Business User driven Scorecards to measure Data Quality using SAP BusinessObjects Information Steward

Asif Pradhan
Learning Points

- SAP BusinessObjects Information Steward Overview

- Building Blocks of a Scorecard
  - Create Rules with associated Data Quality Dimensions
  - Define Rule Bindings
  - Create a Rule task and calculate Data Quality score

- Viewing and Analyzing Data Quality Scores

- Defining and Analyzing a Scorecard
  - Key Data Domains
  - Data Quality Dimensions
  - Rules and Rule Bindings

- Case Study – Building a Scorecard
Business Challenges with Data Quality and Data Integration Projects Today

**IT**

- No environment for business users to collaborate with IT regarding data issues
- No consistent repeatable way to measure and score data quality
- No clear strategy and discipline for improving data quality
- Minimal reuse of data assets; lots of data duplication

**BUSINESS USER**

- Not sure what the business definition is
- No ability to analyze data dependencies across systems
- Lack of visibility to where the numbers or data is coming from
- and many more…
SAP BusinessObjects Information Steward
Product Vision

Empower data stewards with a single environment to discover, assess, define, monitor, and improve the quality of their enterprise data asset.

Discover
- Discover & understand enterprise data
  - Data profiling
  - Metadata management

Define
- Define business terms, validation rules, cleansing rules, models
- Define data ownership:
  - Assign data ownership, accountability, and roles

Monitor & Remediate
- Monitor data quality
- Surface data quality score in business user applications
- Workflows to resolve data quality issues
SAP BusinessObjects Information Steward
Value propositions

Improve information governance and quality

**EMPOWER**

**GOVERN**

**IMPROVE**

**Empower Business Users**
Bridge gap between business and IT with collaborative solution for driving information management initiatives

**Govern Enterprise Information**
Enable effective data governance with the industry’s first combined data profiling, metadata management, and data quality monitoring solution

**Improve Information Transparency**
Give instant visibility into data quality levels and origins with end-to-end impact analysis and data lineage
SAP BusinessObjects Information Steward

**Integrated Environment with Multiple Perspectives**

- **Integrated** metadata management, business glossary, data quality assessment, and data quality monitoring, cleansing package builder
- **“One place”** for data stewards, and business analysts to collaborate and govern their data asset
- Various **perspectives** to understand and analyze trustworthiness of data
- **Easy and secure** access to a rich set of metadata & data sources
Target Personas for Information Steward

**Data Steward**
- Supports Data Governance initiatives, business processes
- Approves validation rules
- Reviews data quality scorecard
- Analyzes Scorecard and its impact
- Creates Cleansing Packages

**Data Analyst/IT**
- Performs data profiling and quality assessment
- Defines validation rules
- Defines and reviews data quality scorecards
- Creates Cleansing Packages
- Analyzes Metadata impact/lineage

**IT Administrator**
- Maintains Information Steward landscape
- Manages user security
- Manages IS projects and connections
- Configures IS application
- Manages IS tasks and utilities
Measuring Data Quality using Validation Rules

- Validation rules are managed in the Rules tab page
- Validation rules are associated to Data Quality Dimensions
- A rule must return a true or false
- Rules are reusable ‘functions’. They can be:
  - reused against multiple tables or views within the same project
  - shared with other projects
  - exported to another instance of Information Steward
  - exported to Data Services
Exporting Rules

- Use Ctrl+click to select multiple rules to export
- Rules can be exported into an XML zip file for import into another instance of Data Insight or as a backup
- Rules can also be exported to Data Services (DS) via DS ATL file or directly to DS repository configured on BI platform CMC
Quality Dimensions in Information Steward

- Each rule has one quality dimension
- In Information Steward 4.0, there are 7 commonly used quality dimensions. Quality dimensions are used to organize rules in the Rules menu, and aggregate rule scores in a scorecard

