

DQ Scoring Framework - User Manual



Version 2.0

December 30, 2019

**Table of Contents**

[Document Version History 4](#_Toc29900771)

[1.0 Purpose 5](#_Toc29900772)

[2.0 Introduction to DQ Scoring Framework 5](#_Toc29900773)

[3.0 Creating Business Asset Type 7](#_Toc29900774)

**[3.1](#_Toc29900775)** [Data Sources 7](#_Toc29900775)

**[3.2](#_Toc29900776)** [Source Attributes 8](#_Toc29900776)

**[3.3](#_Toc29900777)** [Business Terms 9](#_Toc29900777)

[4.0 Rules Configurations 11](#_Toc29900778)

[5.0 System Reference Lists 23](#_Toc29900779)

[5.1 IGX Data Element Group 23](#_Toc29900780)

[5.2 IGX Scoring Config 24](#_Toc29900781)

[5.3 IGX Rule Operator List 25](#_Toc29900782)

[5.4 IGX Comparison Rule Operator List 27](#_Toc29900783)

[5.5 IGX Pass Skip List 28](#_Toc29900788)

[5.6 Regex Rule List 29](#_Toc29900789)

[5.7 IGX SR Thresholds 29](#_Toc29900790)

[5.8 IGX\_Semantic\_List 31](#_Toc29900791)

[5.9 IGX Data Entity Config 31](#_Toc29900792)

[6.0 Creating Relations 32](#_Toc29900793)

[7.0 Weightage 34](#_Toc29900794)

[8.0 Creating Assets 35](#_Toc29900795)

[9.0 Configurations in DQ+ 35](#_Toc29900796)

[9.1 Create Data Source 35](#_Toc29900797)

[9.2 Create Analysis 35](#_Toc29900798)

[9.3 Create Process Model 37](#_Toc29900799)

[9.4 Dashboard URL Configuration 42](#_Toc29900800)

[10.0 How to Execute – DQ+ Process Executions 43](#_Toc29900801)

[11.0 Dashboards 45](#_Toc29900802)

**[11.1](#_Toc29900803)****[Overall DQ Health](#_Toc29900803)** [45](#_Toc29900803)

**[11.2](#_Toc29900804)****[Score Trend by Asset Group](#_Toc29900804)** [46](#_Toc29900804)

**[11.3](#_Toc29900805)****[Failed and Error Records by Data Entity, Element & Rule](#_Toc29900805)** [47](#_Toc29900805)

**[11.4](#_Toc29900806)****[Pass Rate by Hierarchy](#_Toc29900806)** [48](#_Toc29900806)

**[11.5](#_Toc29900807)****[URL Configuration for DQ+ Dashboard](#_Toc29900807)** [48](#_Toc29900807)

[12.0 Data Quality Results in Data Governance 49](#_Toc29900808)

[12.1 Asset Versioning 50](#_Toc29900809)

[12.2 Hierarchy/Score Rollup 50](#_Toc29900810)

# Document Version History

| Author(s) | Version | Amendments/Reasons | Date |
| --- | --- | --- | --- |
| Aditya SD | 1.0 | Initial Draft | 18-Nov-2019 |
| Naveen Seth | 2.0 | Updates for Release 2.0 | 30-Dec-2019 |

# Purpose

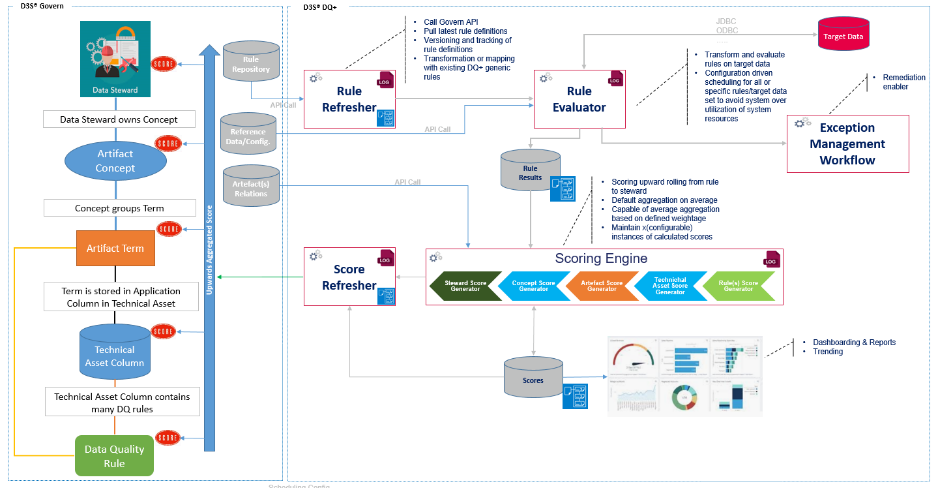
This document describes how to use Infogix solution for data quality scoring, using DQ+ and Data Governance. Intended audience of this document are implementation consultants who wants to implement this solution for specific client.

User of this framework should be well trained on Infogix Data3Sixty® platform including Data Governance and DQ+.

# Introduction to DQ Scoring Framework

DQ Scoring Framework is a solution developed by Infogix. This solution enables to check quality of data based on defined set of data quality rules in Data Governance. As part of the metadata management, various business entities, policies and data quality rules etc. are defined in Data Governance. This framework uses DQ+ to read/fetch metadata and evaluate the quality of actual data, using rules defined in Data Governance. Score of various objects are finally posted back to Data Governance.

Below diagram depicts high level design of this framework.



**Rule Refresher** – Rule Refresher is responsible for fetching Data Governance configurations into DQ+ which will be further used by subsequent processes. . Rule refresher also maintains version of assets and builds/refreshes their hierarchy. It creates version of assets based on their creation, update or changes in their relationships. For every change in the asset, its version and version of assets above in the hierarchy gets increased by 1.

**Rule Evaluator** – Rule Evaluator is responsible to evaluate/validate rules against actual data present in data stores.

**Scoring Engine** – Scoring Engine is responsible for calculating scores of rules, source attributes, data sources, business terms, models and users etc.

**Score Refresher** – Score Refresher is responsible to post score back to Govern at respective level.

**Dashboard Data Preparation** – This component prepares and refreshes data to be used by various dashboards.

# Creating Business Asset Type

Following Business Asset Types need to be created:

* 1. Data Sources

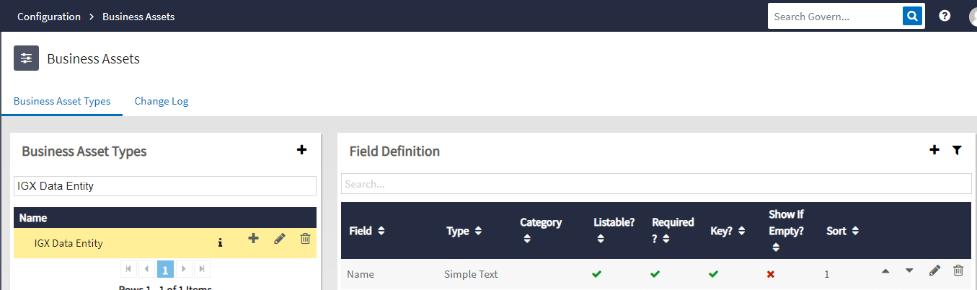
A Business Asset type should be configured to list all the data sources participating in data quality evaluation.

A suitable name can be given for this asset type e.g. “Data Sources”, “Source Files” etc.

Following field(s) must be configured for this asset type:

|  |  |  |
| --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** |
| Name | Name | DQ Results |
| API Name | Name | DQ Results |
| Input Type | Simple Text | Html/Rich text |
| Category | N/A | Data Quality |
| Minimum Value | N/A | N/A |
| Maximum Value | N/A | N/A |
| Type of List | N/A | N/A |
| List Display Format | N/A | N/A |
| Show In Detail Tile | TRUE | TRUE |
| Is Editable | TRUE | FALSE |
| Is Listable | TRUE | FALSE |
| Is Required | TRUE | FALSE |
| Part of Key | TRUE | FALSE |
| Show As Top Level Filter | FALSE | FALSE |
| Allow Multiple Items | FALSE | FALSE |
| Show if Empty | FALSE | FALSE |
| Lookup Type | N/A | N/A |
| Relationship Type | N/A | N/A |
| Table Settings | N/A | N/A |

Refer below screenshot for creating business asset type.



* 1. Source Attributes

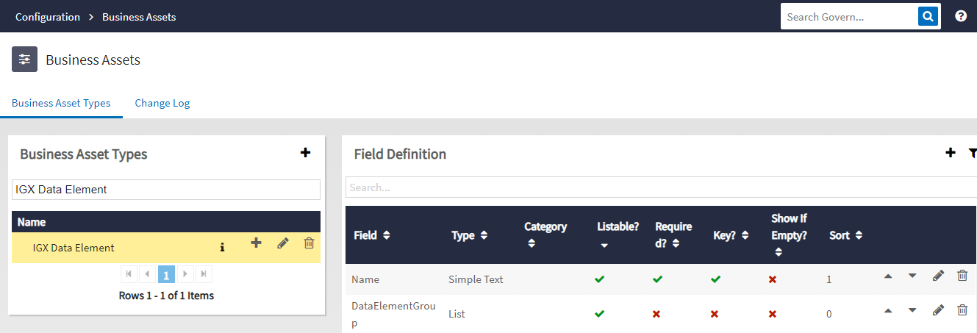
A Business Asset type should be configured to define the consolidated list of attributes of all data sources.

A suitable name can be given for this asset type e.g. “Data Elements”, “Source Fields” etc.

Following field(s) must be configured for this asset type:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** |
| Name | Name | DQ Results | DataElementGroup | Semantic |
| API Name | Name | DQ Results | DataElementGroup | Semantic |
| Input Type | Simple Text | Html/Rich text | List | List |
| Category | N/A | Data Quality | N/A | N/A |
| Minimum Value | N/A | N/A | N/A | N/A |
| Maximum Value | N/A | N/A | N/A | N/A |
| Type of List | N/A | N/A | Reference List Item : IGX Data Element Group | Reference List Item : IGX\_Semantic\_List |
| List Display Format | N/A | N/A | {Group} | {Code} |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | FALSE | TRUE | TRUE |
| Is Listable | TRUE | FALSE | TRUE | TRUE |
| Is Required | TRUE | FALSE | FALSE | FALSE |
| Part of Key | TRUE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | FALSE | FALSE | FALSE | FALSE |
| Allow Multiple Items | FALSE | FALSE | FALSE | FALSE |
| Show if Empty | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A |

Refer below screenshot for creating business asset type.



* 1. Business Terms

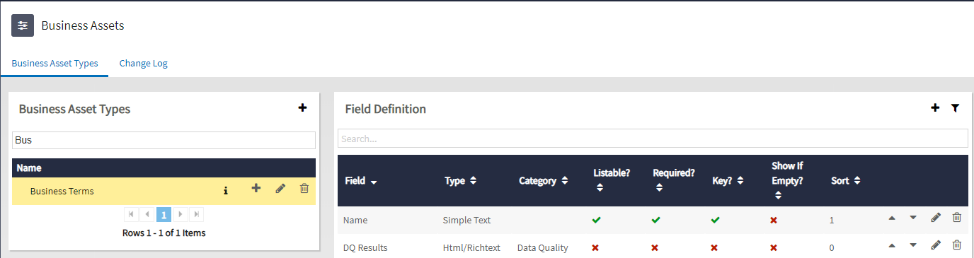
A Business Asset type should be configured to list all the business terms.

