

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

Land Every Legacy ETL Workload on BigQuery.

MigryX converts SAS, Talend, Alteryx, IBM DataStage, Informatica, Oracle ODI, SSIS, Teradata, and SQL dialects directly to **BigQuery SQL, Dataform, Dataproc, Cloud Composer, BigQuery ML, and Vertex AI** — with 95%+ parser accuracy and 99% with AI augmentation.

99%AI
ACCURACY**60%+**COST
SAVINGS**10x**

FASTER

8+SOURCE
ENGINES[Schedule a BigQuery Demo →](#)[Email us →](#)

silver_claims.sqlx · Dataform

```
-- Source: SAS DI Studio
-- Target: Dataform + BigQuery
config { type:"table" }
```

```
SELECT
  claim_id, region, amt,
  CURRENT_TIMESTAMP() ts
FROM ${ref("bronze_claims")}
WHERE claim_id IS NOT NULL
QUALIFY ROW_NUMBER()
  OVER (PARTITION BY claim_id
  ORDER BY ts DESC) = 1
```

Conversion Accuracy

99.2%

Manual BigQuery 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. BigQuery Migration Service handles SQL translation; it doesn't parse SAS macros, DataStage XML, or SSIS .dtsx packages. That's where customers stall.

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 BigQuery with broken lineage, no STTM, and reverse-engineered logic nobody trusts.

\$2M+

ANNUAL LEGACY SPEND

3-5yr

MANUAL REWRITE TIMELINE



Manual Migration to BigQuery

- ✗ 3-5 years of consultant-led rewrites burning budget
- ✗ SAS macros, DataStage XML, SSIS .dtsx logic lost in translation
- ✗ No column-level lineage or STTM — Knowledge Catalog launched empty
- ✗ Pipelines land as raw SQL — no Dataform, no Dataproc
- ✗ No Cloud Composer DAGs, no BigQuery ML — just code dumps
- ✗ Validation is manual — production incidents discovered post go-live



MigryX → BigQuery

- ✓ Custom AST parsers per source — 95%+ deterministic, 99% with AI
- ✓ Output as Dataform SQLX with refs, tags, and assertions
- ✓ Heavy ETL converted to Dataproc PySpark + BigLake/Iceberg
- ✓ Cloud Composer (Airflow) DAGs replace legacy job schedulers
- ✓ STTM and lineage published to Knowledge Catalog with policy tags
- ✓ Native data validation using BigQuery comparison queries

5-STAGE METHODOLOGY

From legacy artifact to production BigQuery — 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 Dataform SQLX, Dataproc PySpark, or Cloud Composer DAGs.

4

Validate

Row-level + aggregate matching using BigQuery-native queries — audit-ready.

5

Govern

STTM published to Knowledge Catalog with policy tags and Lineage API.

Not just code conversion. Production-ready Google Cloud architecture.

Every MigryX migration leverages the full Google Cloud Data Cloud — BigQuery serverless compute, Dataform pipelines, Cloud Composer orchestration, BigLake/Iceberg interoperability, Knowledge Catalog governance, BigQuery ML, and Vertex AI. Production-ready from day one.

GOOGLE CLOUD-NATIVE OUTPUT · AUTO-GENERATED

● Compute & SQL

BigQuery SQL (GoogleSQL)

Serverless

Slot Reservations

UDFs / Remote Functions

Stored Procedures

500+ Function Mappings

● Pipelines & Orchestration

Dataform (SQLX)

Dataproc (PySpark)

Cloud Dataflow

Cloud Composer (Airflow)

Pub/Sub → BQ Subscriptions

Data Transfer Service

● Storage & Tables

Native BigQuery Tables

BigLake

Iceberg Tables

Time Travel

MERGE INTO

Multi-statement Transactions

Change Data Capture

● Governance

Knowledge Catalog

Policy Tags

Column-Level Lineage

Data Profiling

DLP / Data Masking

Row-Level Security

Lineage API

● AI & ML

BigQuery ML

Vertex AI

Gemini in BigQuery

AI Functions

Vector Search

BigQuery Notebooks

Conversational Analytics

● Apps & Sharing

Analytics Hub

Data Clean Rooms

Authorized Views

Looker Studio

Multi-Cloud (Omni)



Precision Parsing

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



BigQuery-Native

Dataform SQLX, Dataproc PySpark, Cloud Composer —
not generic SQL on a warehouse.