<table>
<thead>
<tr>
<th>Quality Dimensions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>The extent to which data objects correctly represent the “real-world” values for which they were designed</td>
</tr>
<tr>
<td>Completeness</td>
<td>The extent to which data is not missing</td>
</tr>
<tr>
<td>Conformity</td>
<td>The extent to which data conforms to specified format</td>
</tr>
<tr>
<td>Consistency</td>
<td>The extent to which distinct data instances provide non-conflicting information about the same underlying data object</td>
</tr>
<tr>
<td>Integrity</td>
<td>The extent to which data is not missing important relationship linkages</td>
</tr>
<tr>
<td>Timeliness</td>
<td>The extent to which data is sufficiently up-to-date for the task at hand</td>
</tr>
<tr>
<td>Uniqueness</td>
<td>The extent of uniqueness in naming and representation of core data objects</td>
</tr>
</tbody>
</table>
Defining Validation Rules

Rule names must be unique and can be a maximum of 50 characters long.

Indicates if the column bound to this parameter will have a score. Typically, parameters not used for filtering should be scored.

Switch from this Basic Rule Definition Editor to Advanced Rule Editor

Filter criteria for rows to be read for processing

Rule script (Data Services ATL script) automatically generated by editor

Basic Rule Editor for comparison, data type, and pattern validation
Binding a Rule

1. Select the connection and table
2. The Binding Editor automatically binds a parameter to a column that has the same name
3. To adjust the threshold, select the marker and then either drag it or use left/right arrow keys
4. Scores below the low threshold can be used for alert notification
Defining a Rule Task
Step 1 of 2

- One or more rules can be bound to a table
- A rule task can only be created at the table level
- A rule task may have more than one table
- When a task is executed, all the rules related to the columns of a table will be executed
- A rule task is defined on the Rule Result tab page
- Select one or more tables, and then click **Calculate Score** to define a rule task
Defining a Rule Task
Step 2 of 2

- Email notification can be sent for rule scores that are below the low threshold
- All rows that failed any rules can be written to a separate failed data connection for further analysis and remediation
- Failed data connection is defined in the BI platform Central Management Console (CMC) → Information Steward (IS) → Data Insight connection
- A maximum of 500 sample failed rows can be stored in the IS repository for analysis on IS UI
- A rule task be executed ‘Now’ on IS or scheduled on BI platform CMC
Viewing and Analyzing Data Quality (DQ) Scores

Information Steward provides built-in dashboards and scorecards, which enable data stewards and data analysts to easily analyze and understand data quality at various levels and from various perspectives.

<table>
<thead>
<tr>
<th>Analyze Data Quality Score</th>
<th>Built-in Analysis Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Rule</td>
<td>Rules DQ Dashboard (in Rules Menu)</td>
</tr>
<tr>
<td>By Table</td>
<td>Table DQ Dashboard (in Working Set → Rule Result)</td>
</tr>
<tr>
<td>By Table Column</td>
<td></td>
</tr>
<tr>
<td>By User-defined Key Data Domain</td>
<td>DQ Scorecards</td>
</tr>
</tbody>
</table>
Analyze Data Quality by Rules

DQ Rules Dashboard

- On Rules tab page, user can view rule score for each table/column binding
- A score is on a scale of 0 to 10. Score of 10 means 100% of the rows processed passed the validation rule
- Click the score to drill into sample failed rows
Analyze Data Quality by Table and Columns
Table DQ Dashboard

Table level score measures number of rows that passed all rules.

Column level score measures number of rows that passed all rules for this column.

Detailed statistics of number of failed and total rows.

Data quality trend over different time range.

Sample failed rows for analysis of the last rule execution.
Scorecard enables a data steward to monitor the quality of key data domains (KDD), using data quality rules for table/columns that are most indicative of the quality of the KDD.

- The same low/high threshold defined for the KDD applies to the quality domain, rule, and rule binding related to the KDD.
- Scores can be weighted at each KDD, quality dimension, rule, and rule binding level.
- Score weights can be fixed or evenly distributed.
- Scorecard value calculation task is scheduled on BI platform CMC.
- User can also view the latest scores using the “Show Latest score” option.
The “Scorecard calculator” task calculates the scorecard for all projects. The KDD scores are calculated based on the latest score of the rule bindings used by the scorecard (at the time this utility is executed).

