MIGRYX

Migrate Everything to Databricks

Databricks Technology Partner | Reference Guide 2026

databricks.migryx.com |

This reference guide provides a comprehensive overview of the MigryX Databricks migration platform—covering supported legacy sources, Databricks target outputs, deep platform integrations, architecture, and platform capabilities.

10+	+95%	85%	Column
LEGACY SOURCES	PARSER ACCURACY	FASTER MIGRATION	LEVEL LINEAGE

Section 1: Overview

MigryX converts SAS, Talend, Alteryx, IBM DataStage, Informatica, Oracle ODI, SSIS, Teradata, and SQL dialects directly to Databricks—PySpark, Delta Lake (MERGE, Z-ORDER, Liquid Clustering), Medallion Architecture, Unity Catalog, DLT Pipelines, and Databricks Workflows—with +95% parsing accuracy and column-level lineage. Parser accuracy up to 99% with AI augmentation.

Section 2: Databricks Targets

1. **PySpark Jobs** — Production-grade PySpark with Auto Loader, CDF CDC, Medallion Architecture layering
2. **Delta Lake Tables** — ACID-compliant, MERGE INTO, OPTIMIZE & Z-ORDER, Liquid Clustering, time travel
3. **Unity Catalog** — Column-level lineage, STTM, attribute tags, fine-grained access controls

4. **Databricks Workflows** — ETL via Asset Bundles (DABs), multi-task DAGs, serverless compute
5. **Delta Live Tables (DLT)** — Streaming/batch DLT pipelines, @dlt.table, Auto Loader CDC, quality expectations
6. **Databricks Notebooks** — Annotated notebooks, %sql/%python, Databricks Connect, lineage comments
7. **MLflow & Feature Store** — SAS models to MLflow, model registry, Feature Store, AutoML, Model Serving
8. **Databricks SQL Warehouse** — Photon-optimized, serverless SQL, 500+ dialect function mappings

Section 3: Supported Migration Sources

1. **SAS** — Base, Macros, PROC SQL, SAS/IML → PySpark, Delta Lake, Databricks SQL, MLflow
2. **Talend** — Studio, Open Studio, tMap, Cloud → PySpark, Workflows, Delta Lake
3. **Alteryx** — Designer, Workflows, Macros, Apps → PySpark, Databricks SQL, Notebooks
4. **IBM DataStage** — Parallel, Server, DataStage X → PySpark, DLT Pipelines, Delta Lake
5. **Informatica** — PowerCenter, IDMC, IICS → PySpark, Unity Catalog, Workflows
6. **Oracle ODI** — Mappings, KMs, Load Plans → PySpark, Delta Lake, Workflows
7. **SSIS** — .dtsx, .ispac, Data Flow, Scripts → PySpark, Workflows, Delta Lake
8. **Teradata** — BTEQ, FastLoad, QUALIFY, Macros → Databricks SQL, PySpark, Delta Lake
9. **Oracle PL/SQL** — Procedures, Packages, Triggers → Databricks SQL, Delta Lake, Python UDFs
10. **SQL Dialects** — 15+ dialects, 500+ function maps → Databricks SQL, SQL Warehouse, Delta Live

11. **SAS DataFlux** — dfPower Studio, DMS, DQ Schemes → PySpark, Great Expectations, Delta Lake
12. **MigryX Compass** — Discovery, lineage, Unity Catalog registration

