

Snowflake

Snowpark Python

Dynamic Tables

Streamlit

Medallion

■■■■■■■■ Premier League ■■ La Liga ■■ Bundesliga ■■ Serie A ■■ Ligue 1

Source de données: football-data.org API

Refresh: Automatique toutes les 6 heures

Architecture: Medallion (RAW → SILVER → GOLD)

■ Table des Matières

- 1. Vue d'ensemble
- 2. Architecture
- 3. Features Snowflake
- 4. Modèle de Données
- 5. Pipeline ETL
- 6. Ordre de Déploiement
- 7. Dashboard Streamlit

1. Vue d'ensemble

SnowGoal est un pipeline de données analytics pour le football européen, construit entièrement avec les fonctionnalités natives de Snowflake. Le projet démontre l'utilisation de 15+ features Snowflake dans un cas d'usage réel.

Ligues Couvertes

Drapeau	Ligue	Code
■■■■■■■■■	Premier League	PL
■■■	La Liga	PD
■■■	Bundesliga	BL1
■■■	Serie A	SA
■■■	Ligue 1	FL1

2. Architecture

Pipeline Flow

■ API → ■ RAW → ■ STAGING → ■ SILVER → ■ GOLD → ■ Streamlit

Diagramme



3. Features Snowflake Utilisées

Feature	Description
■ VARIANT Columns	Stockage JSON semi-structuré pour les données API brutes sans schéma fixe.
■ Streams (CDC)	Capture des changements en temps réel sur les tables RAW pour le traitement incrémental.
■ Dynamic Tables	Rafraîchissement automatique des agrégations sans orchestration externe.
■ Snowpark Python	Procédures stockées en Python pour l'appel API et transformation des données.
■ External Access	Intégration sécurisée avec l'API externe football-data.org.
■ Secrets Management	Stockage sécurisé de la clé API sans exposition dans le code.
■ Tasks & DAG	Orchestration native avec dépendances entre tâches.
■ MERGE	Chargement incrémental avec upsert pour éviter les doublons.
■ Streamlit in Snowflake	Dashboard natif sans infrastructure externe.
■ RBAC	Contrôle d'accès par rôles (Admin, Analyst, Viewer).
■ Masking Policies	Protection des données sensibles (ex: date de naissance).
■ Internal Stages	Stockage des fichiers Python et Streamlit.

4. Modèle de Données

Schemas

Schema	Description	Objets
RAW	Données brutes JSON	RAW_MATCHES, RAW_TEAMS, RAW_STANDINGS...
STAGING	Views LATERAL FLATTEN	V_MATCHES, V_TEAMS, V_STANDINGS...
SILVER	Tables nettoyées	MATCHES, TEAMS, STANDINGS, SCORERS...
GOLD	Dynamic Tables	DT_LEAGUE_STANDINGS, DT_TOP_SCORERS...
COMMON	Objets partagés	Stages, Secrets, Procédures, Integrations

Tables Silver

Table	Colonnes Clés	Description
MATCHES	MATCH_ID, HOME_TEAM, AWAY_TEAM, SCORE	1000000 matches avec scores
STANDINGS	TEAM_ID, POSITION, POINTS, WON, LOST	1000 classements par ligue
TEAMS	TEAM_ID, TEAM_NAME, VENUE, COACH	100 informations équipes
SCORERS	PLAYER_ID, GOALS, ASSISTS, TEAM	Top buteurs
COMPETITIONS	COMPETITION_CODE, NAME, AREA	Infos compétitions

5. Pipeline ETL

Flux de Données

Étape	Source	Destination	Méthode
1. Extract	football-data.org	RAW.RAW_*	Snowpark + External Access
2. Flatten	RAW.RAW_*	STAGING.V_*	Views + LATERAL FLATTEN
3. Transform	STAGING.V_*	SILVER.*	MERGE (incremental)

4. Aggregate	SILVER.*	GOLD.DT_*	Dynamic Tables (auto)
--------------	----------	-----------	-----------------------

Orchestration (Tasks DAG)

```
TASK_FETCH_ALL_LEAGUES (Schedule: every 6 hours)
■
■ Fetch PL → PD → BL1 → SA → FL1
■ (60s delay between each for rate limiting)
■
■■■■ TASK_MERGE_TO_SILVER (After: FETCH)
■
■■■■ Dynamic Tables (Auto-refresh: 30min - 1h)
```

6. Ordre de Déploiement

#	Script	Description
1	00_init/01_database.sql	Database, Schemas, Warehouses
2	00_init/02_file_formats.sql	JSON format, Stages
3	01_raw/01_tables.sql	Tables RAW (VARIANT)
4	01_raw/02_streams.sql	Streams CDC
5	Créer Secret + Upload Python	FOOTBALL_API_KEY + fichiers .py
6	01_raw/03_stored_procedure.sql	Network Rule, Integration, Procedure
7	01_raw/04_fetch_all_procedure.sql	Procedure FETCH_ALL_LEAGUES
8	CALL FETCH_ALL_LEAGUES()	Chargement initial (~5 min)
9	02_staging/01_views.sql	Views FLATTEN
10	03_silver/01_tables.sql	Tables Silver
11	03_silver/02_merge.sql	MERGE RAW → Silver
12	04_gold/01_dynamic_tables.sql	Dynamic Tables
13	05_tasks/01_tasks.sql	Tasks DAG
14	06_security/01_rbac.sql	RBAC + Masking
15	07_streamlit/01_deploy_app.sql	Dashboard Streamlit

7. Dashboard Streamlit

Pages

Page	Description
■ Home	Vue d'ensemble avec stats live (Matches, Teams, Scorers, Leagues)
■ Standings	Classements par ligue avec KPIs et graphiques
■ Top Scorers	Podium des meilleurs buteurs avec stats détaillées
■ Matches	Résultats récents et prochains matches

■ ■ Teams	Détails équipes avec stats Home/Away
-----------	--------------------------------------

Accès

Le dashboard est accessible via **Streamlit in Snowflake** :

Snowsight → Streamlit → SNOWGOAL_DASHBOARD

SnowGoal - Football Analytics Pipeline

Built 100% on Snowflake Native Features

Data Source: football-data.org | Architecture: Medallion (RAW → SILVER → GOLD)