

Data Cosmos_DataFlux

Intelligent Scriptless Ingestion

Part of the Coforge Data Cosmos™ Platform | Hypernova™ Portfolio

Coforge Limited | Data HBU

2026

Solution Overview

Data Cosmos_DataFlux is a dynamic, zero-code data ingestion platform designed to simplify and streamline enterprise data management at scale. It serves as an enterprise-grade, AI-powered data migration and ingestion platform engineered to simplify, automate, and govern large-scale data movement across heterogeneous systems. Acting as a centralized control plane for managing data ingestion workflows, DataFlux leverages AI-driven agents to automate complex tasks such as script generation, schema validation, and error recovery — making data management efficient and accessible without requiring programming expertise.

In modern enterprises, data teams spend up to 60–70% of their time on pipeline plumbing — writing ingestion scripts, debugging schema mismatches, managing scheduling dependencies, and troubleshooting failures. DataFlux eliminates this burden by providing an intelligent, self-healing ingestion fabric that connects any source to any target, generates production-ready pipelines on demand, and continuously monitors data quality — all through a single, unified interface. Whether migrating terabytes from legacy Oracle databases to Azure Data Lake, streaming real-time Kafka events into Databricks Delta Lake, or consolidating multi-cloud data estates, DataFlux delivers consistent, governed, and auditable data movement at enterprise scale.

Key Features

Feature	Description
Zero-Code Platform	Enables business and technical users to configure complex multi-source, multi-target ingestion pipelines through an intuitive UI without writing a single line of code. Users define sources, targets, transformation rules, and scheduling — the platform generates and deploys production-ready pipelines automatically.
AI-Driven Agents	Six specialized AI agents collaborate autonomously to handle every stage of the data ingestion lifecycle — from pipeline generation and schema monitoring to orchestration, quality assurance, failure recovery, and compliance reporting.
Visualization & Monitoring	Interactive dashboards provide real-time visibility into pipeline execution, data quality scores, throughput metrics, and system health — enabling proactive issue identification before SLAs are impacted.
Multi-Source Connectivity	Connects to 30+ source types including RDBMS (Oracle, SQL Server, MySQL, PostgreSQL), NoSQL (MongoDB, Cassandra), cloud warehouses (Snowflake, BigQuery, Redshift), streaming platforms (Kafka, Event Hubs), file systems (SFTP, S3, ADLS), and SaaS APIs.
Multi-Cloud & Hybrid	Deploy across AWS, Azure, or GCP using Terraform with centralized control via a single UI. Supports hybrid scenarios where sources are on-premises and targets are cloud-native.
Extensibility & Integration	Open architecture supporting integration with external orchestrators (Airflow, ADF), governance tools (Purview, Collibra), and CI/CD pipelines for DevOps-native data engineering.

Six Autonomous AI Agents

Agent	Role	What It Does
Orion	Zero-Code Pipeline Generation	Turns high-level configurations (“sync this customer data every night”) into production-ready Python ingestion pipelines with the right settings for environment, dates, batching, and parallelism.
Vega	Schema Drift Detection	Continuously monitors data shapes. When a column is added, renamed, or a data type changes, Vega detects the drift, raises an alert, and auto-reconciles DDL changes in the target to prevent failures.
Andromeda	DAG Orchestration	Builds and manages execution graphs (DAGs) with scheduling, dependencies, concurrency control, and retry logic — ensuring pipelines execute in the right order with the right resources.
Phoenix	Failure Recovery	Detects pipeline failures in real time, performs intelligent root cause analysis, and auto-remediates common failure patterns (connection timeouts, lock conflicts, resource exhaustion).
Sirius	Data Quality Rules	Analyzes data profiles and auto-generates context-aware quality rules. Scores every dataset load across six dimensions — completeness, accuracy, consistency, timeliness, uniqueness, and validity.
Atlas	Audit & Compliance	Maintains a complete audit trail of every data movement — what happened, when, by whom, and why. Generates lineage maps and compliance reports for regulatory oversight (GDPR, HIPAA, SOX, BCBS 239).

How DataFlux Works

1. Discover — Connect to sources and explore schemas, metadata, and sample data across the enterprise
2. Configure — Define per-table migration modes, batching strategies, parallelism, keys, and watermarks

3. Generate — AI auto-creates production-ready Python ingestion pipelines tailored to each source-target pair
4. Target Setup — Configure warehouses, staging areas, and DDL synchronization for target environments
5. Orchestrate — Build DAGs with scheduling, dependencies, concurrency control, and retry policies
6. Execute — Run pipelines on the selected execution engine with full resource optimization
7. Validate — Apply AI-generated DQ rules and score datasets across six quality dimensions
8. Monitor — Track live execution metrics, throughput, latency, and system health in real time
9. Recover — Detect schema drift and auto-remediate failures with intelligent retry and rollback logic
10. Govern — Maintain complete audit trails, data lineage, compliance reporting, and regulatory documentation

