

SAS · TALEND · ALTERYX · DATASTAGE ·  
INFORMATICA · ODI · SSIS · TERADATA →  
DATABRICKS

# Land Every Legacy ETL Workload on the Lakehouse.

MigryX converts SAS, Talend, Alteryx, IBM DataStage, Informatica, Oracle ODI, SSIS, Teradata, and SQL dialects directly to **PySpark, Delta Lake, Unity Catalog, DLT pipelines, and Databricks Workflows** — with 95%+ parser accuracy and 99% with AI augmentation.

99%

AI ACCURACY

60%+

COST SAVINGS

10×

FASTER

8+

SOURCE ENGINES

[Schedule a Lakehouse Demo →](#)

[Email us →](#)

```
silver_claims.py · DLT

# Source: SAS DI Studio job
# Target: Databricks DLT (Silver)
import dlt
from pyspark.sql import functions as F

@dlt.table(
    name="silver_claims",
    comment="Cleansed claims"
)
@dlt.expect_or_drop("valid_id", "id IS NOT NULL")
def silver_claims():
    return dlt.read("bronze_claims")
```

Conversion Accuracy **99.2%**

## THE REALITY CHECK

# Manual Lakehouse migrations stall. Generic transpilers approximate.

Every legacy ETL platform — SAS, Talend, Alteryx, DataStage, Informatica, ODI, SSIS, Teradata — speaks a different dialect. Manual rewrites take years. Generic scanners miss macros, implicit joins, and column-level lineage. Both leave production at risk.

## The True Cost of a Stalled Migration

Legacy ETL licenses, rewrite consultants, and tribal-knowledge risk compound monthly. Most "lift-and-shift" projects miss go-live dates by 12-18 months — and arrive on Databricks with broken lineage, no data contracts, and reverse-engineered logic nobody trusts.

**\$2M+**

ANNUAL LEGACY SPEND

**3-5yr**

MANUAL REWRITE TIMELINE



### Manual Migration to Databricks

- ✗ 3-5 years of consultant-led rewrites burning budget
- ✗ SAS macros, DataStage XML, and SSIS .dtsx logic lost in translation
- ✗ No column-level lineage — Unity Catalog launched empty
- ✗ Pipelines land as raw notebooks — no Bronze/Silver/Gold structure
- ✗ No Asset Bundles, no DLT, no Workflows — just code dumps
- ✗ Validation is manual — production incidents discovered post go-live



### MigryX → Databricks

- ✓ Custom AST parsers per source — 95%+ deterministic, 99% with AI
- ✓ Output organized into Bronze · Silver · Gold Medallion layers
- ✓ Delta Lake with MERGE, OPTIMIZE, Z-ORDER, and Liquid Clustering
- ✓ DLT pipelines, Workflows via Asset Bundles (DABs), Photon-optimized SQL
- ✓ Column-level lineage and STTM published into Unity Catalog
- ✓ Row-level data matching — audit-ready evidence for sign-off

## 5-STAGE METHODOLOGY

# From legacy artifact to production Lakehouse — in five proven steps.

1

### Ingest

SAS scripts, Talend exports, DataStage XML, .dtsx packages, ODI repositories.

2

### Parse + Lineage

Custom parsers build ASTs, expand macros, produce column-level lineage.

3

### Convert

Parser-driven output to PySpark, Delta Lake, Workflows, or DLT — fully documented.

4

### Validate

Row-level + aggregate matching between legacy and Databricks — audit-ready.

5

### Govern

Lineage, STTM, contracts published into Unity Catalog. Merlin surfaces risk.

# Not just code conversion.

## Production-ready Databricks architecture.

Every MigryX migration is engineered for the complete Databricks Lakehouse — Medallion Architecture, Photon-optimized SQL, Unity Catalog governance, Delta Lake storage, and Asset Bundle deployment. Production-ready from day one.

### MEDALLION ARCHITECTURE · AUTO-GENERATED

#### BRONZE

##### Raw Ingestion

Auto Loader-based ingestion preserves source fidelity. Schema inference, file notification mode, and incremental processing built in.

#### SILVER

##### Cleansed & Conformed

Deduplication, type enforcement, business-key resolution, and DLT data-quality expectations from your legacy validation logic.