Catalog-Ready

STTM and lineage published
into Knowledge Catalog with
policy tags and DLP enforcement.

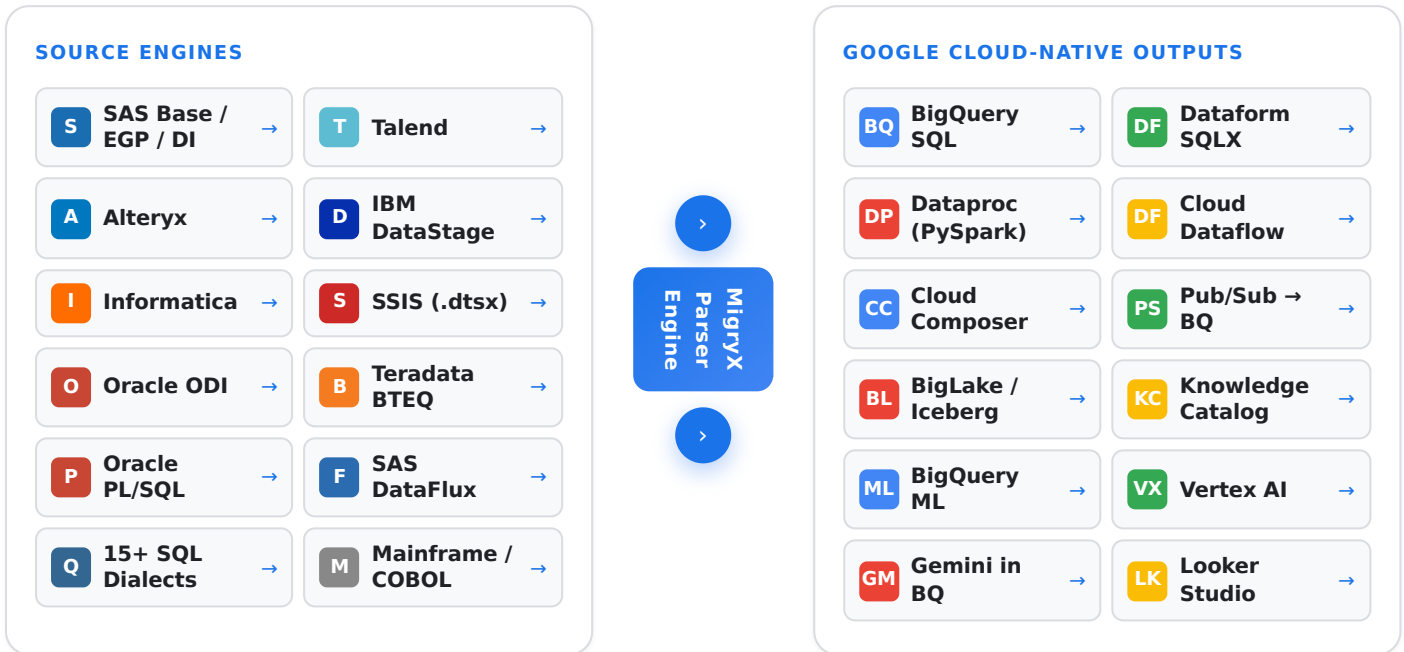


Air-Gapped Capable

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

Every legacy ETL source. One BigQuery 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 → BIGQUERY	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 Google Cloud. Merlin AI resolves what deterministic parsers cannot.

Why Merlin lands cleaner code on BigQuery

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 BigQuery SQL or Dataproc PySpark.



BigQuery-Aware Conversion

Recognizes streaming vs batch vs incremental logic and routes output to Pub/Sub→BQ subscriptions, Cloud Composer, or Dataform incremental tables automatically.



Auto-Generated Documentation

Creates STTM, data contracts, and Dataform SQLX assertions — published into Knowledge Catalog with policy tags, zero manual effort.