Industry Trends and where it can be leveraged

Banking & Financial Services

Use Case	Example
Regulatory Data Consolidation	A global bank consolidates data from 15+ legacy core banking systems (Temenos, Finacle, FIS) into a centralized Azure Data Lake for BCBS 239 regulatory reporting. DataFlux automates nightly batch ingestion of 500+ tables with CDC-based incremental loads, schema drift protection across frequent core banking upgrades, and AI-generated DQ rules ensuring data completeness for risk aggregation. Orion generates pipelines for each source system, while Atlas provides end-to-end lineage for audit trails required by regulators.
Payments & Transaction Analytics	A retail bank streams real-time payment transaction data from Kafka into Databricks Delta Lake for fraud detection analytics. Andromeda orchestrates the streaming pipeline with watermark-based ingestion, while Sirius applies real-time DQ scoring to flag anomalous transaction patterns before they reach downstream fraud models.

Insurance

Use Case	Example
Claims Data Platform Modernization	A multi-line insurer migrates claims data from legacy Guidewire ClaimCenter and Duck Creek systems into Snowflake for a unified claims analytics platform. DataFlux handles schema mapping across disparate claim formats (auto, property, liability), automates CDC pipelines for near-real-time claims status updates, and applies PII detection on policyholder data using Sirius — enabling actuarial teams to build loss-ratio models on trusted, governed data.
Underwriting Data Integration	An underwriting team consolidates third-party risk data feeds (LexisNexis, Verisk, credit bureaus) with internal policy administration data. DataFlux connects to 10+ external APIs and SFTP sources, normalizes heterogeneous formats, and lands cleansed data into Azure Synapse for underwriting decisioning — reducing data preparation time from weeks to days.

Travel, Transportation & Hospitality

Use Case	Example
Real-Time Booking & Revenue Analytics	A global airline ingests real-time booking data from Amadeus/Sabre GDS, loyalty program transactions, and ancillary revenue streams into a Databricks Lakehouse. DataFlux orchestrates 200+ daily pipeline runs across batch and streaming modes, with Phoenix auto-recovering from GDS API timeouts during peak booking hours. Revenue management teams get near-real-time visibility into seat inventory and yield optimization.
Customer 360 Data Consolidation	A hotel chain unifies guest profile data from PMS (Opera), CRM (Salesforce), loyalty platforms, and OTA channels (Booking.com, Expedia) into a centralized Customer 360 on Azure Data Lake. DataFlux handles identity resolution across systems with different guest ID formats and applies automated DQ checks to ensure profile completeness for personalization engines.

Healthcare & Life Sciences

Use Case	Example
Clinical Data Lake for Research	A healthcare provider migrates EMR/EHR data (Epic, Cerner) along with lab results, imaging metadata, and claims data into a HIPAA-compliant Azure Data Lake for population health analytics. DataFlux applies automated PII/PHI detection and masking using Sirius and Atlas agents, ensuring de-identified data is available for research teams while maintaining full audit trails for HIPAA compliance.
Clinical Trial Data Integration	A pharma company consolidates clinical trial data from multiple CROs, CTMS platforms, and EDC systems (Medidata Rave, Oracle InForm) into a unified research data platform. DataFlux automates ingestion of diverse file formats (SAS, CSV, XML) with schema validation, enabling biostatisticians to access harmonized trial data 40% faster.

Cross-Industry Use Cases

Use Case	Description
Real-Time Analytics Pipelines	Stream Kafka clickstream data into Databricks Delta Lake/Postgres with watermark-based ingestion and real-time data quality scoring.
Legacy System Consolidation	Unify data from multiple RDBMS and NoSQL systems into a centralized cloud platform with schema drift protection.
CDC-Based Operational Replication	Enable near-real-time CDC pipelines from Oracle or SQL Server into Snowflake or Redshift for operational reporting.
AI-Driven Data Quality Audits	Auto-generate DQ rules from data profiles and score datasets across six quality dimensions.
Multi-Cloud Data Platform Rollout	Deploy across AWS, Azure, or GCP using Terraform with centralized control via a single UI.

Target Audience

- Enterprise Data Engineering & Platform Teams
- Data Architects and Cloud Migration Leads
- Cloud Program & Delivery Managers
- Data Quality & Governance Teams
- DevOps & Platform Engineers

Azure Compatibility

DataFlux integrates natively with Azure Data Factory, Azure Databricks, Azure Data Lake Storage, Azure Synapse Analytics, Microsoft Fabric, and Microsoft Purview. It is part of the Coforge Data Cosmos Hypernova portfolio for greenfield data programs and advanced analytics, fully designed to leverage Azure-native services for scalable, secure, and governed enterprise data ingestion. DataFlux pipelines can be orchestrated via Azure Data Factory, monitored through Azure Monitor, and governed through Microsoft Purview for end-to-end observability and compliance.