Ensuring trusted data in Snowflake Data Vault for an asset management firm

ETL Validation, Transformation Logic Validation, & Integration with Azure & Snowflake

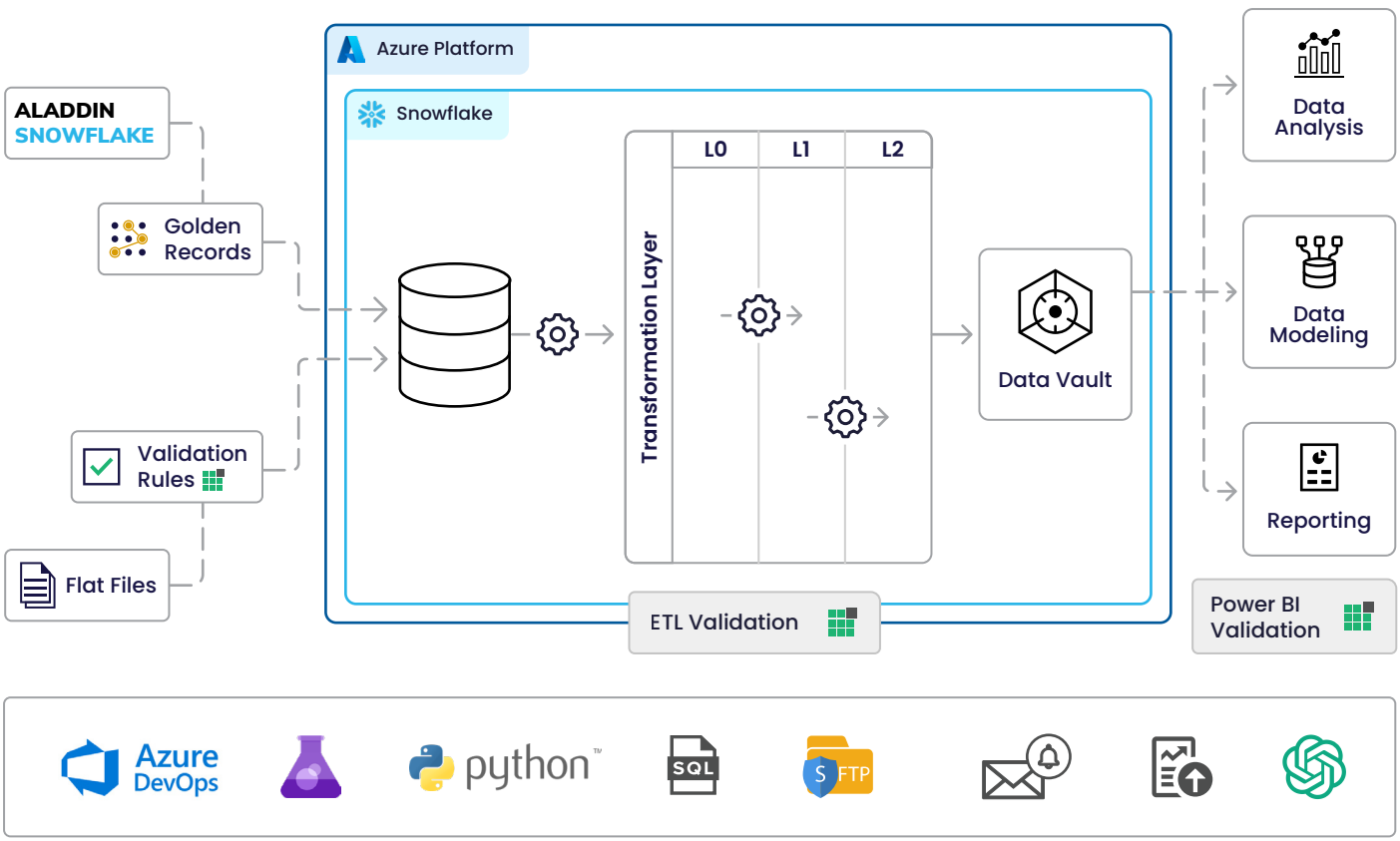
DATAGAPS
THE GREAT DATA COMPANY

Customer Objective

The customer handles a multitude of asset-related datasets (trading and securities) that undergo multiple transformations, which are subject to iterations and changes on a frequent basis. These transformation logic changes and their cascading effects on the subsequent layers have to be validated to ensure that only the intended effects are applied, and no unintentional changes have been implied on related datasets. An additional use-case is also to perform file-level validations in the data lake. The customer also had a major use case of running 10+ suites daily, which have to be validated daily in an automated fashion, as manual testing is too resource-intensive. Finally, the reports created with the transformed datasets also have to be validated against datasets, other reports, and in terms of security and performance.

> “Without confidence in our transformation logic implementation and data integrity, our business basically comes to a standstill. Failure to achieve high-quality data could also result in potential fines by the FCA”

Furthermore, the customer needed a solution to this problem statement that integrates seamlessly with Snowflake and Azure-based pipeline, which includes Azure Test Plans & Azure Dev Ops.



Business Challenges

- ETL Transformation & Logic Validation
- 10+ Test Suite Monitoring
- Data Quality Monitoring of Flat Files
- User-defined business rules
- Flat File Naming-based Validation
- Integration with Snowflake
- Integration with Azure Dev Ops & Azure Test Plans
- Power BI Report Validation

Solutions

- The customer implemented DataOps Suite to help with the multiple objectives with ETL Validation and Data Validation via Data Rules as the main pillars.
- The transformation/logic updates first go through the Suite. The transformation is first deployed in the Suite using the "Code" component that can deploy SQL, Python, Scala, and DB Scripts, and then tested against rules on the validation and effect of downstream and related datasets.
- The customer used the DataOps Suite's Query Builder for easy query creation in conjunction with ADF pipeline functions.
- The customer uses the Suite's functionality to validate their Azure Data Factory functions and transformation logic.
- The 10+ Testing Suites, which are run on a daily basis, were shifted from a manual system to an automated validation system using the Suite's tools.
- Data Quality Rules were applied to the flat files received. The rules are applied to the complete dataset instead of a subset.
- Parameterized Business Rules were created using the low-code / no-code rules system, which was deployed using the import-export functionality.
- The Suite is seamlessly integrated with Azure and Snowflake, with the features of the platforms communicating the suite right out of the box.
- For Power BI reports, BI Validator was deployed to validate reports against the snowflake datasets and other reports. BI Validator is also used to perform Performance and Security monitoring & validation.
- The notification systems and automatic report generation keep track of all the issues the pipeline identifies, which the user accesses retroactively for final touchups.

Technologies

snowflake · Azure DevOps · Jira Software · ALADDIN SNOWFLAKE · SQL

> "DataOps Suite has greatly improved our confidence in our datasets and transformation logic while reducing the time taken to test and subsequently reducing the time-to-deploy of updates and hotfixes. The Suite's integration options with Snowflake and Azure make test cases and validations very easy to deploy, maintain, and manage."
>
> *— QA Manager*

Benefits Realized

Reduce time to run manual daily monitoring suite by **~75%**	**100%** Data Validation achieved with increased overall data quality	Power BI Datasets & Presentation Layers validated
Direct Integration with Snowflake and Azure, results in zero disruption to existing pipelines	**100%** ETL Transformation Validation and Logic Checks applied to all updated transformations	Reduction in Testing time leading to faster implementation and time to deploy