A suitable name can be given for this asset type e.g. “Business Term”, “Business Glossary” etc.

Following field(s) must be configured for this asset type:

|  |  |  |
| --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** |
| Name | Name | DQ Results |
| API Name | Name | DQ Results |
| Input Type | Simple Text | Html/Rich text |
| Category | N/A | Data Quality |
| Minimum Value | N/A | N/A |
| Maximum Value | N/A | N/A |
| Type of List | N/A | N/A |
| List Display Format | N/A | N/A |
| Show In Detail Tile | TRUE | TRUE |
| Is Editable | TRUE | FALSE |
| Is Listable | TRUE | FALSE |
| Is Required | TRUE | FALSE |
| Part of Key | TRUE | FALSE |
| Show As Top Level Filter | FALSE | FALSE |
| Allow Multiple Items | FALSE | FALSE |
| Show if Empty | FALSE | FALSE |
| Lookup Type | N/A | N/A |
| Relationship Type | N/A | N/A |
| Table Settings | N/A | N/A |

Refer below screenshot for creating business asset type.



# Rules Configurations

Following rule types are supported in current release of Data Quality Scoring Framework:

1. IGX Duplicate Check
2. IGX Not Blank
3. IGX Reference Lookup
4. IGX Regex
5. IGX JS Expression
6. IGX Numeric Range
7. IGX Attribute Comparison
8. IGX Conformity Check
9. IGX Consistency Check

Above 9 rules need to be configured in Govern as per the templates below:

**IGX Duplicate Check**

It checks for uniqueness of an attribute values within a source for each run.

The below listed fields needs to be configured in Data Governance:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** | **Field-5** | **Field-6** |
| Name | Name | Dimension | Rule Description | Status | Treat Blank or Null As | DQ Results |
| API Name | Name | Dimension | RuleDesciption | Status | TreatBlankorNullAs | DQ Results |
| Input Type | Simple Text | List | Simple Text | List | List | Html/Rich text |
| Category | N/A | NA | NA | NA | NA | N/A |
| Minimum Value | N/A | N/A | NA | NA | NA | N/A |
| Maximum Value | N/A | N/A | NA | NA | NA | N/A |
| Type of List | N/A | Reference List Item : Rule Dimension | N/A | Reference List Item : Rule Status | Reference List Item : IGX Pass Skip List | N/A |
| List Display Format | N/A | {Code} | N/A | {Code} | {Description} | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | TRUE | FALSE |
| Is Listable | TRUE | TRUE | TRUE | FALSE | TRUE | FALSE |
| Is Required | TRUE | TRUE | TRUE | TRUE | FALSE | FALSE |
| Part of Key | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | N/A | NA | N/A | NA | NA | N/A |
| Allow Multiple Items | N/A | FALSE | N/A | FALSE | FALSE | N/A |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A | N/A |

**IGX Not Blank**

It checks if the attribute value is populated or not.

The below listed fields needs to be configured in Data Governance:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** | **Field-5** | **Field-6** |
| Name | Name | Dimension | Rule Description | Status | DQ Results | Consider All Spaces As Blank |
| API Name | Name | Dimension | RuleDesciption | Status | DQ Results | ConsiderAllSpacesAsBlank |
| Input Type | Simple Text | List | Simple Text | List | Html/Rich text | TRUE/FALSE |
| Category | N/A | NA | NA | NA | N/A | N/A |
| Minimum Value | N/A | N/A | NA | NA | N/A | N/A |
| Maximum Value | N/A | N/A | NA | NA | N/A | N/A |
| Type of List | N/A | Reference List Item : Rule Dimension | N/A | Reference List Item : Rule Status | N/A | N/A |
| List Display Format | N/A | {Code} | N/A | {Code} | N/A | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | FALSE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | FALSE | FALSE | TRUE |
| Is Required | TRUE | TRUE | TRUE | TRUE | FALSE | FALSE |
| Part of Key | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | N/A | NA | N/A | NA | N/A | NA |
| Allow Multiple Items | N/A | FALSE | N/A | FALSE | N/A | NA |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A | N/A |

**IGX Reference Lookup**

Looks for presence of attribute value in the respective reference list.

The below listed fields needs to be configured in Data Governance:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** | **Field-5** |
| Name | Name | Dimension | Rule Description | Status | DQ Results |
| API Name | Name | Dimension | RuleDesciption | Status | DQ Results |
| Input Type | Simple Text | List | Simple Text | List | Html/Rich text |
| Category | N/A | NA | NA | NA | N/A |
| Minimum Value | N/A | N/A | NA | NA | N/A |
| Maximum Value | N/A | N/A | NA | NA | N/A |
| Type of List | N/A | Reference List Item : Rule Dimension | N/A | Reference List Item : Rule Status | N/A |
| List Display Format | N/A | {Code} | N/A | {Code} | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | FALSE |
| Is Listable | TRUE | TRUE | TRUE | FALSE | FALSE |
| Is Required | TRUE | TRUE | TRUE | TRUE | FALSE |
| Part of Key | TRUE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | N/A | NA | N/A | NA | N/A |
| Allow Multiple Items | N/A | FALSE | N/A | FALSE | N/A |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-6** | **Field-7** | **Field-8** | **Field-9** |
| Name | Consider All Spaces As Blank | Reference List | Match Reference List | Treat Blank or Null As |
| API Name | ConsiderAllSpacesAsBlank | ReferenceList | MatchReferenceList | TreatBlankorNullAs |
| Input Type | TRUE/FALSE | List | True/False | List |
| Category | N/A | NA | NA | NA |
| Minimum Value | N/A | N/A | NA | NA |
| Maximum Value | N/A | N/A | NA | NA |
| Type of List | N/A | Reference List | NA | Reference List Item : IGX Pass Skip List |
| List Display Format | N/A | {Code} | N/A | {Description} |
| Show In Detail Tile | TRUE | TRUE | TRUE |  |
| Is Editable | TRUE | TRUE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | TRUE |
| Is Required | FALSE | TRUE | FALSE | TRUE |
| Part of Key | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | NA | NA | N/A | FALSE |
| Allow Multiple Items | NA | FALSE | N/A | NA |
| Show if Empty | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | FALSE |
| Relationship Type | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A |

**IGX Regex**

Checks if the attribute value matches with the regular expression or not. These rules uses the “Regex Rules List” reference list to define the checks.

The below listed fields needs to be configured in Data Governance:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** | **Field-5** | **Field-6** | **Field-7** | **Field-8** |
| Name | Name | Dimension | Rule Description | Status | DQ Results | Treat Blank or Null As | Regular Expression | Match Regular Expression |
| API Name | Name | Dimension | RuleDesciption | Status | DQ Results | TreatBlankorNullAs | RegularExpression | MatchRegularExpression |
| Input Type | Simple Text | List | Simple Text | List | Html/Rich text | List | List | TRUE/FALSE |
| Category | N/A | NA | NA | NA | N/A | NA | N/A | N/A |
| Minimum Value | N/A | N/A | NA | NA | N/A | NA | N/A | N/A |
| Maximum Value | N/A | N/A | NA | NA | N/A | NA | N/A | N/A |
| Type of List | N/A | Reference List Item : Rule Dimension | N/A | Reference List Item : Rule Status | N/A | Reference List Item : IGX Pass Skip List | Reference List Item : Regex Rules List | N/A |
| List Display Format | N/A | {Code} | N/A | {Code} | N/A | {Description} | {Code} | N/A |
| List Edit Format | N/A | N/A | N/A | N/A | N/A |  | {Description} | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | FALSE | TRUE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | FALSE | FALSE | TRUE | TRUE | TRUE |
| Is Required | TRUE | TRUE | TRUE | TRUE | FALSE | FALSE | TRUE | FALSE |
| Part of Key | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | N/A | NA | N/A | NA | N/A | NA | N/A | NA |
| Allow Multiple Items | N/A | FALSE | N/A | FALSE | N/A | FALSE | N/A | NA |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |

**IGX JS Expression**

Checks if the attribute value matches with the JS expression evaluation result.

The below listed fields needs to be configured in Data Governance:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** | **Field-5** | **Field-6** | **Field-7** |
| Name | Name | Dimension | Rule Description | Status | DQ Results | JS Expression | Match JS Expression |
| API Name | Name | Dimension | RuleDesciption | Status | DQ Results | JSExpression | MatchJSExpression |
| Input Type | Simple Text | List | Simple Text | List | Html/Rich text | Simple Text | TRUE/FALSE |
| Category | N/A | NA | NA | NA | N/A | N/A | N/A |
| Minimum Value | N/A | N/A | NA | NA | N/A | N/A | N/A |
| Maximum Value | N/A | N/A | NA | NA | N/A | N/A | N/A |
| Type of List | N/A | Reference List Item : Rule Dimension | N/A | Reference List Item : Rule Status | N/A | N/A | N/A |
| List Display Format | N/A | {Code} | N/A | {Code} | N/A | N/A | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | FALSE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | FALSE | FALSE | TRUE | TRUE |
| Is Required | TRUE | TRUE | TRUE | TRUE | FALSE | TRUE | FALSE |
| Part of Key | TRUE | FALSE | FALSE | FALSE | FALSE | TRUE | FALSE |
| Show As Top Level Filter | N/A | NA | N/A | NA | N/A | N/A | NA |
| Allow Multiple Items | N/A | FALSE | N/A | FALSE | N/A | N/A | NA |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A | N/A | N/A |

**IGX Numeric Range**

Checks if attribute value lies within the specified numeric range.

The below listed fields needs to be configured in Data Governance (refer below 2 tables):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** | **Field-5** | **Field-6** |
| Name | Name | Dimension | Rule Description | Status | DQ Results | Treat Blank or Null As |
| API Name | Name | Dimension | RuleDesciption | Status | DQ Results | TreatBlankorNullAs |
| Input Type | Simple Text | List | Simple Text | List | Html/Rich text | List |
| Category | N/A | NA | NA | NA | N/A | NA |
| Minimum Value | N/A | N/A | NA | NA | N/A | NA |
| Maximum Value | N/A | N/A | NA | NA | N/A | NA |
| Type of List | N/A | Reference List Item : Rule Dimension | N/A | Reference List Item : Rule Status | N/A | Reference List Item : IGX Pass Skip List |
| List Display Format | N/A | {Code} | N/A | {Code} | N/A | {Description} |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | FALSE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | FALSE | FALSE | TRUE |
| Is Required | TRUE | TRUE | TRUE | TRUE | FALSE | FALSE |
| Part of Key | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | N/A | NA | N/A | NA | N/A | NA |
| Allow Multiple Items | N/A | FALSE | N/A | FALSE | N/A | FALSE |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A | N/A |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-7** | **Field-8** | **Field-9** | **Field-10** | **Field-11** |
| Name | Lower Bound | Lower Bound Inclusive | Upper Bound | Upper Bound Inclusive | Within Range |
| API Name | LowerBound | LowerBoundInclusive | UpperBound | UpperBoundInclusive | WithinRange |
| Input Type | Decimal Number | TRUE/FALSE | Decimal Number | TRUE/FALSE | TRUE/FALSE |
| Category | N/A | N/A | N/A | N/A | N/A |
| Minimum Value | N/A | N/A | N/A | N/A | N/A |
| Maximum Value | N/A | N/A | N/A | N/A | N/A |
| Type of List | N/A | N/A | N/A | N/A | N/A |
| List Display Format | N/A | N/A | N/A | N/A | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Required | TRUE | FALSE | TRUE | FALSE | FALSE |
| Part of Key | FALSE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | NA | NA | NA | NA | NA |
| Allow Multiple Items | NA | NA | NA | NA | NA |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A |

**IGX Attribute Comparison:**

Compares the attribute value with the comparison attribute value.