- The task is scheduled to run daily by default.
- Scorecard (KDD score) Calculator task scheduled only in BI platform CMC.
- Choose BI platform CMC → Applications → IS → Actions → Manage Utilities → KDD Score Calculator → Actions → Schedule.
- Scorecard and rules results are retained for three years by default; retention period can be modified in BI platform CMC.
Analyze DQ User-defined Key Data Domain Scorecard – Summary View

- View scores computed now or computed by scheduled scorecard task
- Scores computed now are not stored in the IS repository.
- Click **Show more…** or **Details** to analyze the scorecard
Analyze DQ User-defined Key Data Domain Scorecard – Detail View

Select KDD, Quality Dimension, Rule, and Rules Binding to analyze data quality based on any combination of these attributes

View Rule Definition

Switch to Workspace

Analyze sample failed rows for the rule binding

Analyze impact of failed data
Scorecard – Case Study

Edna is a data steward responsible for the Sales Operations application in her company.

She is also responsible for the data quality of Customer and Product data.

Using existing validation rules previously created in Information Steward, she decides to leverage these rules to create her own scorecards for the Customer and Product data domains.

Let’s explore how Edna creates a Scorecard for the Customer data domain.
Scorecard – Case Study

Edna’s Sales Operations Project has two tables: Customer and Product.
Scorecard – Case Study

Edna reviews existing Validation rules previously created in Information Steward.
Scorecard – Case Study

Edna binds rules to the Customer table columns.

*Rule*: **Customer Email Address Type must be Valid**
Scorecard – Case Study

Edna binds rules to the Customer table columns.

*Rule*: **Customers must belong to countries serviced**
Scorecard – Case Study

Edna binds rules to the Customer table columns.

*Rule:* **Customer Address information must be entered**
Edna binds rules to the Customer table columns.

*Rule:* **Customer title must be entered**
Scorecard – Case Study

Edna binds rules to the Customer table columns.

Rule: Customer Postal Code must be Valid
Scorecard – Case Study

Edna binds rules to the Customer table columns.

*Rule:* **Customer Telephone Number format must conform to standard**
Scorecard – Case Study

Edna creates a rule task for the rules that were bound to the Customer table. She then runs the rule task to generate scores for the rules that will be used in the scorecard.
Scorecard – Case Study

Edna wants to create a Customer scorecard, corresponding to the Key Data Domains (KDD) that are pertinent to the sales operations data quality performance measurement.

In this case, Key Data Domain = **Customer**

She will leverage the rule bindings which were created in this project for her scorecard.
Scorecard – Case Study

Edna adds Quality Dimensions: **Accuracy, Completeness, and Conformity.**
Scorecard – Case Study

Edna’s Scorecard Setup depicts the **Key Data Domain**: Customer, **Quality Dimensions**: Accuracy, Completeness, and Conformity, as well as the **rules** and **rule bindings** defined earlier.
Scorecard – Case Study

Edna wants to analyze her Scorecard.

In the Workspace/Scorecard drop-down list, Edna selects Scorecard.
Scorecard – Case Study

Edna analyzes Scorecard details by exploring Data Quality Dimensions, associated Validation Rules, and Rule Bindings.
Scorecard – Case Study

Edna analyzes Scorecard details.

*Data Quality Dimension: Accuracy*
Scorecard – Case Study

Edna analyzes Scorecard details.

*Data Quality Dimension: Completeness*
Scorecard – Case Study

Edna analyzes Scorecard details.

*Data Quality Dimension: Conformity*
Information Steward allows business users to understand data quality from various levels and perspectives.

Scorecards enable data stewards to monitor the quality of key data domains (KDD), using validation rules established by the business.

Use scorecards to analyze data quality based on any combination of key data domains, quality dimensions, rules, and rule bindings.

Scorecards may be explored in a summary view or in detail by examining specific data quality rules.
Enterprise Information Management with SAP

- Understand the big picture of SAP’s enterprise information management offerings
- Explore step-by-step instructions for working with SAP Data Services
- Learn how to perform the most important tasks in SAP Information Steward, SAP NetWeaver Information Lifecycle Management, SAP Master Data Governance, and more
- All royalties donated to Doctors Without Borders
Thank you for participating.

Please provide feedback on this session by completing a short survey via the event mobile application.

SESSION CODE: 0216

Learn more year-round at www.asug.com