Optimization Recommendations

Surfaces partitioning, clustering, and slot reservation hints from the parsed estate — plus BQ ML candidates for SAS PROC code.

```
gold_claims_dataform.sqlx

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

-- MERLIN OUTPUT: Dataform SQLX
-- Resolved: &env_schema → 'prod'
-- Routed: aggregation → Gold layer
config {
  type: "table",
  schema: "gold",
  tags: ["daily"]
}
SELECT region,
  SUM(amt) AS total_amt
FROM ${ref("silver_claims")}
GROUP BY region;
```

Conversion Accuracy **99.2%**

Three products. One BigQuery 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 the Google Cloud Knowledge Catalog before you migrate.

DISCOVERY

MigryX Compass

Map your estate before you migrate. Extracts column-level lineage, STTM, and dependency graphs — and publishes them directly into the Knowledge Catalog.

- Inventory + complexity scoring
- MIGRATE / ARCHIVE / DELETE classification
- Dependency graphs & risk scoring
- STTM published with Policy Tags
- 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 Knowledge Catalog.

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

AI ENGINE

Merlin AI

Domain-trained AI that understands legacy semantics, routes output to Dataform, Dataproc, or Cloud Composer, and pushes accuracy to 99%.

- Macro & parameter resolution
- BigQuery-aware code routing
- Partitioning & 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



On-Premise & Air-Gapped

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



GKE & Cloud Run

Containerized on Docker, GKE, or Cloud Run. 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 GOOGLE CLOUD ADOPTERS ACROSS

Fortune 500

Financial Services

Healthcare

Government

Insurance

Retail

Telecom

Pharmaceutical

Your competitors are already on Google Cloud. 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. BigQuery 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 BigQuery, Dataform, and Gemini.



Up to 54% Lower TCO

BigQuery's serverless architecture and compute autoscaling deliver up to 54% lower TCO than alternative cloud data warehouses. Eliminate legacy SAS, Informatica, and DataStage licenses entirely.



Gemini-Native AI

SAS analytical models land as BigQuery ML and Vertex AI deployments. Gemini-powered AI Functions, Conversational Analytics, and Vector Search ready to use — agentic AI from day one.



Streaming + Batch Unified

Pub/Sub BigQuery Subscriptions stream data in real-time. Dataform handles incremental batch. Cloud Composer orchestrates everything — no Kafka, no separate streaming stack.



Governance Built-In

Knowledge Catalog provides automated metadata harvesting, data profiling, policy 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.



Open Lakehouse with BigLake

BigLake + Iceberg tables provide read/write interoperability with BigQuery, Spark, and OSS engines — zero data movement. Open formats, enterprise-grade engines.

FOR DECISION MAKERS

See your legacy code land on BigQuery — 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 Dataform SQLX with full Knowledge Catalog lineage so you can evaluate accuracy, structure, and governance yourself — zero commitment.

60%+
COST SAVINGS

10x
FASTER

99%
AI ACCURACY

<1hr
TO DEPLOY

[Schedule Your BigQuery Demo →](#)

Or send a sample — we'll convert to Dataform/BigQuery and return Knowledge Catalog lineage free.
hello@migryx.com · [\(617\) 512-9530](tel:(617)512-9530)

MigryX for  Google BigQuery

Stop rewriting legacy. Start running on BigQuery.

Join the enterprises that have already landed millions of lines of SAS, Talend, DataStage, Informatica, and SSIS code on Google Cloud — Dataform-native, Knowledge Catalog-governed, Gemini-ready — in weeks, not years.

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



VP of Data Engineering

Fortune 500 Financial Services

"We submitted DataStage jobs Tuesday. Thursday we had Dataform SQLX and Cloud Composer DAGs running with full STTM. Nothing else comes close."



Chief Data Officer

Global Healthcare Enterprise

Free Proof of Concept — On BigQuery



Send a sample of SAS, Talend, Alteryx, DataStage, Informatica, ODI, SSIS, or BTEQ. We'll convert it, deploy to a BigQuery project, and return Knowledge Catalog lineage — **free, no commitment. Start your POC →**

60%+

COST SAVINGS

10×

FASTER

99%

AI ACCURACY

Schedule Your BigQuery Demo →

bigquery.migryx.com | hello@migryx.com | (617) 512-9530