

Snowflake CI/CD POC



Background

- Snowflake is increasingly being used for Transformation and Data warehousing. But it doesn't offer any out-of-the box CI/CD solutions.
- Datalytx is being evaluated for DataOps (CI/CD, Orchestration) platform
- Extraction and Loading are done through Azure Data Factory and Transformations are carried out in Snowflake via Streams, Tasks and Stored Procedure

Problem

- Azure DevOps does not support deployment Snowflake native features (Tasks & Stored Procedures)
- Knowledge in JavaScript is required to write stored procedure in snowflake
- Avoid rework if any DataOps tool is selected for Snowflake Projects

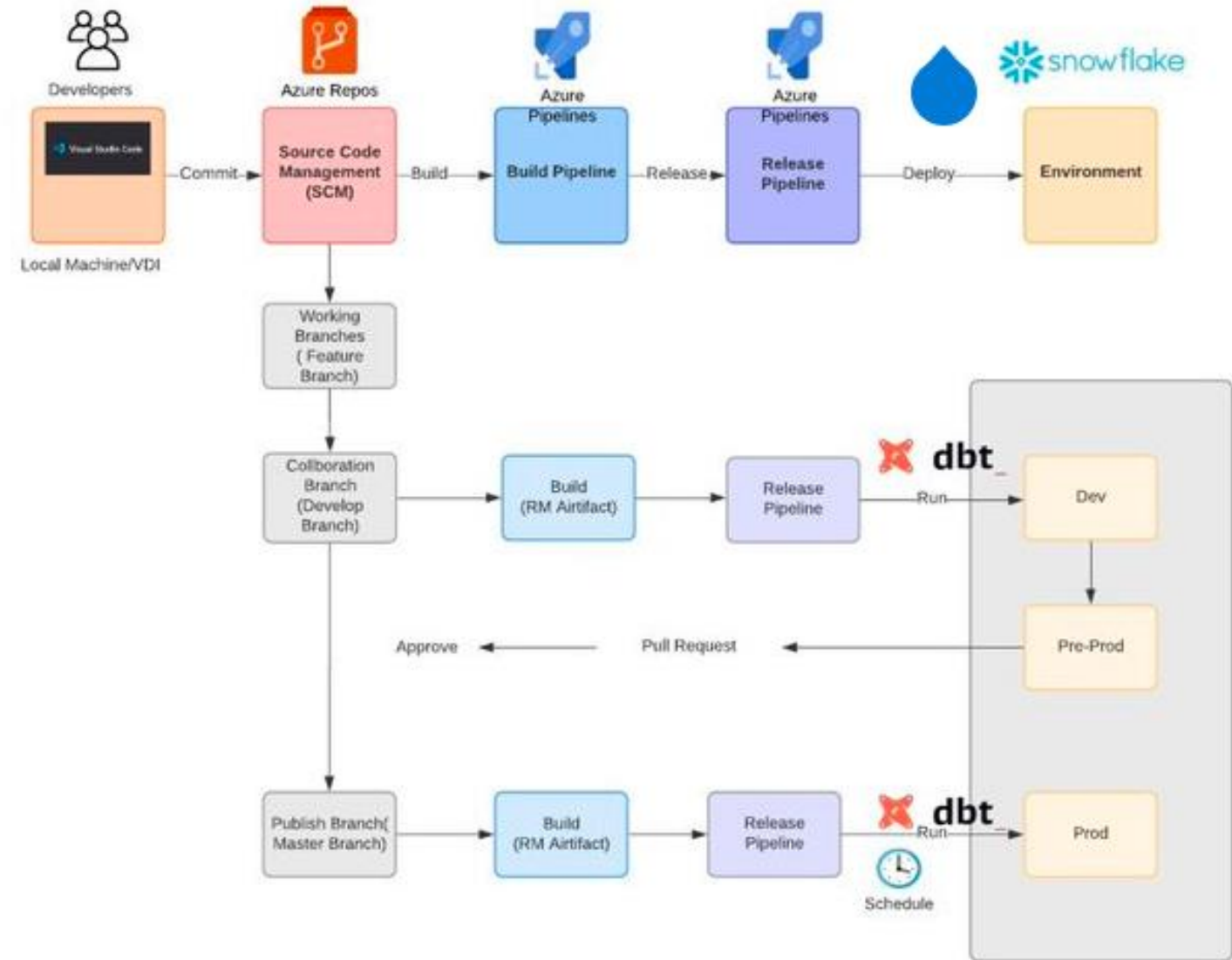
Solution

- DBT – a transformation tool is used to execute data transformations in Snowflake
- DBT is integrated with Azure DevOps & Snowflake Deployments are automated
- DBT - Models are created using Jinja template which are executed in Snowflake

Benefits

- Able to perform continuous integration / Continuous delivery for Snowflake projects
- DBT Models are reusable and can be run against any Cloud data warehousing tools with minimal changes
- Business analysts and data scientists tend to speak in SQL and it's much easier to collaborate with them. Anyone with SQL and basic knowledge of Python can easily develop solution using DBT
- Data Lineage for Pipeline can be easily generated to understand which data sources are involved in a certain transformation and the flow of the data from source to target.
- All the computational work is pushed towards DW(Snowflake)
- Deploy Analytics code faster with software engineering best practices (Environment , Package management and Continuous integration)

Architecture



Next Steps

Enhance this Solution with Airflow for Data pipeline orchestration and Monitoring