Usage Scenario – Data Profiling, Data Quality Assessment, Data Quality Monitoring

Purpose

This article will discuss SAP Information Steward backend servers and services that are leveraged in Data Profiling, Data Quality Assessment and Data Quality Monitoring scenarios.

Overview

This scenario is implemented using Data Insight module within Information Steward. Within these scenarios, following backend servers and services are leveraged.

Backend servers for Data Profiling, Data Quality Assessment and Monitoring

We will now take a look at each service within the context of this usage scenario.

Metadata Browsing Service

Metadata Browsing Service is built on top of robust enterprise connectivity layer within SAP Data Services. Within Data Insight, when you add a table to your project workspace, Information Steward uses Metadata Browsing service to display the metadata structures in ‘Browse Metadata’ dialog (as shown in the diagram below) and consequently import metadata into Information Steward repository.
View Data Service

View Data Service also leverages enterprise connectivity layer within SAP Data Services in order to display data within Information Steward. Note that within Information Steward, as a part of data discovery scenario, data stewards and data analysts want to understand how data looks but do not necessarily want to see all of the records. View Data Service fetches a limited set of records (typically 500) from the source system. Any filtering or sorting criteria defined within Information Steward UI is applied on all of the records within source but only top 500 rows can be viewed within Information Steward. This functionality is leveraged within following user interactions:

1. Add ‘Table’ to the project workspace - View Data within ‘Browse Metadata’ dialog
2. View Data from a table in Data Insight project ‘Workspace Home’ tab.
3. View More Failed Data from the Scorecard screen or Rule Results screen.
Data Services Job Server

Data Services Job Server is the core ETL platform component that executes ‘data profiling’ and ‘data validation’ tasks. These tasks are expressed as ETL jobs in Data Services where source is same as ‘Data Profiling Source’ defined in Information Steward. These jobs populate Information Steward repository with profile results, rule results as well as sample data associated with these results. In addition, if ‘Failed Data Connection’ is defined for Information Steward rule tasks, data validation job populates failed data repository with all the records that failed the validation rules.

Note that Data Services job server has to be associated with Information Steward repository using Data Services Server Manager in order to make it available for execution of ‘data profiling’ and ‘data validation’ tasks. More than one instances of Data Services job servers can be associated with Information Steward repository for load balancing and high availability.
Information Steward Task Scheduling Service

Information Steward Task Scheduling Service processes ad-hoc and scheduled ‘data profiling’ and ‘data validation’ tasks. This service is hosted within Information Steward job server which is deployed in BI platform. Information Steward job server leverages scheduling service of BI platform. It distributes data profiling and data validation tasks to all available instances of Data Services job server.
Information Steward Administrative Task Service

Information Steward repository within itself contains Data Services repository tables. Data Profiling and Data Validation tasks are internally stored as jobs within Data Services portion of Information Steward repository. Data Insight connections, validation rules, file formats and virtual views are stored using native Data Services constructs like data stores, functions etc.

Information Steward Administrative Task Service is used to manage ‘Data Services’ portion of Information Steward repository and also perform tasks on Data Services like ‘test rules’, ‘export rules’ etc. Administrative Task Service is deployed as a part of ‘primary installation’ of Information Steward. It leverages local Data Services instance to perform these tasks.

For further details around architecture overview, please refer to SAP Information Steward Administrator Guide.

Related Content

Related Documents
Related SAP Notes/KBA