The below listed fields needs to be configured in Data Governance (Refer the 2 tables below):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** | **Field-5** |
| Name | Name | Dimension | Rule Description | Status | DQ Results |
| API Name | Name | Dimension | RuleDesciption | Status | DQ Results |
| Input Type | Simple Text | List | Simple Text | List | Html/Rich text |
| Category | N/A | NA | NA | NA | N/A |
| Minimum Value | N/A | N/A | NA | NA | N/A |
| Maximum Value | N/A | N/A | NA | NA | N/A |
| Type of List | N/A | Reference List Item : Rule Dimension | N/A | Reference List Item : Rule Status | N/A |
| List Display Format | N/A | {Code} | N/A | {Code} | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | FALSE |
| Is Listable | TRUE | TRUE | TRUE | FALSE | FALSE |
| Is Required | TRUE | TRUE | TRUE | TRUE | FALSE |
| Part of Key | TRUE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | N/A | NA | N/A | NA | N/A |
| Allow Multiple Items | N/A | FALSE | N/A | FALSE | N/A |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-6** | **Field-7** | **Field-8** | **Field-9** |
| Name | Treat Blank or Null As | Rule Operator | Rule Value | Rule Value Case Sensitive |
| API Name | TreatBlankorNullAs | RuleOperator | RuleValue | RuleValueCaseSensitive |
| Input Type | List | List | List | TRUE/FALSE |
| Category | NA | NA | NA | N/A |
| Minimum Value | NA | NA | NA | N/A |
| Maximum Value | NA | NA | NA | N/A |
| Type of List | Reference List Item : IGX Pass Skip List | Reference List Item : IGX Comparison Rule Operator List | Reference List Item : IGX Data Element | N/A |
| List Display Format | {Description} | {Description} | {Name} | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | TRUE |
| Is Required | FALSE | TRUE | TRUE | FALSE |
| Part of Key | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | NA | NA | NA | NA |
| Allow Multiple Items | FALSE | FALSE | FALSE | NA |
| Show if Empty | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A |

**IGX Conformity Check**

Compares the attribute value with the respective string or numeric value.

The below listed fields needs to be configured in Data Governance (Refer 2 tables below):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** | **Field-5** |
| Name | Name | Dimension | Rule Description | Status | DQ Results |
| API Name | Name | Dimension | RuleDesciption | Status | DQ Results |
| Input Type | Simple Text | List | Simple Text | List | Html/Rich text |
| Category | N/A | NA | NA | NA | N/A |
| Minimum Value | N/A | N/A | NA | NA | N/A |
| Maximum Value | N/A | N/A | NA | NA | N/A |
| Type of List | N/A | Reference List Item : Rule Dimension | N/A | Reference List Item : Rule Status | N/A |
| List Display Format | N/A | {Code} | N/A | {Code} | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | FALSE |
| Is Listable | TRUE | TRUE | TRUE | FALSE | FALSE |
| Is Required | TRUE | TRUE | TRUE | TRUE | FALSE |
| Part of Key | TRUE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | N/A | NA | N/A | NA | N/A |
| Allow Multiple Items | N/A | FALSE | N/A | FALSE | N/A |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-6** | **Field-7** | **Field-8** | **Field-9** |
| Name | Treat Blank or Null As | Rule Operator | Rule Value | Rule Value Case Sensitive |
| API Name | TreatBlankorNullAs | RuleOperator | RuleValue | RuleValueCaseSensitive |
| Input Type | List | List | List | TRUE/FALSE |
| Category | NA | NA | NA | N/A |
| Minimum Value | NA | NA | NA | N/A |
| Maximum Value | NA | NA | NA | N/A |
| Type of List | Reference List Item : IGX Pass Skip List | Reference List Item : IGX Comparison Rule Operator List | Reference List Item : IGX Data Element | N/A |
| List Display Format | {Description} | {Description} | {Name} | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | TRUE |
| Is Required | FALSE | TRUE | TRUE | FALSE |
| Part of Key | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | NA | NA | NA | NA |
| Allow Multiple Items | FALSE | FALSE | FALSE | NA |
| Show if Empty | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A |

**IGX Consistency Check**

Compares attribute value with the comparison value if pre-condition is met.

The below listed fields needs to be configured in Data Governance (Refer 2 tables below):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** | **Field-5** | **Field-6** | **Field-7** |
| Name | Name | Dimension | Rule Description | Status | DQ Results | Treat Blank or Null As | Pre-Condition Attribute |
| API Name | Name | Dimension | RuleDesciption | Status | DQ Results | TreatBlankorNullAs | PreConditionAttribute |
| Input Type | Simple Text | List | Simple Text | List | Html/Rich text | List | List |
| Category | N/A | NA | NA | NA | N/A | NA | NA |
| Minimum Value | N/A | N/A | NA | NA | N/A | NA | NA |
| Maximum Value | N/A | N/A | NA | NA | N/A | NA | NA |
| Type of List | N/A | Reference List Item : Rule Dimension | N/A | Reference List Item : Rule Status | N/A | Reference List Item : IGX Pass Skip List | Reference List Item : IGX Data Element |
| List Display Format | N/A | {Code} | N/A | {Code} | N/A | {Description} | {Name} |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | FALSE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | FALSE | FALSE | TRUE | TRUE |
| Is Required | TRUE | TRUE | TRUE | TRUE | FALSE | FALSE | TRUE |
| Part of Key | TRUE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | N/A | NA | N/A | NA | N/A | NA | NA |
| Allow Multiple Items | N/A | FALSE | N/A | FALSE | N/A | FALSE | FALSE |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A | N/A | N/A |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-9** | **Field-10** | **Field-11** | **Field-12** | **Field-13** |
| Name | Pre-Condition Value | Pre-Condition Value Case Sensitive | Rule Operator | Rule Value | Rule Value Case Sensitive |
| API Name | PreConditionValue | PreConditionValueCaseSensitive | RuleOperator | RuleValue | RuleValueCaseSensitive |
| Input Type | Simple Text | TRUE/FALSE | List | List | TRUE/FALSE |
| Category | N/A | N/A | NA | NA | N/A |
| Minimum Value | N/A | N/A | NA | NA | N/A |
| Maximum Value | N/A | N/A | NA | NA | N/A |
| Type of List | N/A | N/A | Reference List Item : IGX Comparison Rule Operator List | Reference List Item : IGX Data Element | N/A |
| List Display Format | N/A | N/A | {Description} | {Name} | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | TRUE | TRUE |
| Is Required | TRUE | FALSE | TRUE | TRUE | FALSE |
| Part of Key | FALSE | FALSE | FALSE | FALSE | FALSE |
| Show As Top Level Filter | N/A | NA | NA | NA | NA |
| Allow Multiple Items | N/A | NA | FALSE | FALSE | NA |
| Show if Empty | FALSE | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A | N/A |

# System Reference Lists

Create below reference lists in Data Governance with the given names. These reference lists are used by various rules and scoring configurations.

## IGX Data Element Group

Sometimes data information is divided among multiple source attributes e.g. address may be represented as address line-1, address line-2, city, state and zip. An address is considered valid only if all these fields are valid. Such type of validation checks can be achieved by defining a group for such attributes and the type of validation required on the group.

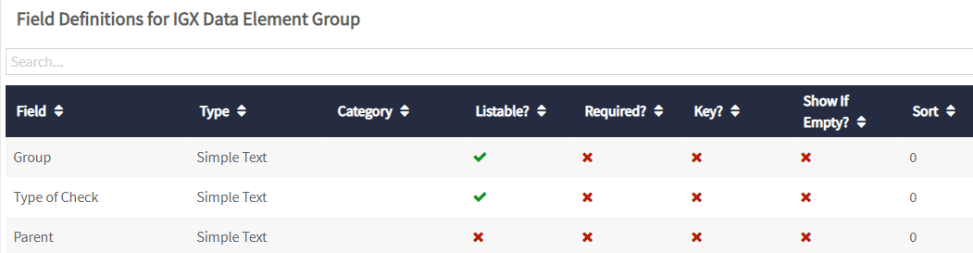
Following two type of group checks are supported:

1. AtLeastOneMustPass – Any one of source attribute in the group has valid values as per rules defined on them then consider entire group as valid. This scenario is relevant for Phone, Email, IdProof etc.
2. AllMustPass - Every source attributes of the group should have valid values as per defined rules on them then only consider entire group as valid else consider all the attributes of the group as invalid. This scenario is relevant for Address etc.

Below fields need to be configured for IGX Data Element Group:

|  |  |  |  |
| --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** |
| Name | Group | Type of Check | Parent |
| API Name | Group | TypeofCheck | Parent |
| Input Type | Simple Text | Simple Text | Simple Text |
| Category | N/A | N/A | N/A |
| Minimum Value | N/A | N/A | N/A |
| Maximum Value | N/A | N/A | N/A |
| Type of List | N/A | N/A | N/A |
| List Display Format | N/A | N/A | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | FALSE |
| Is Required | TRUE | TRUE | FALSE |
| Part of Key | N/A | N/A | N/A |
| Show As Top Level Filter | N/A | N/A | N/A |
| Allow Multiple Items | N/A | N/A | N/A |
| Show if Empty | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A |

Below is screenshot from Data Governance



## IGX Scoring Config

This reference list contains system configuration parameters for the solution.

