create to create an attribute.
   e. Select glossary terms that you want to associate with the attributes.
   f. Click Save.

Checklist for Axon

Configure Axon according to the following settings:

2. Create a standard data quality rule for the glossary object:
   a. Open a glossary object, click the Edit actions menu, and select New Standard Data Quality Rule.
   b. Enter the details of the standard rule, and set the Axon Status field to Active.
   c. In the Technical Rule Reference field, select a rule specification or rule mapplet from Informatica Data Quality.
   d. Select the Create Local Rules Automatically check box.
   e. Click Save to associate the standard rule in Axon with the rule specification or rule mapplet in Informatica Data Quality.

When you save the standard rule, Axon sends the Enterprise Data Catalog fields to Informatica Data Quality and retrieves the data quality scores. To view the scores, open the standard rule in Axon and go to the Reports tab. The Latest Score column displays the data quality scores.
**Note:** When Informatica Data Quality connects to an Enterprise Data Catalog resource for the first time, you must enter the resource password in Informatica Data Quality for the connection to be successful. Open the Axon Agent log file to determine the Enterprise Data Catalog resource for which you must enter the password. The Axon Agent file is `<Axon_Agent_Home>/logs/dq-axon/dqagent.log`.

The following image shows a sample Axon Agent log file:

![Axon Agent Log File Sample](image)

**Author**

Abhilash Scariya

**Acknowledgements**

Pradeep G.N.