LTIMindtree

Accelerate Migration from SAS to PySpark on Dataproc with Scintilla

An automated migration solution for your data platform modernization journey.

Challenges with legacy SAS

In today's data-driven landscape, organizations are facing mounting pressure to extract meaningful insights from ever-growing datasets. However, legacy analytics solutions like SAS, while powerful, often present significant challenges:

High upfront costs: Traditional SAS licensing models require significant upfront investments.

Limited scalability: SAS environments can face difficulties scaling efficiently to handle massive data volumes or complex analytical tasks.

Closed ecosystem: The closed SAS ecosystem restricts integration with modern, open-source technologies and tools, hindering agility and innovation.

As a result, organizations are seeking alternatives like PySpark.

Why migrate to PySpark on Dataproc?

- **Cost reduction:** Dataproc offers a pay-as-you-go model, reducing infrastructure and licensing costs compared to traditional SAS deployments.

- **Modern ecosystem:** Python's vast data science libraries and community provide more flexibility and advanced analytics capabilities.

- **Scalability and performance:** PySpark leverages distributed computing on Dataproc, enabling faster processing of large datasets than SAS.

- **Cloud integration:** Seamlessly integrate with other GCP services like BigQuery, Cloud Storage, and Vertex AI for a comprehensive data platform.

- **Open-Source advantage:** PySpark' s open-source nature fosters collaboration and customization and prevents vendor lock-in.

Navigating the complex path to Modernization

Migrating large SAS codebases to PySpark is **highly complex and time-consuming,** particularly for advanced analytics.

SAS is procedural, while PySpark uses distributed computing and functional programming, requiring a **shift in both mindset and approach.**

Market players offering SAS to PySpark conversion prioritize a **tactical approach,** which overlooks thorough testing and having the right architecture fundamentals.

Introducing Scintilla

LTIMindtree accelerates your SAS to PySpark migration journey with Scintilla, a robust solution that automates 60-80% of your SAS ETL code conversion. Combined with LTIMindtree's expertise, large-scale SAS modernization programs are seamlessly executed, significantly reducing risk.

LTIMindtree value proposition using Scintilla

- **Complete ownership:** Comprehensive migration service offering with full lifecycle support.

- **Automation:** High automation coverage ratio (60% to 80%).

- **Accurate:** Less than 10% manual intervention for converted code.

- **Transparent:** 100% open standards-based.

- **Proven:** Multiple engagements with leading global customers.

- **Flexible approach:** Start with a pilot, prove value, and calibrate the tool as appropriate.

- **Continuously improved:** Continuously enhanced through multiple engagement experiences.

- **Portable:** Containerized code base with ease of deployment.

Key components

- **Mapping dictionary:** A dictionary of SAS to PySpark equivalent keywords and datatype.

- **Function library:** A library of PySpark equivalent statements or commands.

- **Pattern repository:** Repository with a catalog of SAS patterns supported for conversion.

- **Conversion engine:** The core engine for conversion leveraging the mapping dictionary and function library.

- **Lineage and logging:** Process log of the conversion including conversion percentage, exceptions, and unconverted code blocks.

Our Success Story

SAS to PySpark modernization for a leading healthcare technology company

The healthcare tech provider aimed to update its outdated on-prem data system using SAS, which has over 6000+ ETL scripts, to a modern setup on GCP. Despite starting its modernization process, it hasn't met the expected outcomes in terms of time, effort, and quality.

LTIMindtree, with its experience in delivering SAS Platform modernization solutions and its mindful automation approach leveraging Scintilla, was best positioned to solve these challenges. We proposed viable solutions for the target state architecture and designed the overall migration program. LTIMindtree 's distinctive migration framework effectively leveraged Scintilla-led automation to expedite the SAS code analysis and Pyspark conversion process, reduce risks and manual efforts, and increase the predictability of the entire migration program.

The results were impressive. LTIMindtree achieved **~ 75% automation** through Scintilla and not only significantly **accelerated the migration** but also **delivered thoroughly tested and performant Pyspark ETLs.** Given its success, LTIMindtree is currently migrating all the 6000+ scripts for the customer.

Engagement journey with LTIMindtree

Execute a PoC*: Observe Scintilla's automation and accuracy in action through a quick PoC to convert a set of your selected SAS scripts to PySpark.

Productionize: Create a migration roadmap and strategy and begin migrating SAS workloads of a selected domain into production, in a phased manner.

Scale: Leverage LTIMindtree's T-shirt-sized pricing and factory model to migrate from SAS to GCP completely, for all SAS workloads across domains.

*Leverage LTIMindtree's Quick Start PoC Package, developed in partnership with Google, for a fast and affordable PoC.

How to get started?

Reach out to Data.gcpcoe@ltimindtree.com to get started with Scintilla.

LTIMindtree