Configure the below field in govern as per the template below:

|  |  |
| --- | --- |
| **Field Attribute** | **Field-1** |
| Name | Description |
| API Name | Description |
| Input Type | Simple Text |
| Category | N/A |
| Minimum Value | N/A |
| Maximum Value | N/A |
| Type of List | N/A |
| List Display Format | N/A |
| Show In Detail Tile | TRUE |
| Is Editable | TRUE |
| Is Listable | TRUE |
| Is Required | TRUE |
| Part of Key | N/A |
| Show As Top Level Filter | N/A |
| Allow Multiple Items | N/A |
| Show if Empty | FALSE |
| Lookup Type | N/A |
| Relationship Type | N/A |
| Table Settings | N/A |

Following entries need to be created in the reference list. For the value in description column refer remarks.

|  |  |  |
| --- | --- | --- |
| **Code** | **Description** | **Remark** |
| RE\_RULE\_NOT\_BLANK | IGX Not Blank | Put the Description value as defined here in the description column. |
| RE\_RULE\_REGEX | IGX Regex |
| RE\_RULE\_NUMERIC\_RANGE | IGX Numeric Range |
| RE\_RULE\_DUPLICATE\_CHECK | IGX Duplicate Check |
| RE\_RULE\_REFERENCE\_LOOKUP | IGX Reference Lookup |
| RE\_RULE\_CONSISTENCY | IGX Consistency Check |
| RE\_RULE\_JSEXP | IGX JS Expression |
| RE\_ATTRIBUTE\_COMPARISON | IGX Attribute Comparison |
| RE\_CONFORMITY\_CHECKS | IGX Conformity Check |
| AT\_DATA\_ENTITY | IGX Data Entity | Put the value of description as name of business asset type, configured in [Data Sources Section.](#DataSources) |
| AT\_DATA\_ELEMENT | IGX Data Element | Put the value of description as name of business asset type, configured in [Source Attributes Section](#SourceAttributes). |
| AT\_BUSINESS\_TERM | Name of the Asset Type defined for Business Term e.g. IGX Business Term | Put the value of description as name of business asset type, configured in section [business terms](#BusinessTerms). |
| PE\_DATA\_ENTITY\_TO\_DATA\_ELEMENT\_PREDICATE\_UID |  | Put the UID value of predicate used to relate data sources with source attributes. |
| PE\_DATA\_ELEMENT\_TO\_RULE\_PREDICATE\_UID |  | Put the UID value of predicate used to relate the source attribute with rule. |
| PR\_RULE\_TO\_LOWEST\_LEVEL\_PREDICATE |  | Put the name of predicate used to relate the source attribute with rule. |
| PR\_HIERARCHY\_PARTICIPANT\_PREDICATES |  | Put the comma separated list of predicate names which needs to be used to build the hierarchy. |
| SR\_SCORE\_POST\_DAYS |  | No. of days/executions of scores to be posted to Govern |

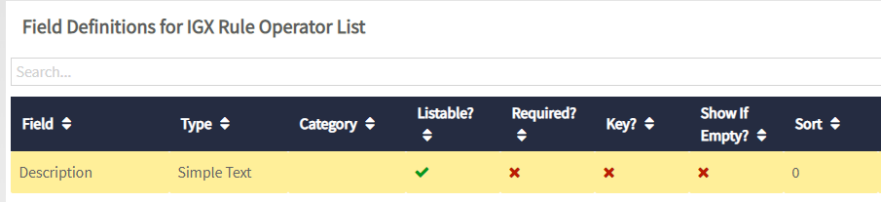
## IGX Rule Operator List

This reference list contains list of comparison operators to be used in rule definition.

Configure the below field in govern as per the template below:

|  |  |
| --- | --- |
| **Field Attributes** | **Field-1** |
| Name | Description |
| API Name | Description |
| Input Type | Simple Text |
| Category | N/A |
| Minimum Value | N/A |
| Maximum Value | N/A |
| Type of List | N/A |
| List Display Format | N/A |
| Show In Detail Tile | TRUE |
| Is Editable | TRUE |
| Is Listable | TRUE |
| Is Required | TRUE |
| Part of Key | N/A |
| Show As Top Level Filter | N/A |
| Allow Multiple Items | N/A |
| Show if Empty | FALSE |
| Lookup Type | N/A |
| Relationship Type | N/A |
| Table Settings | N/A |

Below is screenshot from Data Governance



Following entries need to be created in the reference list with the given values.

|  |  |
| --- | --- |
| **Code** | **Description** |
| Contains | Contains |
| Does Not Contain | Does Not Contain |
| Does Not Match Regular Expression | Does Not Match Regular Expression |
| Ends With | Ends With |
| Is Equal To | Is Equal To |
| Is Greater Than | Is Greater Than |
| Is Greater Than or Equals To | Is Greater Than or Equals To |
| Is In Delimited List | Is In Delimited List |
| Is Less Than | Is Less Than |
| Is Less Than or Equal To | Is Less Than or Equal To |
| Is Not Equal To | Is Not Equal To |
| Is Not In Delimited List | Is Not In Delimited List |
| Is Not Null/Blank | Is Not Null/Blank |
| Is Null/Blank | Is Null/Blank |
| Matches Regular Expression | Matches Regular Expression |
| Starts With | Starts With |

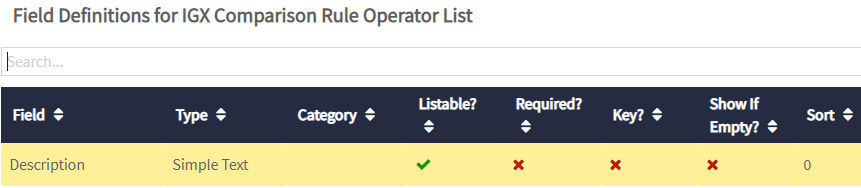
## IGX Comparison Rule Operator List

This reference list contains list of comparison operators (between two attributes) to be used in rules definition.

Configure the below field in Data Governance as per the template below:

|  |  |
| --- | --- |
| **Field Attribute** | **Field-1** |
| Name | Description |
| API Name | Description |
| Input Type | Simple Text |
| Category | N/A |
| Minimum Value | N/A |
| Maximum Value | N/A |
| Type of List | N/A |
| List Display Format | N/A |
| Show In Detail Tile | TRUE |
| Is Editable | TRUE |
| Is Listable | TRUE |
| Is Required | TRUE |
| Part of Key | N/A |
| Show As Top Level Filter | N/A |
| Allow Multiple Items | N/A |
| Show if Empty | FALSE |
| Lookup Type | N/A |
| Relationship Type | N/A |
| Table Settings | N/A |

Below is the screen shot from Data Governance



Following entries need to be created in the reference list with the given values.

|  |  |
| --- | --- |
| **Code** | **Description** |
| Is Equal To | Is Equal To |
| Is Greater Than | Is Greater Than |
| Is Greater Than or Equals To | Is Greater Than or Equals To |
| Is Less Than | Is Less Than |
| Is Less Than or Equal To | Is Less Than or Equal To |
| Is Not Equal To | Is Not Equal To |

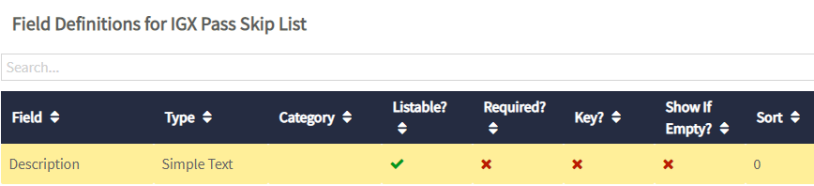


## IGX Pass Skip List

This reference contains list of available option for treating the Null/Blank Value of an attribute. Configure the below field in govern as per the template below:

|  |  |
| --- | --- |
| **Field Attribute** | **Field-1** |
| Name | Description |
| API Name | Description |
| Input Type | Simple Text |
| Category | N/A |
| Minimum Value | N/A |
| Maximum Value | N/A |
| Type of List | N/A |
| List Display Format | N/A |
| Show In Detail Tile | TRUE |
| Is Editable | TRUE |
| Is Listable | TRUE |
| Is Required | TRUE |
| Part of Key | N/A |
| Show As Top Level Filter | N/A |
| Allow Multiple Items | N/A |
| Show if Empty | FALSE |
| Lookup Type | N/A |
| Relationship Type | N/A |
| Table Settings | N/A |

Below is the screenshot from Data Governance



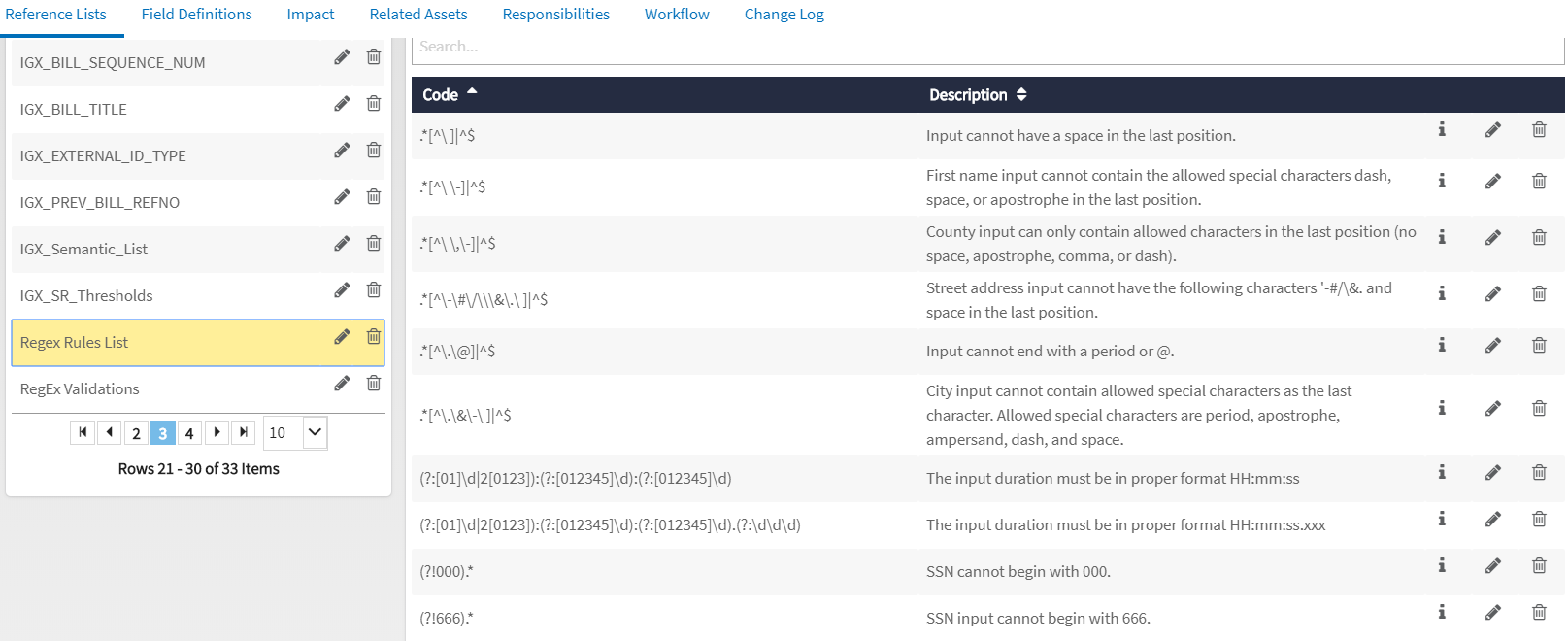
Following entries need to be created in the reference list with the given values.



## Regex Rule List

This reference list contains Pre Defined list of regex checks which can be re-used to define the Regex based rules. Any new regex check needs to be added to this list before using it in Regex rule.

