

Feature Metadata and Versioning Documentation

(Snowflake)

Introduction

This document outlines the implementation of a feature store using Snowflake, focusing on metadata management and version control for engineered features.

Feature Metadata

Each feature is stored with associated metadata to ensure traceability and ease of management. Below are the key metadata attributes:

Attribute	Description
Feature Name	Unique name of the feature
Description	Explanation of the feature
Source	Origin of the data (e.g., table, pipeline)
Data Type	Type of data stored (e.g., INTEGER, FLOAT, STRING)
Last Updated	Timestamp of the last modification
Owner	Person or team responsible for the feature
Version	Versioning information to track changes
Creation Date	Timestamp of initial creation

Feature Store Implementation in Snowflake

Features are stored in a Snowflake table with metadata columns for tracking. Below is an example schema:

```
create or replace TABLE PROJECT.PUBLIC.FEATURE_METADATA (  
    FEATURE_NAME VARCHAR(16777216) NOT NULL,  
    DESCRIPTION VARCHAR(16777216),  
    SOURCE_TABLE VARCHAR(16777216),  
    VERSION VARCHAR(16777216),  
    LAST_UPDATED TIMESTAMP_NTZ(9) DEFAULT CURRENT_TIMESTAMP(),  
    primary key (FEATURE_NAME)  
);
```

Feature Versioning

Versioning is crucial to track changes in feature definitions over time. The following versioning strategy is implemented:

1. **Incremental Versioning:** Each update to a feature results in a new version (e.g., v1.0, v1.1).
2. **Timestamp-based Tracking:** Changes are logged with timestamps.
3. **Audit Table:** Historical versions of features are stored separately for traceability.

Version Control Table Example

```
CREATE TABLE feature_version_history (  
  feature_name STRING,  
  version STRING,  
  change_description STRING,  
  modified_by STRING,  
  modified_at TIMESTAMP  
);
```

Querying Feature Metadata

Retrieve metadata for a specific feature:

```
SELECT * FROM feature_store WHERE feature_name = 'Avg Spend Per Tenure';
```

Retrieve version history for a feature:

```
SELECT * FROM feature_version_history WHERE feature_name = 'Avg Spend Per Tenure' ORDER  
BY modified_at DESC;
```

This implementation ensures that feature metadata and versioning are well-managed in Snowflake, supporting efficient feature retrieval for training and inference workflows.