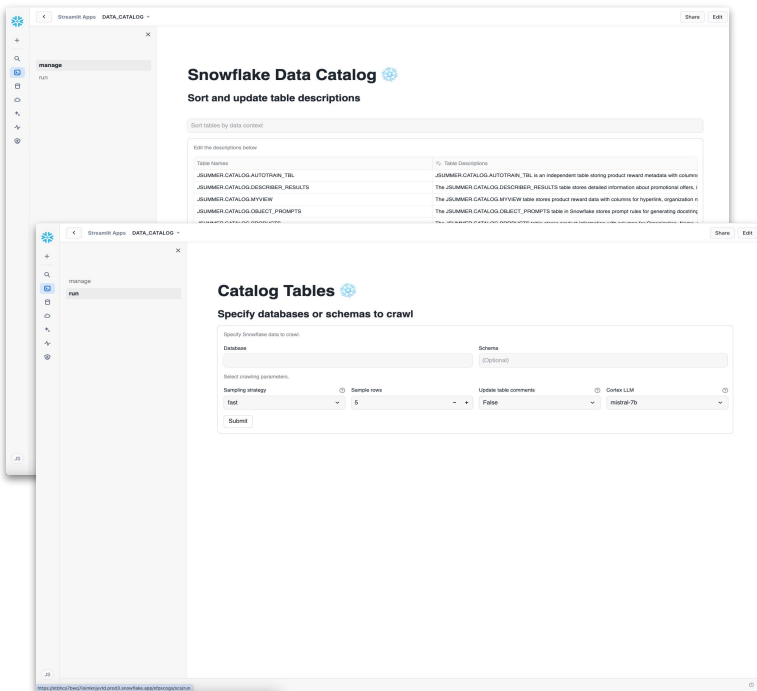


Solution Brief

Data Crawler



Solution Overview

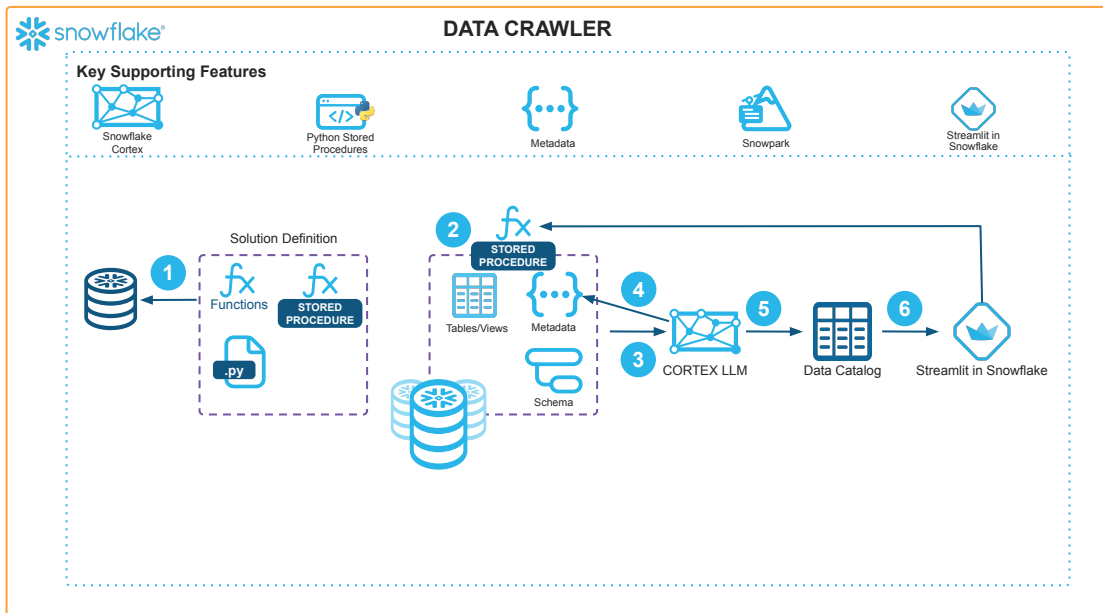
The Data Crawler is an LLM-powered utility that prompts a Cortex Large Language Model (LLM) to generate a natural language description of each table and view contained in a Snowflake database and/or schema. The output of the utility is a catalog table containing natural language summaries of tables' contents which can be easily reviewed, searched, and revised by team members. The utility can also update table/view comments directly.

Solution Benefits

- Provides a self-service method for cataloging and commenting data at scale with asynchronous stored procedures
- Updates Snowflake table/views with comprehensive descriptions
- Data access for crawling follows existing role-based security protocols
- Results can easily be searched, managed, and updated using Streamlit in Snowflake interface



Reference Architecture



Utility Process & Logic

- 1 Register Solution Definitions (scripts, artifacts)
- 2 Stored Procedure Crawls Database/Schema Metadata
- 3 Table/View Data passed to LLM to Generate Descriptions
- 4 Table/View Comments Updated with Descriptions
- 5 Descriptions Recorded in Catalog Table
- 6 App for Managing Descriptions and Running Crawler