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| CODE | Regular Expression for the Rule |
| DESCRIPTION | Short Description of rule |



The Regex Rule List is embedded below



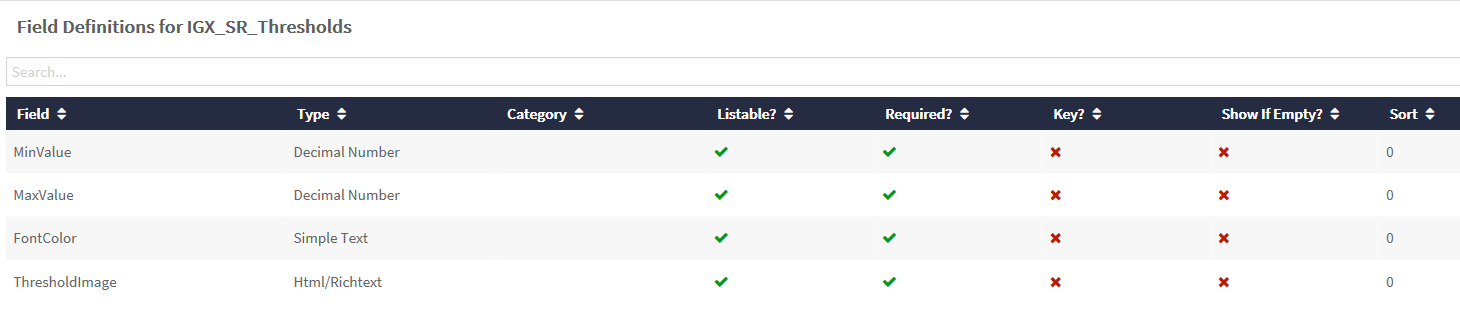
## IGX SR Thresholds

This reference list contains score slabs and their respective Font Color and Images to highlight the scores if they fall within the range. These slabs are used while publishing the DQ Score in Govern by the score refresher module.

Configure the below field in Data Governance as per the template below:

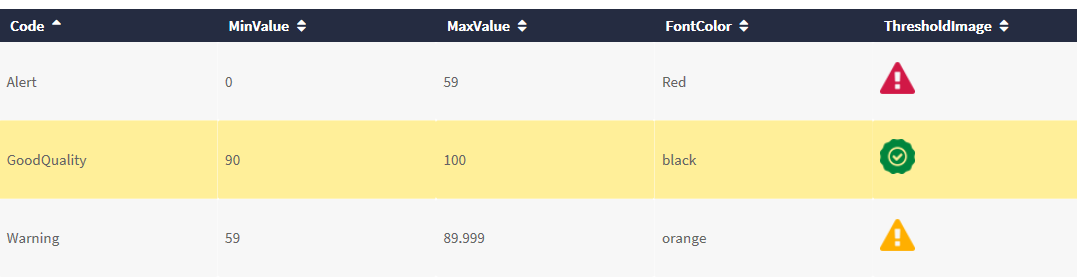
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field Attribute** | **Field-1** | **Field-2** | **Field-3** | **Field-4** |
| Name | MinValue | MaxValue | FontColor | ThresholdImage |
| API Name | MinValue | MaxValue | FontColor | ThresholdImage |
| Input Type | Decimal Number | Decimal Number | Simple Text | Html/Rich text |
| Category | N/A | N/A | N/A | N/A |
| Minimum Value | N/A | N/A | N/A | N/A |
| Maximum Value | N/A | N/A | N/A | N/A |
| Type of List | N/A | N/A | N/A | N/A |
| List Display Format | N/A | N/A | N/A | N/A |
| Show In Detail Tile | TRUE | TRUE | TRUE | TRUE |
| Is Editable | TRUE | TRUE | TRUE | TRUE |
| Is Listable | TRUE | TRUE | TRUE | TRUE |
| Is Required | TRUE | TRUE | TRUE | TRUE |
| Part of Key | N/A | N/A | N/A | N/A |
| Show As Top Level Filter | N/A | N/A | N/A | N/A |
| Allow Multiple Items | N/A | N/A | N/A | N/A |
| Show if Empty | FALSE | FALSE | FALSE | FALSE |
| Lookup Type | N/A | N/A | N/A | N/A |
| Relationship Type | N/A | N/A | N/A | N/A |
| Table Settings | N/A | N/A | N/A | N/A |

Below is the screen shot from Data Governance



Note :- In ThresholeImage Field user should embed the image to be displayed if the score falls within the given range. No other characters/contents should be added to this field.

Following are example entries that can be created in the reference list with the given values.



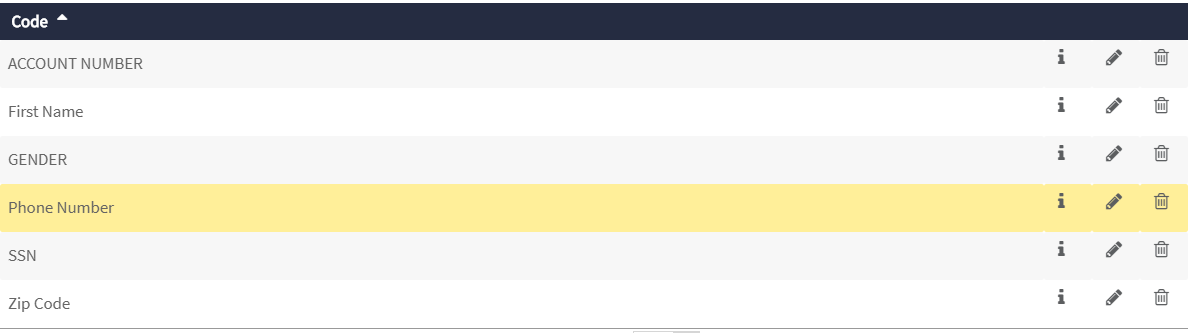
The above Threshold Images used are attached below

## IGX\_Semantic\_List

This reference list contains various types of Semantics which can be defined for different Source Column to elaborate their nature of content. These semantics can be further linked (Via relationship) to different DQ Rules. These rules are automatically applied to the Source Columns which had defined semantic.

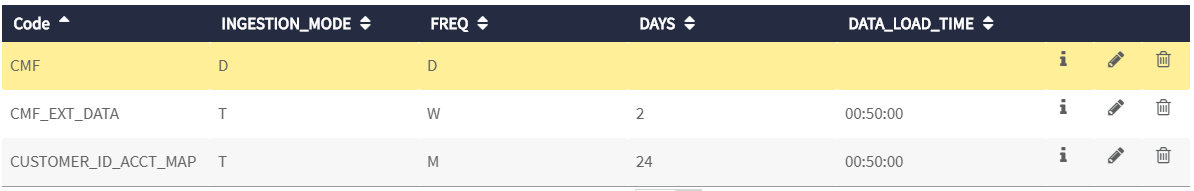
|  |  |
| --- | --- |
| **Field Name** | **Description** |
| CODE | Regular Expression for the Rule |



## IGX Data Entity Config

This reference list will contain the configuration of source or Data Entity on which DQ rules will be applied. Reference list will have following structure

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| CODE | Exact Name Data Source as defined in [Data Sources](#DataSources) |
| INGESTION\_MODE | Based on nature of sources, data can be ingested into system to perform data quality check. Following are various ingestion modes supported  D – Direct. This mode works for the data sources which already exist on DQ+ and their data keeps updating directly in the source.  L – Manually loaded into DQ+ through an adhoc process and their logs are populated accordingly to indicate the new load.  T – It is used for Data Sources which lands in to DQ+ system on a predefined time. No log of such data load is available |
| FREQ | D – Daily  W – Weekly  M – Monthly |
| DAYS | It is the days on which file is expected e.g.  If Freq is ‘W’  1,3,5 – represents Monday, Wednesday and Friday  If Freq is ‘M  10,20 – 10th and 20th of the month |
| DATA\_LOAD\_TIME | Time by which data is assumed to have got loaded  Format: HH24:MI |



# Creating Relations

User needs to setup relationship between following:

* 1. [Data Source](#DataSources) and [Source Attributes](#SourceAttributes)
  2. [Source Attributes](#SourceAttributes) and [Rules](#RulesConfigurations)
  3. IGX\_Semantic\_List and Rules
  4. [Source Attributes](#SourceAttributes) and [Business Terms](#BusinessTerms)
  5. [Business Terms](#BusinessTerms) and Models

While creating a relationship type between Business Terms and Model, below custom field needs to be configured:

|  |  |
| --- | --- |
| **Field Attribute** | **Field-1** |
| Name | Weightage |
| API Name | Weightage |
| Input Type | Decimal Number |
| Category | N/A |
| Minimum Value | 0 |
| Maximum Value | 1 |
| Type of List | N/A |
| List Display Format | N/A |
| Show In Detail Tile | TRUE |
| Is Editable | TRUE |
| Is Listable | TRUE |
| Is Required | FALSE |
| Part of Key | FALSE |
| Show As Top Level Filter | FALSE |
| Allow Multiple Items | FALSE |
| Show if Empty | FALSE |
| Lookup Type | N/A |
| Relationship Type | N/A |
| Table Settings | N/A |

**Note**:

1. It is recommended to use different predicates for above relations.

2. Owners should be assigned for the Models. It is recommended not to assign owners at any other assets.

3. While establishing the relationship, weightage can be assigned. Refer [section#7.0](#Weightage) for more details.

4. Each Business Term should at least be related to one source attributes.

# Weightage

When creating a relation between Business Term and Model, user can assign weightage to each business term. This weightage, if defined, will be used in computing the score of the Model.

**Note**: Sum of weightage for all business terms related to a model should always be 1. If configured, Weightage should be assigned to all the Business Terms related to a Model.

# Creating Assets

Assets need to be created for following asset types along with their relationships.

* [Data Sources](#DataSources)
* [Source Attributes](#SourceAttributes) – Naming convention of source attributes is <DataSource>.<SourceAttribute>
* [Business Terms](#BusinessTerms)
* [Rules](#BusinessTerms)
* Model
* Responsibilities
* User
* Group

**Note**:

1. Assets of any other types may be configured as per project requirements.
2. Currently there is a limitation of DQ+ due to which if any record in Java Script takes more than 60 secs to process then it throws error in analysis execution. Due to this limitation processes related to building Hierarchy , Asset Versioning & Scoring may throw an error if the number of source attributes and business terms are huge in nature and may require some performance optimization. (Zendesk Ticket #80783 opened for same).

# Configurations in DQ+

Configure following environment variables in DQ+ to point Data Governance. Please ensure connectivity between Data Governance and DQ+ before configuring.

**governURL** – Data Governance application URL (for e.g. <https://infogix-fda.dev.data3sixty.com>)

**governAuth** – API authorization key to call Data Governance APIs using a service account. To get this key reach out to administrator/infrastructure team.

For each input [data source](#DataSources), following steps need to be performed:

## Create Data Source

For each data source on which data quality is to be evaluated, create DS and configure it as per the specific requirement of the project.

## Create Analysis

Purpose of this analysis is to identify the data and transform into standard layout, so that Rule Evaluator can execute the DQ rules on it. Such separate analysis to be developed for each data source for which DQ is to be ascertained.

