

Intelligent and automated transformation of SAS analytic workloads to PySpark

Offloading SAS workloads to a distributed processing paradigm can help enterprises cut costs while establishing a vendor-agnostic, easily scalable analytics platform. However, migrating SAS workloads, including procedures, reports, and datasets, to PySpark on the cloud or on-premises is complex, as the scripts contain thousands of conditional statements (logic), macros, statistical algorithms, advanced SAS algorithms, procedural statements, and outputs. It is also difficult to ascertain the accuracy and precision of migrated SAS analytical models.

Migrate SAS workloads to PySpark with LeapLogic

LeapLogic, Impetus' automated cloud accelerator, streamlines and de-risks the complete transformation of SAS workloads to Achieving up to 95% automation in the initial iteration, LeapLogic facilitates end-to-end assessment, transformation, validation, and operationalization of workloads. Its intelligent, pattern-based transformation engine effortlessly manages diverse code patterns, constructs, dialects, functions, etc., converting them into a cloud-native equivalent for enhanced compatibility and performance.

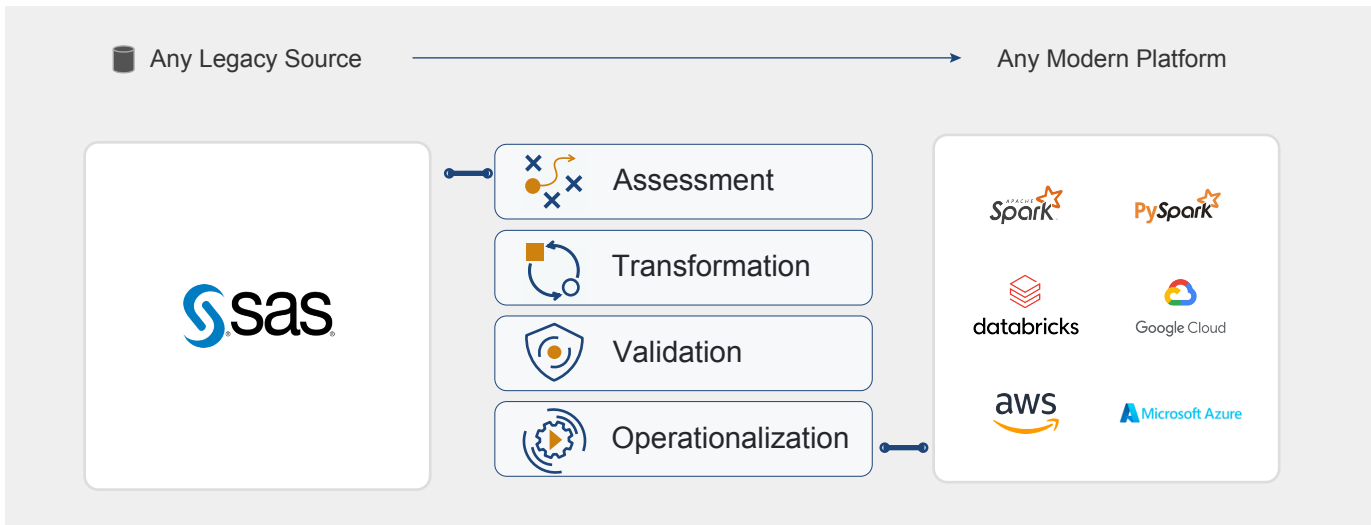
Key Benefits

Compared to traditional approaches, LeapLogic enables:

- **4x** faster transformation
- **1.5x** faster validation
- **2x** cheaper than manual migration
- **50%** migration acceleration
- **100%** risk compliance
- **100%** SLA adherence

Watch how LeapLogic automates the transformation of:

- ▶ SAS to PySpark
- ▶ SAS to Amazon EMR
- ▶ SAS to Azure HDInsight
- ▶ SAS to Databricks
- ▶ SAS to Snowpark