#### GOLD

##### Business-Ready Aggregates

Aggregated Delta tables with MERGE INTO upserts, OPTIMIZE & Z-ORDER compaction, and Change Data Feed enabled.

#### ● Compute & SQL

Photon Engine

Serverless SQL Warehouses

Job Clusters

ANSI SQL Compliance

500+ Function Mappings

#### ● Orchestration

Databricks Workflows

Asset Bundles (DABs)

Multi-Task DAGs

DLT Pipelines

@dlt.table · @dlt.expect

#### ● Storage & Tables

Delta Lake

MERGE INTO

Z-ORDER

Liquid Clustering

Change Data Feed

Time Travel

#### ● Governance

Unity Catalog

Column-Level Lineage

STTM Publishing

Attribute Tags

Data Contracts

Lineage API

#### ● ML & AI

MLflow Tracking

Model Registry

Feature Store

AutoML Baselines

Model Serving



#### Precision Parsing

Custom ASTs per source. SAS macro expansion, DataStage XML, Talend .item, SSIS .dtsx — full fidelity.



#### AI Augmentation

Merlin AI resolves ambiguity, infers business intent — pushes accuracy to 99%.



#### Unity Catalog Native

Lineage, STTM, and data contracts published directly to Unity Catalog — fine-grained access included.

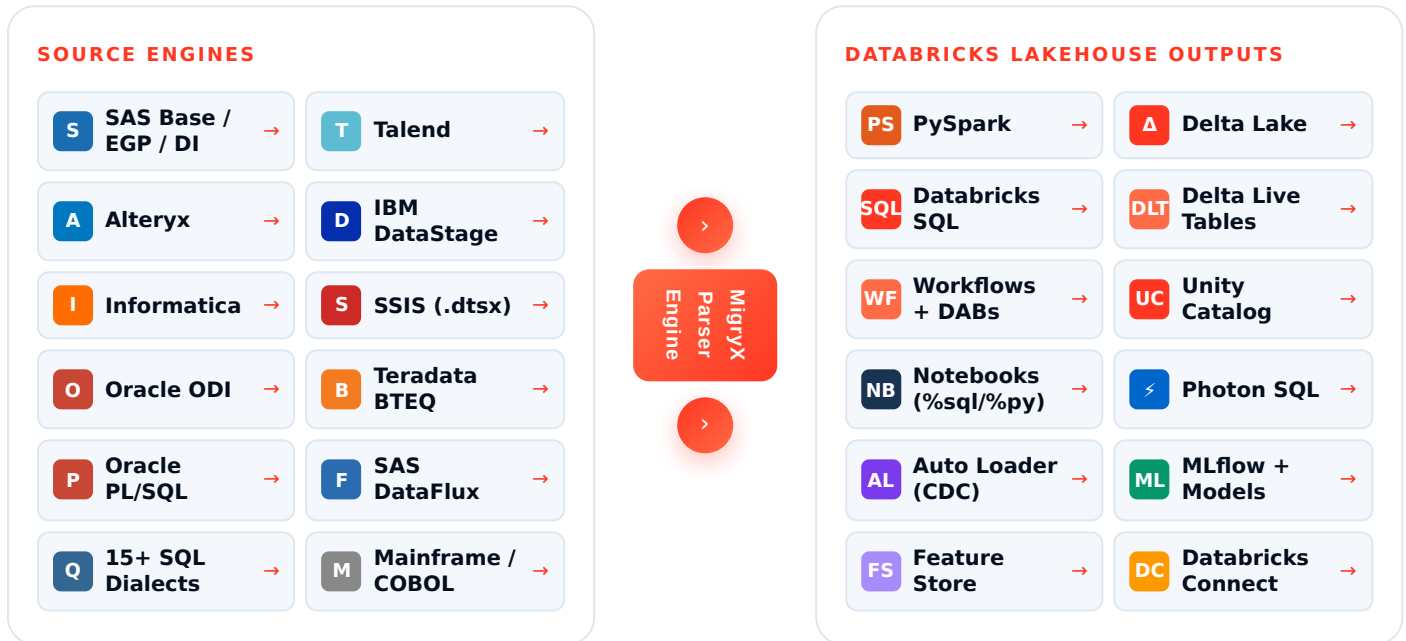


#### Air-Gapped Capable

Runs entirely in your network. Zero external calls. Deploy in <1 hour.

