KOSHA.AI - Document Curator & Analyst: End User Training Guide

Welcome!

This guide will help you understand and effectively use the Smartello Document Curator & Analyst application within your Snowflake environment. This powerful tool is designed to simplify how you manage, document, understand, and interact with your data.

What is KOSHA.AI Document Curator & Analyst?

It's an application that runs directly inside Snowflake, helping you:

- Keep your data documentation up-to-date automatically.
- Easily onboard new data solutions.
- Define and monitor data quality.
- Understand who has access to what data.
- Optimize your Snowflake queries and warehouse usage.
- Ask questions about your data in plain English.
- Generate starter code for data pipelines.

Accessing the Application

You will access the Document Curator & Analyst through your Snowflake account. Your Snowflake administrator will provide you with the necessary access and URL.

Understanding Your Role & Permissions

The features you see in the application depend on your assigned "persona" (e.g., Admin, Data Engineer, Business Analyst). The sidebar will show your current User, Snowflake Role, and Application Persona.

Navigating the Application

The main navigation is in the **sidebar on the left**. Features are grouped into categories:

- **Manage Onboarding:** For registering and tracking data solutions.
- **Data Quality:** For ensuring your data is accurate and reliable.
- **Productivity and Documentation:** For configuring automated documentation and generating code.
- **Governance and Optimization:** For understanding data access and improving Snowflake performance.
- **Semantic Layer & Analysis:** For making data easier to query using natural language.

COPYRIGHT OF SMARTELLO ANALYTICS | 2025

- **Manage App:** (Usually for Admins) For application-level settings.
- **Help:** Access to this guide and other helpful information.

Simply click on a **Category**, then click on an **Option** within that category to access a specific feature.

Key Features & How to Use Them

Here's a breakdown of the main features you might use:

1. Managing Data Onboarding

- **Add New Onboarding:**
    - **Purpose:** To register a new application, report, or data solution that uses Snowflake data.
    - **How to Use:**
        1. Navigate to Manage Onboarding > Add New Onboarding.
        2. Fill in the Solution ID (a unique identifier), Functional Area (e.g., Sales, Finance), Solution Name, Data Product Owner (email), and Data Steward (email).
        3. Optionally, add Roles that access this solution.
        4. Click Submit Onboarding.
    - **Next Steps:** After submitting, you'll use Configure Onboarding to specify which database objects (views, procedures, functions) belong to this solution.
- **Configure Onboarding:**
    - **Purpose:** To tell Smartello which specific views, procedures, and functions are part of an already onboarded solution and how their documentation should be generated.
    - **How to Use:**
        1. Navigate to Productivity and Documentation > Configure Onboarding.
        2. Select the Solution ID you want to configure.
        3. The existing details (Business, Solution Name, Owners) will be displayed. You can update owner information here.
        4. Use the expander sections (Configure Views, Configure Procedures, Configure Functions):

COPYRIGHT OF SMARTELLO ANALYTICS | 2025

        - Select the Database and Schema.
        - Choose the specific Views, Procedures, or Functions from the multi-select list.
        - For each object type (especially views and procedures), you can enter a specific **Prompt**. This prompt guides the AI in generating relevant documentation (e.g., "Focus on the business purpose and key calculations for this view"). For procedures/functions, you can also define **Code Quality Rules**.
    5. Click Onboard Objects to save your selections and prompts.

- **Approve Onboarding:** (Typically for Managers/Admins)
    - **Purpose:** To review and approve or reject new onboarding requests.
    - **How to Use:**
        1. Navigate to Manage Onboarding > Approve Onboarding.
        2. View the list of onboarded records.
        3. Select a Solution ID to see its details.
        4. Set the Approval Status (Approved/Rejected) and add Comments.
        5. Click Update Approval Status.
- **Scan History:**
    - **Purpose:** To see a log of when documentation scans were run.
    - **How to Use:**
        1. Navigate to Manage Onboarding > Scan History.
        2. View the summary and detailed logs of past scan activities.

2. Modifying AI-Generated Documentation

- **Purpose:** To review and manually edit the documentation that the AI has generated for your views, procedures, or functions.
- **How to Use:**
    1. Navigate to Productivity and Documentation > Modify Documentation.
    2. In the sidebar, select the Solution ID, Object Type (View, Procedure, or Function), and the specific Object Name.
    3. The current AI-generated documentation will appear in a text area.

COPYRIGHT OF SMARTELLO ANALYTICS | 2025

4. Make your edits directly in the text area.
5. Click Save Documentation. Your changes are saved, and the previous version is archived.

3. Data Quality (SmartDQ)

- **My DQ Dashboard:**
    - **Purpose:** To get an overview of your data quality.
    - **How to Use:**
        1. Navigate to Data Quality > Data Quality (main tab is Dashboard).
        2. Select Databases, Schemas, and Tables to filter the dashboard.
        3. View metrics like Number of Records Processed, Error Records, Warning Records, and trends over time.
- **Define DQ Rules:**
    - **Purpose:** To set up specific data quality checks for your table columns (e.g., a column should not be empty, a number must be within a certain range).
    - **How to Use:**
        1. Navigate to Data Quality > Data Quality and select the Define DQ Rules tab.
        2. Select the Database, Schema, Table, and Column you want to apply a rule to.
        3. Existing rules for that column will be displayed.
        4. Under "Show Rules Library," click on a pre-defined rule type (e.g., Nullability, Min, Max) to pre-fill an instruction.
        5. Modify the instruction in the text area (e.g., "Add a test for Min by checking if the value is less than 10 with on fail config set to error.").
        6. Click Process Instruction. The AI will try to configure the rule based on your text.
        7. Once rules are configured, you can click Run Analysis now to execute them.
- **SmartDocs (View Documentation):**

COPYRIGHT OF SMARTELLO ANALYTICS | 2025

- **Purpose:** To see detailed AI-generated documentation for your views, including business descriptions, technical details, and improvement suggestions.
- **How to Use:**
    1. Navigate to Data Quality > Data Quality and select the SmartDocs tab.
    2. Select the Database, Schema, and View.
    3. The AI-generated documentation will be displayed in expandable sections.