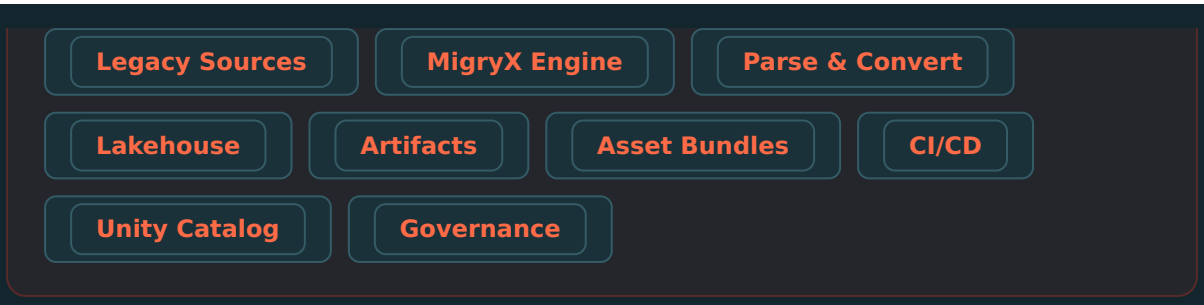
Section 4: Deep Platform Integrations

1. **Photon Engine Optimization** — Vectorized column ops, predicate pushdown, Photon-compatible patterns
2. **Serverless Compute** — Serverless SQL Warehouses and Jobs compute, auto-provisioned
3. **LakeFlow Connect** — Managed CDC ingestion replacing legacy connectors
4. **Mosaic AI & Model Serving** — SAS model conversion, A/B testing, Feature Engineering tables
5. **Asset Bundles (DABs)** — CI/CD packaging, version-controlled YAML, environment promotion
6. **Unity Catalog Governance** — Lineage, STTM, data classification, row/column security
7. **Databricks SQL & AI/BI Dashboards** — Report migration, parameterized queries, alert triggers
8. **Delta Sharing** — Cross-org data exchange, recipients/providers/shares
9. **Databricks Apps** — Migration dashboards as native Streamlit/Gradio workspace apps

Section 5: Architecture Flow

The PDF guide includes a visual pipeline from legacy sources through the MigryX engine to the lakehouse, artifacts, CI/CD, and governance. Stage labels in the original document:

Pipeline stages (as in PDF diagrams)



Section 6: How It Works

1. **Ingest** — Upload source artifacts
2. **Parse & Analyze** — Custom AST parsers, macro expansion, column-level lineage
3. **Convert** — Parser-driven conversion to PySpark, Delta Lake, Databricks SQL, Workflows, DLT
4. **Validate** — Row-level and aggregate data matching, audit-ready evidence
5. **Govern** — Publish lineage, STTM, data contracts to Unity Catalog

Section 7: Platform Capabilities

1. **Custom-Built Parsers** — Purpose-built per source, full fidelity, deterministic
2. **Medallion Architecture** — Bronze/Silver/Gold auto-layering
3. **Delta Lake Native Output** — MERGE INTO, Z-ORDER, Liquid Clustering, CDF, time travel
4. **Unity Catalog Lineage** — Column-level STTM, tags, Lineage API
5. **Merlin AI & MLflow** — AI optimization, SAS models to MLflow
6. **On-Premise & Air-Gapped** — Behind firewall, DAB packaging, SOX/GDPR/BCBS 239

Section 8: Why MigryX vs Generic Tools

Feature	MigryX	Generic
Custom parser per source	YES	NO
Column-level lineage to Unity Catalog	YES	~
Native Delta Lake & DLT output	YES	NO
Databricks Workflows generation	YES	NO
SAS macro expansion	YES	NO
Medallion Architecture generation	YES	NO
Asset Bundles (DABs) CI/CD	YES	NO
On-premise / air-gapped	YES	NO
Deterministic AST-based parsing	YES	NO
MLflow model migration	YES	NO

Section 9: Business Impact

85%
FASTER DELIVERY

70%
RISK REDUCTION

60%
LOWER COSTS

+95%
PARSER ACCURACY

MigryX delivers measurable business value by dramatically accelerating migration timelines, reducing risk through deterministic parsing and validation, lowering costs with automated conversion, and ensuring governance compliance via Unity Catalog integration.

Ready to migrate?

databricks.migryx.com |

This HTML version is derived from the content of `MigryX-Databricks-Reference-Guide.pdf`. Diagram-heavy pages in the PDF are summarized in Section 5; refer