# Every legacy ETL source. One Databricks Lakehouse target.

Purpose-built parsers for each source language. SAS macro expansion, DataStage XML, Talend .item files, SSIS .dtsx — full fidelity, deterministic output, no approximation. The same proven methodology applies to every source.



SOURCE → DATABRICKS	PARSER ACCURACY	+ AI AUGMENTED	LINEAGE DEPTH
SAS (Base / Macro / PROC SQL)	95%+	99%	Column-level
Talend (.item / Std Jobs / tMap)	95%+	99%	Column-level
IBM DataStage (DSX / ISX)	95%+	99%	Column-level
Informatica PowerCenter / IICS	95%+	99%	Column-level
SSIS (.dtsx / .ispac)	95%+	99%	Column-level
Teradata BTEQ / FastLoad / TPump	95%+	99%	Column-level

# Custom parsers get you to 95%. Merlin AI gets you to 99%.

Generic ETL scanners approximate lineage. MigryX parses it exactly — every macro, every column, every dialect — then lands it natively on Databricks. Merlin AI resolves what deterministic parsers cannot.

## Why Merlin lands cleaner code on Databricks

Merlin understands legacy code semantics — not just syntax. It resolves SAS macro variables, DataStage runtime parameters, and Informatica session expressions that line-by-line transpilers silently mistranslate.



### Macro & Parameter Resolution

Resolves SAS macros, DataStage job parameters, Informatica \$\$ variables, and ODI substitution APIs into deterministic PySpark.



### Medallion-Aware Conversion

Recognizes ingestion vs cleansing vs aggregation logic and routes output into Bronze, Silver, or Gold layer notebooks automatically.



### Auto-Generated Documentation

Creates STTM mappings, data contracts, and inline lineage comments — published into Unity Catalog with zero manual effort.



### Optimization Recommendations

Surfaces Z-ORDER candidates, Liquid Clustering opportunities, and Photon-suitable queries from the parsed estate.

```
gold_claims_dlt.py

# — INPUT: SAS PROC SQL —————
# PROC SQL;
#   CREATE TABLE claims_summary AS
#   SELECT region, SUM(amt) FROM
#   &env_schema..claims GROUP BY 1;

# — MERLIN OUTPUT (Gold layer DLT) —
# Resolved: &env_schema → 'prod'
# Routed: aggregation → Gold
import dlt
from pyspark.sql import functions F

@dlt.table(
    name="gold_claims_summary",
    table_properties={
        "delta.autoOptimize":"true"
    })
def gold_claims():
    return (dlt.read("silver_claims")
            .groupBy("region")
            .agg(F.sum("amt")))
```

**Conversion Accuracy** **99.2%**

# Three products. **One Lakehouse migration.**

Map your estate, convert your code, govern your outputs — every transformation, end-to-end, on your infrastructure. Compass extracts column-level lineage from your source and publishes it directly into Unity Catalog before you migrate.

## DISCOVERY

### MigryX Compass

Map your estate before you migrate. Extracts column-level lineage, STTM, and dependency graphs from any source — and publishes them directly into Unity Catalog.

- Inventory + complexity scoring
- MIGRATE / ARCHIVE / DELETE classification
- Dependency graphs & risk scoring
- STTM published to Unity Catalog
- Wave-by-wave migration roadmap

## LINEAGE

### MigryX Atlas

Universal cross-platform lineage spanning SAS, SQL, Python, ETL tools, and 30+ languages — feeding rich, parser-generated metadata into Unity Catalog.

- Column-level traceability across stacks
- Cross-platform STTM & impact analysis
- GDPR / SOX / BCBS 239 audit reports
- Visual lineage explorer
- Unity Catalog Lineage API integration

## AI ENGINE

### Merlin AI

Domain-trained AI that understands legacy semantics, routes output into Bronze/Silver/Gold, and pushes accuracy to 99% on Databricks.

- Macro & parameter resolution
- Medallion-aware code routing
- Photon & Liquid Clustering hints
- Auto-documentation + data contracts
- Conversational Q&A on your estate

**99%**

AI-AUGMENTED ACCURACY

**10x**

FASTER THAN MANUAL

**60%+**

COST SAVINGS

**8+**

SOURCE ENGINES SUPPORTED



#### On-Premise & Air-Gapped

Your source code never leaves your network. Fully isolated, zero external dependencies. Deploy in under an hour.