4. User & Role Governance

- **Purpose:** To understand who has access to what in Snowflake (users, roles, permissions on objects).
- **How to Use:**
    1. Navigate to Governance and Optimization > User & Role Governance.
    2. Choose to view governance by User, Role, or Object.
    3. **User/Role View:** Select a specific user or role from the dropdown. An interactive tree will show their role assignments, inherited roles, and permissions. Click the ▶ icons to expand sections.
    4. **Object View:** Select a Database, Schema, Object Type, and Object Name. Click Analyze Object Permissions to see which roles and users have access to that specific object.

5. Warehouse & Query Optimization

- **Purpose:** To monitor your Snowflake warehouse costs and get AI help to make your queries run faster and cheaper.
- **How to Use:**
    1. Navigate to Governance andOptimization > Warehouse & Query Optimization.
    2. **Warehouse Costs:** See a chart of credit usage by warehouse for the last 30 days.
    3. **Longest Running Queries:** View a list of the slowest queries from the last 7 days.
        - Expand a query to see its text and details.

COPYRIGHT OF SMARTELLO ANALYTICS | 2025

        - Click Optimize with Cortex AI. The AI will analyze the query and provide suggestions to improve it. *Always review AI suggestions carefully before applying them.*

6. Semantic Model Generator

- **Purpose:** To create a "semantic model" – a special file that helps another AI (Cortex Analyst) understand your data's business meaning so you can ask questions in plain English.
- **How to Use:**
    1. Navigate to Semantic Layer & Analysis > Semantic Model Generator.
    2. Enter a Model Name and Model Description.
    3. **Tables & Views Tab:** Add tables or views from your Snowflake databases/schemas to be part of this model.
    4. **Metrics Tab:** Define key business calculations (e.g., "Total Sales" as SUM(orders.amount)).
    5. **Relationships Tab:** Define how your selected tables are related (e.g., customers.customer_id joins to orders.customer_id). You can add manually or try Auto-Suggest Relationships.
    6. **Generate Model Tab:** Click Generate Enhanced YAML. The model definition will appear.
        - You can Download YAML File or Save to Snowflake Stage (your admin might need to deploy it from the stage).

7. Chat with Data

- **Purpose:** To ask questions about your data using natural language (e.g., "What were total sales last month?") and get answers directly from Snowflake.
- **Prerequisite:** A Semantic Model (see above) must be created and *deployed* by an administrator.
- **How to Use:**
    1. Navigate to Semantic Layer & Analysis > Chat with Data.
    2. Select a Deployed Semantic Model from the dropdown.
    3. Type your question in the chat input box at the bottom (e.g., "Show me top 5 customers by revenue").

COPYRIGHT OF SMARTELLO ANALYTICS | 2025

4. The AI (Cortex Analyst) will try to answer your question and may show you the SQL query it used.
5. Your chat history for the selected model is displayed.

8. Code Generation

- **Purpose:** To get AI assistance in generating starter code for data ingestion pipelines (e.g., code to load data from S3 into Snowflake).
- **How to Use:**
    1. Navigate to Productivity and Documentation > Code Generation.
    2. In the sidebar, configure:
        - Onboarded Application ID (the solution this code is for).
        - LLM Model to use.
        - Data Source Type (e.g., S3, SQL Server) and its details.
        - Snowflake Target table names (staging and final).
        - Processing Tool (e.g., PySpark, SQL Script).
        - Standard Features to include (e.g., logging, error handling).
    3. In the main area, either:
        - **Describe the logic:** Write a text description of what the code should do.
        - **Upload Flow Diagram:** Upload an image (PNG, JPG) of your data flow.
    4. Click Generate Code. The AI will generate code based on your input. *Always review and test AI-generated code thoroughly.*

Tips for Best Results:

- **Be Specific with Prompts:** When configuring documentation or DQ rules, clear and detailed prompts or instructions yield better AI results.
- **Review AI Output:** AI is a powerful assistant, but always review documentation, code, or query suggestions before relying on them or implementing them in production.
- **Patience with ACCOUNT_USAGE:** Features relying on SNOWFLAKE.ACCOUNT_USAGE views (like governance and some optimization metrics) might have a data lag of a few hours.

COPYRIGHT OF SMARTELLO ANALYTICS | 2025

- **Use the Help Text:** Most input fields and buttons have a help icon or tooltip. Hover over them for more information.

Troubleshooting:

- If a page is not loading correctly, try a browser refresh.
- If you encounter errors, note down the error message and the steps you took.
- Contact your internal Smartello administrator or support team for assistance.

We hope this guide helps you make the most of the Smartello Document Curator & Analyst!

COPYRIGHT OF SMARTELLO ANALYTICS | 2025