leaplogic

Automated ETL workload transformation from Ab Initio to Google Cloud

VIDEO TRANSCRIPT

The freedom of the cloud is on the horizon, but the path to modernization is challenging.

We change that!

LeapLogic can accelerate the migration and modernization of legacy data warehouse, ETL, Hadoop, analytics, and reporting systems to cloud-native stacks with up to 95% automation.

Let's take a deep dive into LeapLogic's end-to-end automated transformation that simplifies the migration of legacy workloads to modern platforms – starting with a comprehensive assessment, followed by automated transformation, validation, and all the way up to operationalization.

In this video, we will showcase how LeapLogic can automate workload modernization by considering an example of transforming Ab Initio scripts to a Google Cloud-native environment using PySpark or DBT.

Prerequisite

Before diving into the assessment or transformation of workloads, there are crucial prerequisites to address.

These involve extracting the right set of extensions for LeapLogic, such as KSH, XFR, or DML files, from the Ab Initio graph. This can be achieved either through the command line, using shell scripts with native Ab Initio commands, or with the help of the GDE UI, allowing the extraction of files containing business logic.

The exported KSH files are then configured onto an assessment, setting the stage for further analysis.

Moving forward, our focus shifts to the DBT input file, containing native Ab Initio code. The file outlines steps like reformat and replicate, enabling us to understand and

analyze the components to make informed decisions for the next stages.

Assessment

Now, let's explore how LeapLogic assesses the source code.

In this example, the extracted Ab Initio files serve as inputs, and our Subject Matter Experts (SMEs) add configurations with use case-driven names before executing the assessment.

A previously executed assessment by LeapLogic shows that all components are supported, resulting in a 100% analysis. Complexity matrices categorize the overall complexity of all files, helping us understand automation percentages and potential manual interventions.

The assessment analysis covers input and output sources, expected components, and LeapLogic's automated conversion percentage. During the comprehensive assessment, SMEs identify patterns, unrecognized elements, and repetitive patterns for fine-tuning before the transformation phase.

LeapLogic also offers in-depth assessment reports, providing a data-driven dive into business logic. These reports encompass code-level and schema-level optimization recommendations and future uploads of orchestration scripts, allowing for workflow-level optimization suggestions.

The Lineage section further addresses scenarios like single jobs with input and output, common inputs in multiple jobs, outputs from one job as inputs for another, and dependencies between multiple jobs or objects. This helps prioritize independent workloads, aiding in phased and iterative capacity planning. Subsequently, a migration plan is developed with fixed durations, allowing for a clear understanding of migration costs and delivery timelines.

The assessment delivers valuable insights regarding inventory, complexity, dependency, and lineage. Control flow lineage aids in capacity planning, and technical debt is explored through active versus static inventory analysis.

Transformation

As we move to the transformation stage with LeapLogic, our understanding of dependent objects and workloads guides the initial steps.

In this example of Ab Initio to Google Cloud transformation, the first phase involves migrating schemas, replicating the schema, and migrating test data to the target platform. This ensures post-conversion objects are available for validation, allowing us to reconcile data and synchronize source and target platform objects.

The ETL conversion process first involves configuring Ab Initio as a source and selecting the target platform, whether DBT or PySpark.

To fast-track things, let’s have a look at one of our pre-created pipelines. Ideally, this graph will be converted into DBT-native scripts while retaining the business logic.

Looking closer at the source file, we identify components like reformat, replicate, and sequels, including Delete statements and procedure calls.

To explore how this code appears when transitioning to a DBT environment, we navigate to the target state folders to observe the expected structure. DBT-native structures have distinct left-hand and right-hand components. As a reference table, we focus on logic related to processed flags, replication, and updates on the table. The Directed Acyclic Graph (DAG) transforms into a model, with DBT supporting pre-hooks and post-hooks for executing pre-sequels and post-sequels.

We can also examine the LeapLogic migration output here to check how many components are automatically transformed. Here, typical folder structures supported by DBT are also showcased.

This encapsulates the packaged code generated by LeapLogic, ready for execution on the target platform. We can go to DBT Cloud to run and perform further unit validation on the code. Additionally, LeapLogic supports function validation, offering advanced steps for reconciling processed data. This provides a comprehensive view of how the target code will appear.

As we look closer at DBT's capabilities, it's crucial to recognize that DBT is a powerful transformation tool, supporting scalable analytics workflows. However, its limitations arise when dealing with external scripts, API invocations, or JDBC connections. In such cases, LeapLogic seamlessly integrate PySpark, which can run on Google Cloud-native services like Dataproc.
The strategic move allows the execution of logic within the Google Cloud environment.

Configuring PySpark results in a distinct package showcasing a modified folder structure. Here, we shift away from DBT-native and consider the execution of PySpark code on Google Cloud.

In the job section, we encounter multiple stages in Ab Initio. Examining the input graph, we notice a unique structure containing certain native components that can’t be directly converted into DBT.

As we visualize the graph, it becomes evident that multiple components make calls to an external source JDBC connection, subsequently inserting data into a database within the environment. In this scenario, a Python-based Spark or BigQuery code can simulate business logic. The workflows intricately stitch together each stage, replicating Ab Initio's stages onto PySpark-equivalent components, maintaining consistent naming conventions.

As we look at the output, we observe that these stages have been successfully replicated onto equivalent PySpark components. This configuration can be adapted to run in Python, allowing the flexibility to push the logic onto Spark.
This adaptability is crucial in deciding the solution architecture on the target side.

Here, we can see that LeapLogic offers diverse solutions for end-to-end workload transformation. With LeapLogic, we can transform legacy workloads to both DBT and PySpark-equivalent, seamlessly integrating with Google Cloud for efficient execution.

Conclusion

In this example of Ab Initio to Google Cloud transformation, we can see that regardless of the chosen target code—LeapLogic plays a crucial role in automating, accelerating, and packaging the code for deployment on the selected platform or tool.

LeapLogic automates rigorous validation tests and handles orchestration for the end-to-end transformation of legacy workloads to cloud-native stacks such as AWS, Azure, Google Cloud, Databricks, and Snowflake without any business disruption.

Not just that – LeapLogic also assists with target-specific optimization and capacity planning, ensuring the optimized performance of transformed workloads in a modern stack.

Explore LeapLogic's automation capabilities for the end-to-end transformation of data warehouse, ETL, Hadoop, analytics, and reporting systems to cloud-native stacks – faster, at a lower cost, and with minimal risk. It's more than the next step. It's a leap into the future of your business.