#### Docker & Kubernetes

Containerized on Docker, Kubernetes, or OpenShift. Scales with your workload — no infrastructure surprises.



#### Enterprise Security

SOC 2 certified. Role-based access control, complete audit trails, and compliance reporting built in.

#### TRUSTED BY DATABRICKS ADOPTERS ACROSS

Fortune 500

Financial Services

Healthcare

Government

Insurance

Retail

Telecom

Pharmaceutical

# Your competitors are already on the Lakehouse. Are you still rewriting SAS by hand?

Every month on legacy ETL compounds the debt — in licensing, talent scarcity, compliance risk, and missed AI/ML opportunity. The Databricks Data Intelligence Platform is the destination. MigryX is how you get there.

⚡ **The talent cliff is real:** The pool of engineers who know SAS, BTEQ, DataStage, and legacy ETL shrinks every year. The longer you wait, the more expensive your migration becomes — and the harder it is to find anyone who can do it. Modern data engineers want PySpark and Delta Lake.



### Eliminate Legacy Licensing

SAS, Informatica, and DataStage licenses consume millions annually. Cutover to Databricks delivers more capability for a fraction of the cost — typically 60%+ savings.



### Unlock AI & ML on Your Data

SAS analytical models and scoring logic land as MLflow experiments with Model Registry, Feature Store, and Model Serving — production AI from day one.



### Streaming + Batch in One Place

DLT pipelines unify streaming and batch. Auto Loader CDC ingestion, declarative @dlt.table, data quality expectations, enhanced autoscaling — out of the box.



### Governance Built-In

Unity Catalog provides fine-grained access, attribute tags, and lineage from the moment your migrated tables land. Audit-ready GDPR, SOX, BCBS 239 reporting.



### Preserve Decades of Logic

Decades of SAS macros and Informatica mappings encode institutional knowledge. MigryX preserves every transformation with column-level lineage — nothing guessed.



### Photon-Optimized Performance

Legacy SQL transpiled to Photon-optimized Databricks SQL on serverless SQL Warehouses with 500+ dialect function mappings. Faster queries, lower cost.

FOR DECISION MAKERS

## See your legacy code land on Databricks — live, in 30 minutes.

No slides. No generic demos. We migrate a sample of your actual SAS, Talend, DataStage, Informatica, ODI, or SSIS code into Bronze/Silver/Gold notebooks with full Unity Catalog lineage so you can evaluate accuracy, structure, and governance yourself — zero commitment.

**60%+** COST SAVINGS    **10×** FASTER    **99%** AI ACCURACY    **<1hr** TO DEPLOY

[Schedule Your Lakehouse Demo](#) →

Or send us a sample — we'll parse it, convert to PySpark/Delta, and return Unity Catalog lineage free.  
[hello@migryx.com](mailto:hello@migryx.com) · (617) 512-9530

# MigryX

for  Databricks

## Stop rewriting legacy. Start running on the Lakehouse.

Join the enterprises that have already landed millions of lines of SAS, Talend, DataStage, Informatica, and SSIS code on Databricks — Medallion-structured, Unity Catalog-governed, Photon-optimized — in weeks, not years.

"MigryX cut our SAS-to-Databricks timeline from 3 years to 4 months. The Unity Catalog lineage alone justified the entire investment."



**VP of Data Engineering**  
Fortune 500 Financial Services

"We submitted DataStage jobs Tuesday. Thursday we had Bronze/Silver/Gold notebooks running on Databricks with full STTM. Nothing else comes close."



**Chief Data Officer**  
Global Healthcare Enterprise

### Free Proof of Concept — On Databricks



Send a sample of SAS, Talend, Alteryx, DataStage, Informatica, ODI, SSIS, or BTEQ. We'll convert it, deploy it as Asset Bundles to a Databricks workspace, and return Unity Catalog lineage — **free, no commitment. Start your POC →**

**60%+**

COST SAVINGS

**10×**

FASTER

**99%**

AI ACCURACY

[Schedule Your Lakehouse Demo →](#)

[databricks.migryx.com](https://databricks.migryx.com) | [hello@migryx.com](mailto:hello@migryx.com) | (617) 512-9530