Step 1: Assessment

- Lists entire inventory to provide actionable insights for disparate workloads
- Architecture and tech stack components
- Assesses usage patterns to categorize workloads into three major buckets

ETL usage	SAS statistical	Advanced SAS algorithms
<ul style="list-style-type: none"> • SQL operations + SAS procedural • Operations like pyscan, pysum, upcase, intnx, input, put, intck, ifn, catx, cats, lowercase, pymeans, etc. 	<ul style="list-style-type: none"> • Base operations + SAS procedural • Operations like proc_append, proc_datasets, process_dataset, hash_table, proc_contents, proc_tabulate, union-AllDatasets, proc_stdize, proc_summary, calculate_stats_with_func, calculate_Stats_For_Type_1ToN_GroupByCols, NULLIFZERO, proc_transpose, log_message, etc. 	<ul style="list-style-type: none"> • Base operations + SAS procedural • Operations like proc_corr, proc_princomp, proc_surveyselect, proc_discrim, proc_glmmod, proc_princomp, proc_reg, etc.

Step 2: Transformation

- Transforms SAS scripts and migrates schema and dataset to PySpark or any other cloud-native variant such as Amazon EMR, Azure HDInsight, GCP Dataproc, Databricks stack, etc.
- End-to-end packaging
 - Transforms core business logic to cloud-native wrappers or orchestrators
 - Repackages to production-ready jobs on target
 - Ensures end-to-end execution on the production environment after system integration testing
- Extensible tool and methodology
 - Converts code to a variety of formats

Accurate and precise transformation of SAS analytical models on the target

- SAS programs → Python programs
- Macros → Python functions
- PROC SQLs → Spark SQL
- Functions / PROCs → PySpark / Python common functions in reusable library
- SYMPUT and SYM → Spark constructs
- SAS inbuilt functions → Python inbuilt and custom functions in reusable library

LeapLogic generates production-ready code and validates and optimizes workloads for the target platform.

Step 3: Validation

- Pipeline-based automated validation
- Data-driven code validation
- Automated SQL/query and data-level validation
- Cell-to-cell validation reports
- Data type and entity-level matching
- Auto-generation of reconciliation scripts

Designed for cloud transformation, LeapLogic overcomes legacy limitations to efficiently assess, migrate, validate, and operationalize complex SAS analytics workloads to PySpark or any other cloud-native variant such as Amazon EMR, Azure HDInsight, Google Cloud Dataproc, Databricks stack, etc. Fueled by an intelligent grammar engine, it

Step 4: Operationalization

- Productionization and go-live
- Infrastructure as code
- Execution using cloud-native orchestrators
- Automated DevOps, including CI/CD
- Target environment stabilization
- Smooth cut-over

accelerates cloud modernization, unlocking advanced analytics and supporting long-term business goals.

Enterprise success stories

F500 payment card company

- **650+ SAS scripts** migrated to PySpark

Leading data analytics company

- **4982 SAS scripts** with 1.8 million lines of code converted to PySpark
- **60%+ auto-conversion** of SAS workloads to Azure Databricks

To automate SAS analytics workloads to PySpark,
write to us at info@leaplogic.io

BOOK A DEMO 

leaplogic

LeapLogic, an Impetus product, automates the end-to-end transformation of legacy data warehouse, ETL, BI/reporting, analytics, Mainframe, and Hadoop workloads to native cloud platforms for optimum cost and performance. Impetus Technologies helps enterprises solve the data, AI, and cloud puzzle by combining unmatched cloud and data engineering expertise. Recognized as an AWS Advanced Consulting Partner, Elite Databricks Consulting Partner, and Elite Snowflake Services Partner, Impetus has been the 'Partner of Choice' for several Fortune 500 enterprises for transforming their digital nuclei and driving unmatched innovation and growth for over a decade.

To learn more, visit www.leaplogic.io or write to inquiry@impetus.com.

© 2024 Impetus Technologies, Inc. All rights reserved. Product and company names mentioned herein may be trademarks of their respective companies. MAR 2024