There may be different scenarios of data identification and example of some of them are listed below:

1. Whole data in the dataset to be evaluated for DQ
2. Only new data in the dataset to be evaluated for DQ
   1. New data loaded in data source after last run
   2. New data based on some log
   3. New data based on insert/update date available in data source

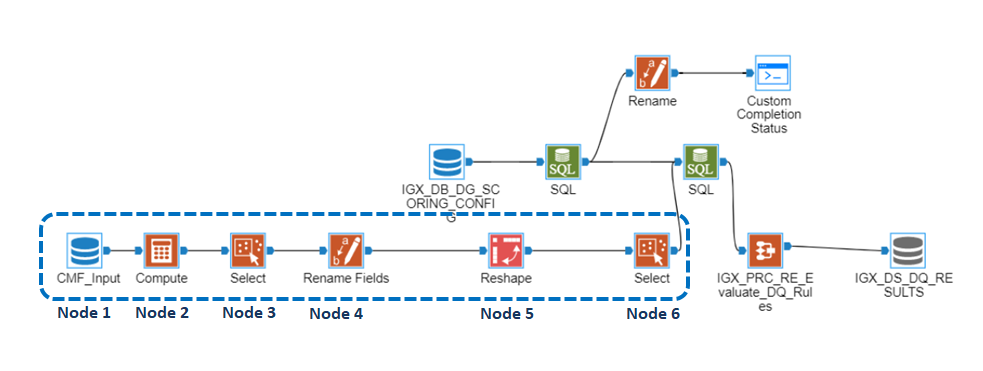
Based on the scenarios the analysis needs to be customized as below-

**IGX\_PRC\_EVALUATE\_DQ \_< DATA Entity>:** This analysis reads the input data and prepares the transformed key value pair for each record & column in the input. And then it calls the **IGX\_PRC\_DQ\_RULES\_EXECUTOR,** using Nested Analysis node.

User can create a copy of existing analysis –

“DQ Govern Scoring\ Sample\Analysis\IGX\_PRC\_ EVALAUTE\_DQ\_CMF”

And make the required changes in it. Changes are applicable only to the 6 nodes highlighted in below screen shot.



Custom configurations to be done in above highlighted nodes –

Node 1: Data Input Node - This node reads in the data from source input. The user needs to configure here the data load range as per the applicable scenario.

In the same node add a new field by name “**Data Key**” and assign the unique identifier column(s) available in input data.

Node 2: Compute Node - Converts those attributes to String data type, on which data quality is to be evaluated. Assigns these String values to new attributes with original names suffixed with string "\_1". This node is required to convert non string data types to string data types.

Node 3: Select Node - Select the required fields (Retain the field with names that are suffixed with "\_1" and ignore their corresponding original fields)

Node 4: Rename Node - Rename all the attributes that are selected in previous node (Remove the suffix "\_1" from attributes name)

Node 5: Reshape Node - Convert all input columns, on which data quality is to be evaluated, to rows and create key value pair. Select option “Convert Columns to Rows” in drop down – “Reshape Operation to perform”. Add all the fields except Data Key in section – “Fields to Move to Rows”. Put the field name as FieldName and values as FieldValue.

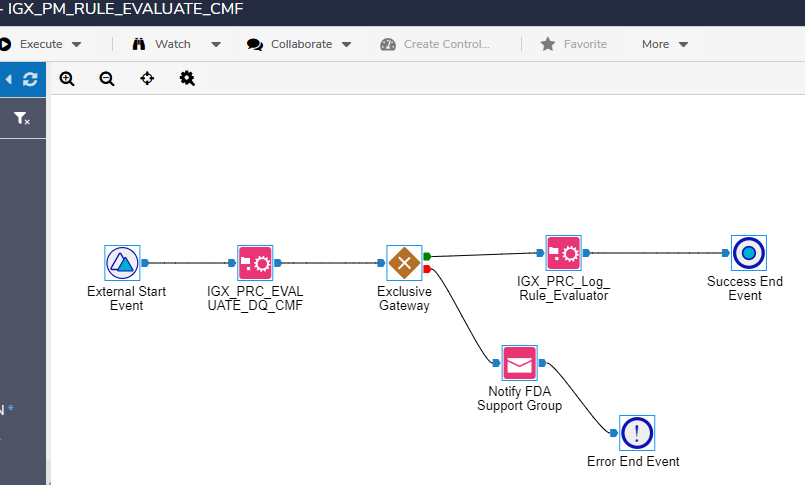
Node 6: Select Node - Select the Data Key, Field Name and Field Value and add fields: **Batch\_Identifier** Assign the unique batch identifier to the column, if available in data set, else assign SUBSTRING( CURRENT\_TIMESTAMP(),0,10 ).

**PM\_DATA\_ENTITY** as RUNTIME('PM\_DATA\_ENTITY').

**Run\_Date** as RUNTIME('PM\_START\_TS')

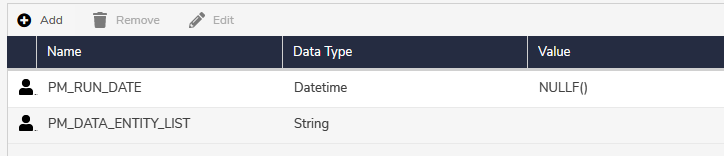
## Create Process Model

1. Create Process Models IGX\_PM\_RULE\_EVALUATE\_<DATA SOURCE> each for executing the analysis defined in [Create Analysis section above](#_Create_Analysis) for each Data Source in the system.



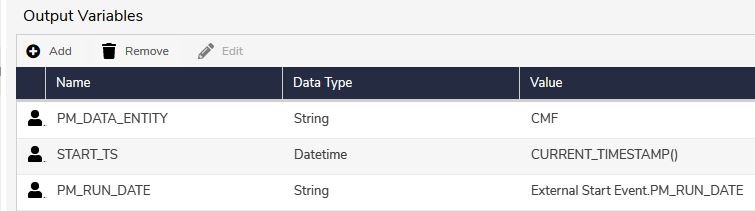
Step 1 :- External Start Event :-

Below Variables to be set in External Start Event.



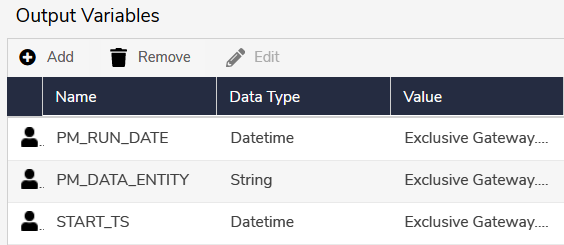
PM\_DATA\_ENTITY\_LIST to be set as Data Type String and as an External Variable

Step 2 :- IGX\_PRC\_EVALUATE\_DQ\_CMF - This will execute the analysis defined for Data Source in [Create Analysis](#_Toc24980323) section . Define parameter PM\_RUN\_DATE as in External Start Event (Step 1) and PM\_DATA\_ENTITY as “CMF” and START\_TS as CURRENT\_TIMESTAMP.



Step 3:- Check for PM\_RUN\_DATE in exclusive Gateway – If PM\_RUN\_DATE returned by analysis “IGX\_PRC\_EVALUATE\_DQ\_CMF” as not null then proceed further otherwise send notification to FDA Support Group.

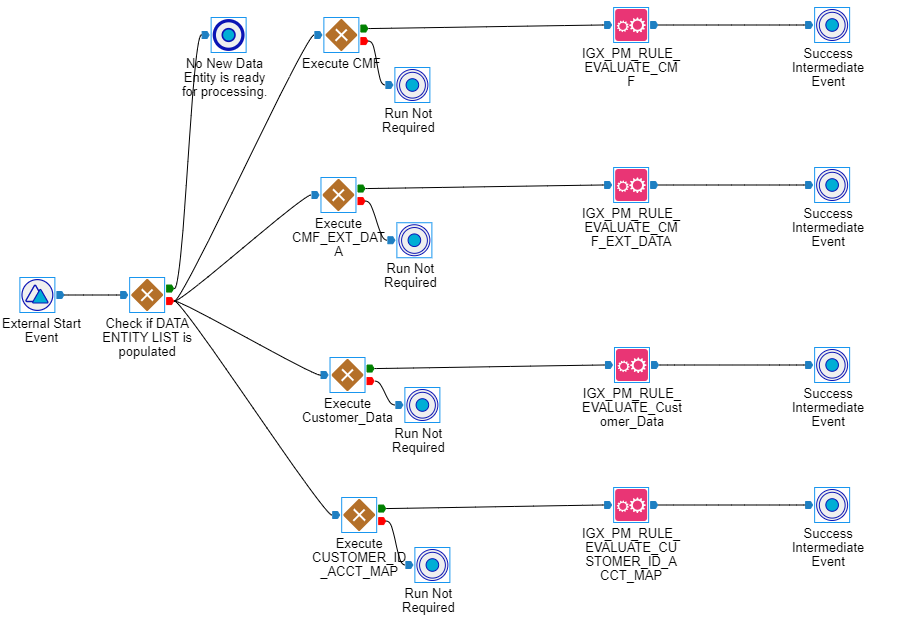
Step 4:- IGX\_PRC\_Log\_Rule\_Evaluator – This will execute the analysis for updating the Evaluator Log for Data Source “CMF”. Carry the parameters defined in Step 3.



Step 5:- Success END Event – It represents end of successfully execution of the Process Model.

Repeat Steps 1 to Steps 5 for each Data Source defined in the system.

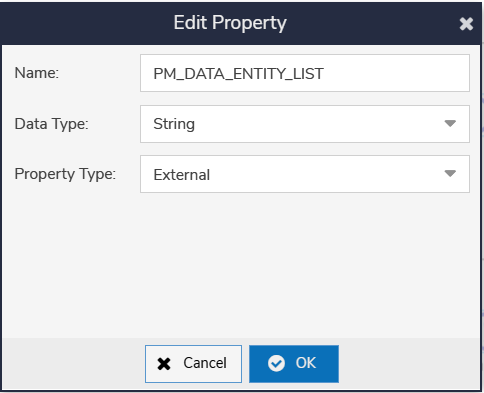
1. Create Process Model IGX\_PM\_RULE\_EVALUATOR for executing the Process Models defined each for Data Source in Point 1 above. This Process Model should be created in such a way that it should execute all Process Models defined separately for each Data Source in Parallel.



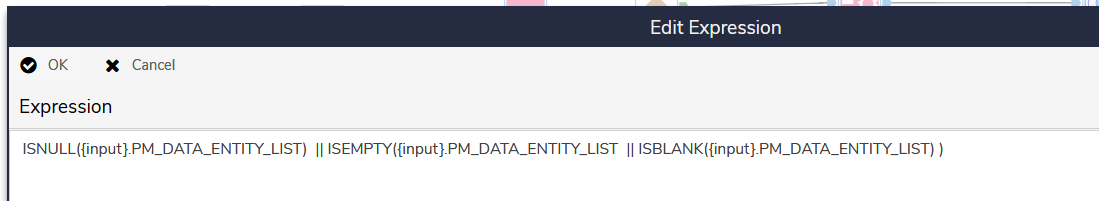
Custom changes to be applied to the PM:

Step 1 :- External Start Event **–** It triggers the execution process. It carries below runtime external parameter from previous analysis execution-

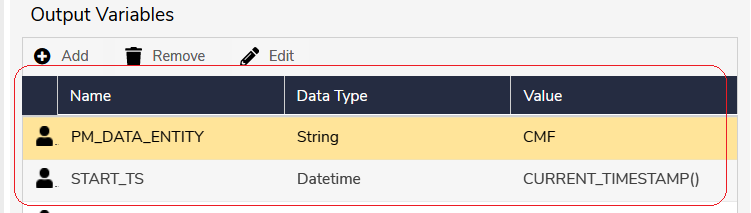
* PM\_DATA\_ENTITY\_LIST – Configure it as shown in below screen shot –



Step 2 :- - Check if DATA ENTITY List is populated – This is the exclusive gateway and will check for the validity of external parameter . If it is NULL then it will exit out otherwise it will go to next branches.



