Project Team Member : Ch. Mubashir, Sultan, Waqar.

Course: DS5-1 – Data Warehousing and Business Intelligence.

Project Proposal

Title: Implementation of a dbt in Snowflake for Pakistan's Investment Data Analytics (PSX & Mutual Funds).

1. Introduction

Pakistan's financial market produces massive volumes of data daily — including stock trades from the Pakistan Stock Exchange (PSX) and mutual fund performance reports from major asset management companies. Managing and transforming this data into meaningful business insights requires a robust, scalable, and automated data pipeline.

This project implements a data warehouse pipeline in Snowflake integrated with dbt (Data Build Tool) to automate data modeling and transformation. The project will analyze Pakistan's investment options, including stocks, mutual funds, and market indices, providing actionable insights through Power BI visualizations.

2. Objectives

- Develop an optimized data pipeline for investment tracking using dbt and Snowflake, focusing on Pakistan's financial ecosystem (PSX & Mutual Funds).
- Configure a scalable Snowflake data warehouse with defined roles, warehouses, and schemas for secure and efficient data management.
- Implement a dbt project to automate data modeling, transformation, testing, and documentation.
- Integrate multiple data sources such as PSX datasets, mutual fund NAV data, and Kaggle financial data into a unified data warehouse.
- Design modular SQL + Jinja models to clean, aggregate, and structure investment data for analytical use.
- Automate deployment and validation using dbt commands (dbt run, dbt test) and Snowflake task scheduling.
- Leverage Snowsight Workspace for collaborative model development, testing, and performance monitoring.
- Connect Power BI to Snowflake to create dynamic dashboards for investment insights and performance visualization.

- Enable version control and CI/CD integration for continuous improvement, reproducibility, and rollback management.
- Ensure secure data governance by managing access control, role-based permissions, and monitoring activity logs.

3. Problem Statement

Pakistan's investment data — including stock transactions, daily market indices, and mutual fund NAVs — is dispersed across various platforms such as PSX APIs, fund websites, and financial reports. Manual aggregation and reporting make it difficult for analysts and investors to gain timely insights.

This project addresses that gap by developing an automated Snowflake–dbt–Power BI pipeline that efficiently ingests, transforms, and visualizes financial data for decision-making.

4. Tools and Technologies

- Snowflake – Cloud Data Warehouse
- dbt (Data Build Tool) – Data modeling, transformation, testing, and documentation
- Snowsight Workspace – Integrated development and visualization environment in Snowflake
- Power BI – Business Intelligence and visualization tool for interactive dashboards
- GitHub – Version control and CI/CD integration
- Python (Optional) – For automation and orchestration scripts
- SQL + Jinja – For creating modular and reusable transformation logic
- Data Sources – PSX APIs, Mutual Fund NAV datasets (NBP Funds, MCB Arif Habib, Al Meezan, etc.), and Kaggle datasets

5. Project Workflow / Methodology

1. Create and Configure Snowflake Account

- Set up roles (SYSADMIN, DEVELOPER, ANALYST).
- Create virtual warehouses for compute resources.
- Define databases and schemas (raw, staging, analytics).

2. Initialize dbt Project

- Run dbt init to create the project structure.
- Configure dbt_project.yml and profiles.yml for Snowflake connection.
- Define source tables for PSX and mutual fund data.

3. **Build Data Models**

- Create staging models to clean and standardize PSX and mutual fund data.
- Develop intermediate models for joins, aggregations, and transformations.
- Design mart models for final business-ready datasets.
- Apply dbt macros, tests, and documentation.

4. **Deploy Project on Snowflake**

- Execute transformations using dbt run and validate with dbt test.
- Generate documentation using dbt docs generate.
- Schedule automated tasks using Snowflake CLI or tasks.

5. **Use Snowsight Workspace**

- Edit, test, and version dbt models directly within Snowflake.
- Visualize query results and monitor performance.
- Collaborate on changes and manage model history.

6. **Enable Version Control and CI/CD**

- Integrate dbt project with GitHub for version control.
- Automate testing and deployment using pipelines.
- Support rollback and maintain version traceability.

7. **Power BI Visualization**

- Connect Power BI to Snowflake data models.
- Build dashboards showing PSX index trends, mutual fund performance, and sector analysis.
- Provide real-time insights for investment decisions.

8. **Monitor and Manage Access**

- Configure role-based access control (RBAC).
- Set up query monitoring and logging.
- Ensure secure and compliant data operations.

6. Expected Outcomes

- Automated dbt project deployed on Snowflake.
- Centralized and analytics-ready data warehouse for PSX & Mutual Funds.
- Clean, tested, and documented data models.
- Interactive Power BI dashboards for market and fund analysis.

- Scalable, version-controlled, and auditable data engineering pipeline.

7. Data Sources

Data for this project will be collected from the Pakistan Stock Exchange (PSX), publicly available Mutual Fund NAV datasets (NBP Funds, Al Meezan, UBL Funds, etc.), and verified Kaggle datasets containing financial indicators for Pakistan. This ensures the project reflects real investment data for meaningful analysis.

8. Reference

YouTube Reference: https://youtu.be/OLXkGB7krGo?si=PJVOMW0k_gzrzy40

9. Conclusion

This project demonstrates a real-world data warehousing and BI implementation using Snowflake, dbt, and Power BI tailored for Pakistan's investment ecosystem. It highlights how modern data engineering practices can optimize financial data analysis, enable data-driven decision-making, and deliver real-time insights into the stock and mutual fund markets.