



Modernize SAS to PySpark with SAS2PY

SAS2PY: Automated Migration from SAS to PySpark

Transform your legacy SAS environment into a modern, PySpark-based data ecosystem with SAS2PY. Our platform ensures a smooth transition, enhanced by comprehensive data lineage tracking and AI-assisted development tools.

- **Automated Code Conversion:** Migrate extensive SAS codebases swiftly and accurately.
- **Data Lineage & Flow Tracking:** Maintain complete visibility of data transformations and dependencies.
- **Data Matching:** Ensure data integrity and consistency throughout the migration process.
- **Merlin AI Developer Assistant:** Enhance code quality and development efficiency with AI-driven insights.

Comprehensive Migration Capabilities

SAS2PY offers end-to-end migration solutions with unparalleled flexibility:

- **Automated Workflows:** Convert SAS, SQL, and ETL processes into PySpark workflows without manual intervention.
- **Scalability:** Effortlessly manage complex migrations involving millions of lines of code and large datasets.
- **Data Lineage & Flow Tracking:** Gain end-to-end visibility into your migration journey with detailed lineage tracking, ensuring complete transparency.
- **AI/ML Support:** Enable seamless machine learning workflows with PySpark's MLlib integration.





Automated Migration Process

Legacy Analysis

- **Component Identification:** Automatically detect SAS scripts, SQL queries, and ETL workflows.
- **Migration Roadmap:** Develop actionable insights to streamline the migration process.
- **Workload Profiling:** Prioritize high-impact components for efficient migration.

Code Conversion

- **Script Transformation:** Convert SAS and SQL scripts into optimized PySpark workflows.
- **Schema Adaptation:** Align legacy schemas with PySpark-compatible formats, ensuring seamless integration.
- **ETL Modernization:** Revamp ETL pipelines to leverage PySpark's distributed processing capabilities.

AI-Powered Validation & Testing

- **Generative AI Validation:** Utilize AI to ensure operational parity between legacy and new workflows.
- **Automated Regression Testing:** Confirm consistency and accuracy across systems.
- **Performance Optimization:** Refine workflows for optimal performance in production environments.

Data Matching and Validation

Ensuring data integrity during migration is paramount. SAS2PY provides comprehensive validation to eliminate inconsistencies:

- **Schema-Level Matching:** Map source schemas to PySpark architectures, validating data structures and constraints.
- **Row-Level Validation:** Compare individual records using efficient matching techniques, resolving discrepancies automatically.
- **Aggregate Data Matching:** Validate metrics like counts and averages to ensure parity between source and target datasets.



Merlin: AI Developer Assistant

Merlin AI enhances development efficiency throughout the migration process:

- **Real-Time Code Insights:** Provide immediate feedback and best practice recommendations.
- **Advanced Debugging:** Streamline troubleshooting to minimize errors and accelerate development.
- **Optimization Suggestions:** Enhance code performance and maintainability with AI-driven advice.

Proven Results

SAS2PY delivers measurable benefits:

- **Cost Reduction:** Achieve up to 70% savings by automating migration processes.
- **Accelerated Timelines:** Complete migrations up to 10 times faster, reducing project durations from months to weeks.
- **Enhanced Scalability:** Efficiently manage enterprise-scale migrations involving extensive codebases and datasets.
- **Data Accuracy:** Ensure data integrity with advanced validation tools and AI-powered testing.

Frequently Asked Questions

- 1. How does SAS2PY handle ETL migrations to PySpark?** SAS2PY automates the transformation of legacy ETL workflows into PySpark scripts, optimizing for distributed data processing and seamless integration.
- 2. Can SAS2PY migrate machine learning models to PySpark?** Yes, SAS2PY converts legacy models into formats compatible with PySpark's MLlib, facilitating end-to-end machine learning workflows.
- 3. What advantages does SAS2PY offer over other migration tools?** SAS2PY provides unparalleled automation, speed, and accuracy, reducing costs by up to 70% and accelerating migration timelines by a factor of 10. Its comprehensive features, including data lineage tracking and AI-assisted development, ensure a seamless and efficient transition.