Step 3:- Execute CMF (CMF is an example and is based on Data Sources defined in [Data Source](#DataSources) section)- This is the exclusive gateway and will check if the external Parameter PM\_DATA\_ENTITY\_LIST containing Data Entity “CMF” then it will proceed further otherwise it will exit. Also define 2 parameters “PM\_DATA\_ENTITY” as “CMF” and START\_TS as CURRENT\_TIMESTAMP.



Step 4:- “IGX\_PM\_RULE\_EVALUATE\_CMF” . It will execute the Process Model defined in Point 1 above for CMF Data Source.

Step 5 :- Success Intermediate Event

Repeat Steps 3 to Steps 5 by branching out in parallel from Step 2 for each Data Source Defined in the system.

## Dashboard URL Configuration

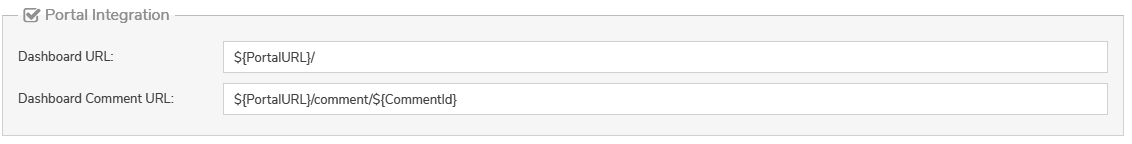
1. Go to Admin and select Settings
2. Check Portal Integration checkbox and add the URL value for configured & connected Govern Environment

               e.g. <https://infogix-fda.data3sixty.com>

1. Go to Admin and select Environments
2. Select the Corresponding Environment
3. In Details Tab , add the URL value for configured & connected Govern Environment in Portal URL Box

               e.g. <https://infogix-fda.data3sixty.com>

1. Go to the Dashboard in pipeline and click on edit stage
2. Click on the settings button at Right upper corner of the Dashboard
3. Go To Settings tab
4. Enable checkbox for Portal Integration as shown below

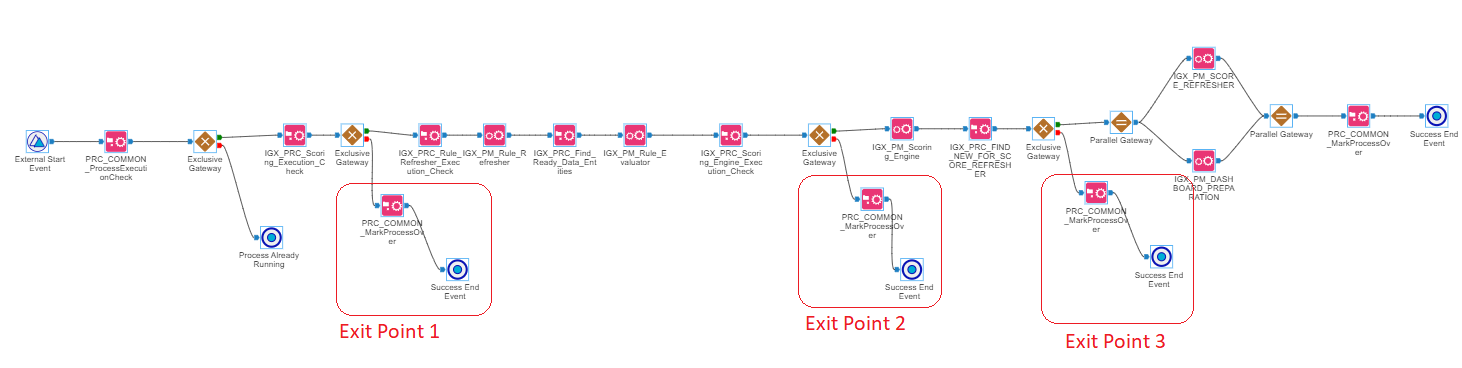


# How to Execute – DQ+ Process Executions

The Process Model “IGX\_PM\_EVALUATE\_DATA\_QUALITY” need to be executed/scheduled (frequency can be few hours everyday)

The Process Model will execute following process models in given sequence:

1. **IGX\_PM\_Rule\_Refresher** - For Rule Refresher
2. **IGX\_PM\_Rule\_Evaluator** (process model(s) created as per process model defined in [Create Process Model section](#CreateProcessModel)) – For Rule Evaluator per source
3. **IGX\_PM\_Scoring\_Engine** – For Scoring Engine
4. **IGX\_PM\_SCORE\_REFRESHER** – For Score Refresher
5. **IGX\_PM\_DASHBOARD\_PREPARATION** – For Dashboard Preparation

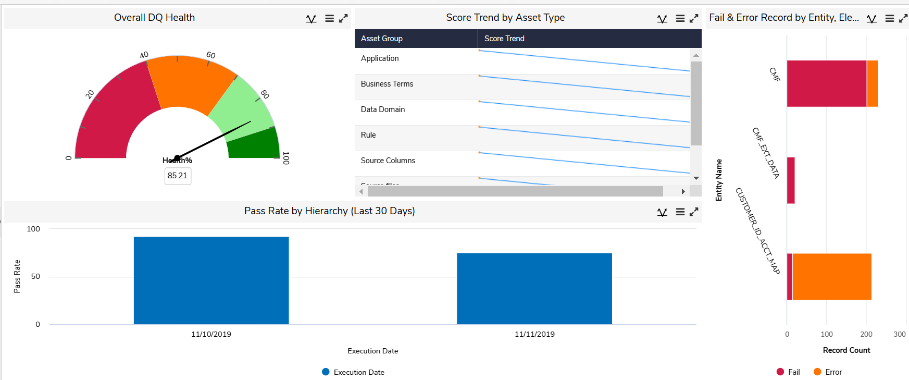


**Note**:

1. The Process Model will check for Duplicate Instance Check first and if the process is already running then it will exit automatically. The PROCESS\_LIST DataSource created in “Common” path needs to be populated with Process Model name as “IGX\_PM\_EVALUATE\_DATA\_QUALITY” for this check to be performed successfully.
2. The Process Model is having 3 exits points.
3. The First Exit point is at the beginning where if Scoring Engine already ran for the day then it will exit and will not execute any other module/components. The scoring engine is expected to execute only once a day.
4. The Second Exit point is after Rule Evaluator. If any Data Source is not yet evaluated (as per the defined frequency for the day) then the process will exit from this point and will not proceed for scoring engine.
5. The third exit point is after Scoring Engine where if Score Refresher is already ran for the day. If it already ran then it will exit out otherwise it will execute score refresher and Dashboard Processes.

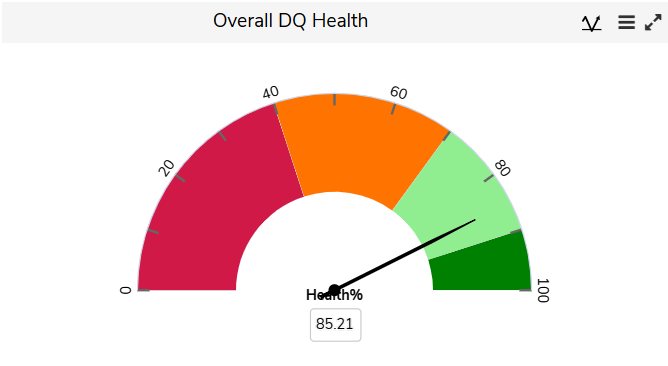
# Dashboards

DQ+ Scoring Dashboard allows user to monitor the overall data quality health and facilitate user to investigate cause of failure and errors in data quality evaluation.



DQ Scoring Dashboard is having below dashlets:

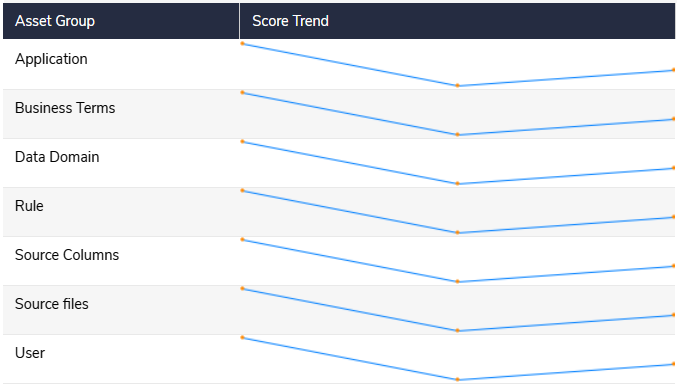
### **Overall DQ Health**



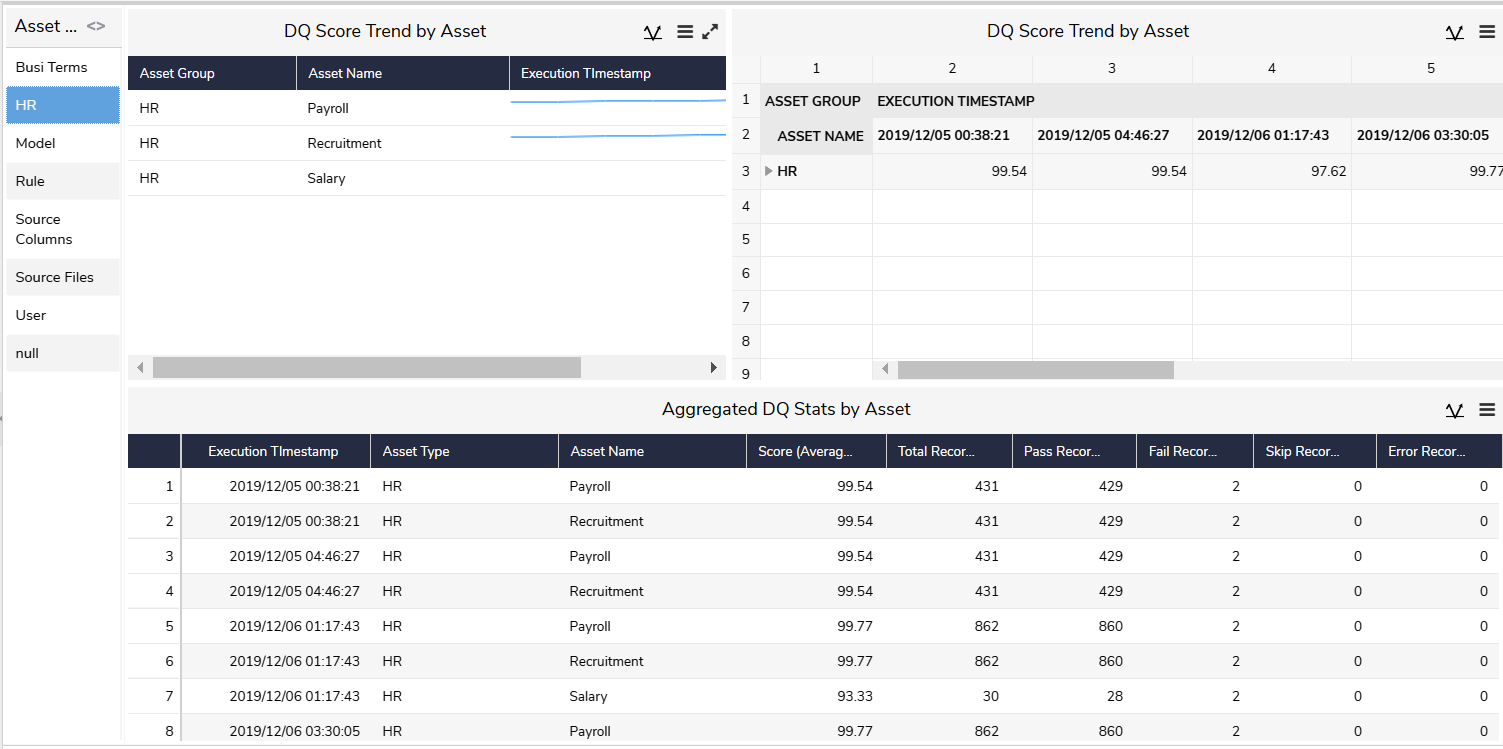
This widget shows the overall DQ score for the entire system. The score is computed as simple average of scores (for their latest DQ runs) for all Source Attributes. The widget shows a gauge on scale of 0 to 100. 0 to 40 shows as Red, 40 to 70 as Orange, 70 to 90 as light green and more than 90 as Dark Green.

This is a non-drillable widget

### **Score Trend by Asset Group**

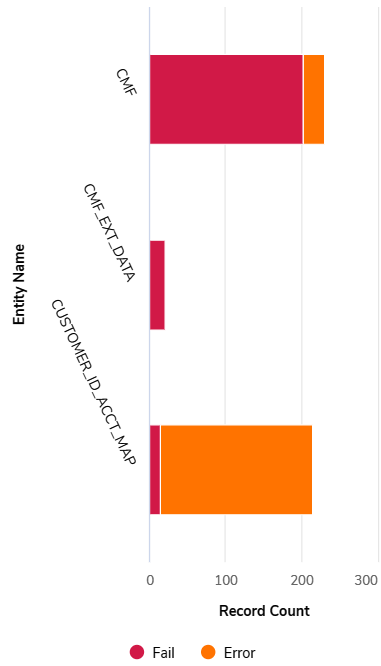


This widget shows the Score trend for last 1 month of different Asset Groups. It is a linked widget, if clicked it opens up another tab with more detailed options as shown in below screenshot.



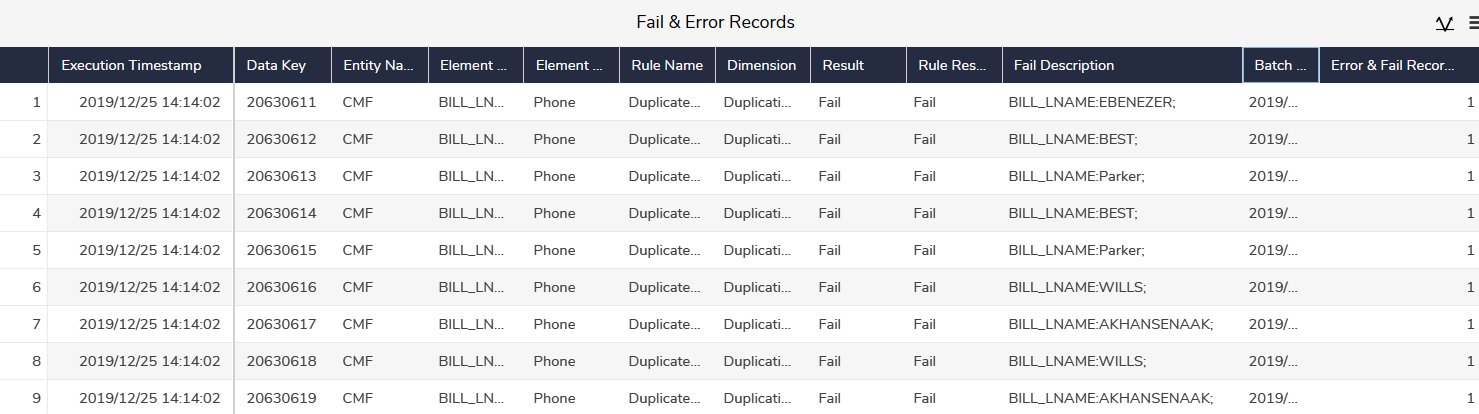
The detailed view allows the user to filter by available Asset Group. The upper view will show the graphical trend for the Assets qualified as per filter criteria and the bottom part will show the stats for Total Records, Pass, Fail, Error and Skip record counts.

### **Failed and Error Records by Data Entity, Element & Rule**

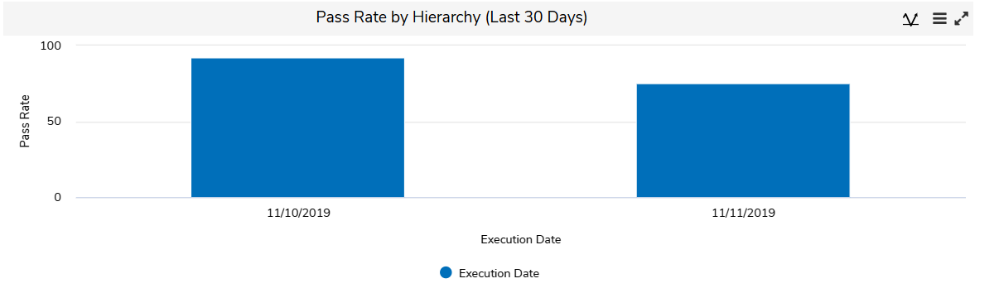


This widget is a stacked column chart that shows count of failed and error records per data source for its last execution. At starting level, the counts are shown by Data Source. User can drill down the counts by Source Attributes and then by Rules, as shown in below screenshots, to investigate the root cause of failure.

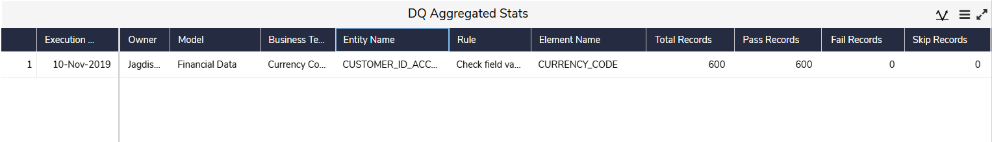
On clicking the Rule Name, a tabular listing displays the detailed records of the source, for which the rule evaluation resulted as Failed or Error.



### **Pass Rate by Hierarchy**



This widget shows the passing rate of different DQ evaluations executed in last 30 days. At starting level, the values are shown by Execution Date. User can drill down it further by Owner to Model to Business Terms to Rule and then a detailed tabular list will appear showing the pass, fail, error and skip counts.



### **URL Configuration for DQ+ Dashboard**

1. Go to Administration and select Dashboard
2. create the dashboard by clicking ‘+’ sign
3. Enter Report Type , Name & Fusion Type
4. Enter the URL of DQ+ Dashboard here in “URL” Box to link the same

               e.g. <https://cafe-dev.infogix.com/desktop/index.html?tenantId=Infogix&portalmode=true#viewdashboard/28247>

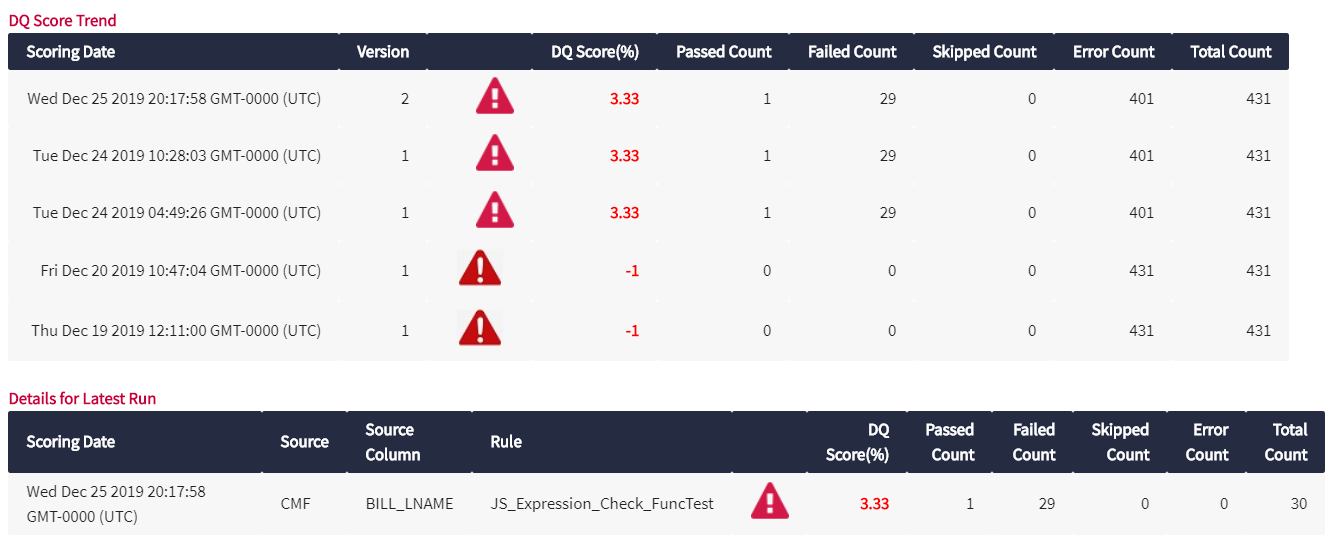
Here TenantID is the Logging-in User name and the above URL is the URL taken from one of the corresponding DQ+ Environment

# Data Quality Results in Data Governance

DQ+ Scoring solution calculates scores of assets and these scores are posted back to Data Governance. Scores of last n executions will be posted in Data Governance for each asset. For data sources, source attributes, business terms and rules detailed scores for the latest run are also posted.

Scores of assets can be seen on asset’s detail screen under Data Quality pane in DQ Results field. Number of executions for which scores need to be posted is configured in [IGX Scoring Config](#IGXScoringConfig) against parameter “SR\_SCORE\_POST\_DAYS”.

Refer below screenshot for DQ Results.



## Asset Versioning

Versions of assets are created based on their creation, update or changes in their relationships. If there is any change in rule definition or asset relationship then the asset version will be incremented by 1 and the cascading effect (w.r.t. asset version) will be there to the related assets in the hierarchy.

## Hierarchy/Score Rollup

We create relationships between assets. While calculating score of assets (other than source attributes), these relationships are transformed as hierarchy of assets, which helps in rolling up scores.

For example, we may have relationship like below:

Rule(s) -> Data Element(s) -> Business Term(s) -> Model(s)

In above example, score of Business Term will be calculated using scores of Data Elements related to it.

Similarly, score of Business Terms and their assigned weightage, if any, will be considered and rolled up to get score for Model.

Refer below diagram for hierarchy and score rollup.

