

Data Cosmos_Code_Conversion_Agent

Multi-Agent AI Solution for Code Conversion

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

Coforge Limited | Data HBU

2026

Solution Overview

Data Cosmos_Code_Conversion_Agent is a multi-agent AI system designed to accelerate large-scale data platform modernization. It serves as an enterprise-grade code conversion engine that automatically converts DDL, DML, SQL stored procedures, and transformation code from legacy platforms to modern cloud-native stacks — preserving business logic with high fidelity and delivering full auditability. Part of Coforge Data Cosmos™, the innovation backbone combining platforms, agentic accelerators, and services, it accelerates execution across every phase of the data modernization lifecycle.

In modern enterprises, migrating legacy ETL and database code to cloud-native platforms is slow, manual, and error-prone — teams spend weeks manually rewriting SSIS packages, Informatica mappings, and Oracle/SQL Server stored procedures, risking loss of business logic and lengthy validation cycles. Code Conversion Agent eliminates this burden through a 4-stage multi-agent pipeline that maps lineage, converts code to target syntax, validates logic fidelity, and automatically fixes failures. It ingests source code from SSIS, Informatica, Oracle, SQL Server, and other RDBMS/EDW platforms and converts to 20+ modern targets including dbt on Snowflake/Databricks/BigQuery, PySpark on Databricks, and native SQL for Snowflake, BigQuery, and Azure Synapse — converting full mappings in minutes instead of weeks.

Key Features

Feature	Description
Multi-Agent Conversion Pipeline	A 4-stage pipeline — Lineage Agent, Conversion Agent, Validation Agent, and AutoFix Loop — automates end-to-end code conversion with confidence scoring per conversion unit.
Broad Source & Target Coverage	Ingests SSIS, Informatica PowerCenter, Oracle PL/SQL, SQL Server T-SQL, and other RDBMS/EDW platforms; converts to 20+ targets including dbt, PySpark, and native SQL for Snowflake, BigQuery, and Synapse.
Multi-LLM Flexibility	Supports Google Gemini, OpenAI GPT-4o, Anthropic Claude, and local Ollama with customizable prompts to enforce naming standards and coding patterns — no vendor lock-in.
Business Logic Preservation	Preserves business logic, naming conventions, and data types during conversion, with the Validation Agent checking semantic equivalence between source and target outputs.
Enterprise Auditability	Delivers pre/post-validation snapshots, confidence scores, validation reports, and auto-generated lineage diagrams — providing full traceability for modernization audits.
Batch & Scalable Conversion	Converts dozens of pipelines in parallel with cost transparency — tracking estimated LLM token usage for budget planning across large-scale modernization programs.

4-Stage Multi-Agent Pipeline

Agent	Role	What It Does
Lineage Agent	Data Flow Mapping	Generates end-to-end data flow diagrams showing source-to-target transformations, dependencies, and system touchpoints — providing visual validation context before conversion begins.
Conversion Agent	Target Syntax Translation	Translates source code into target-platform syntax, handling DDL conversion, stored procedure translation, ETL logic mapping, and transformation rule adaptation — preserving business logic, naming conventions, and data types.
Validation Agent	Correctness & Logic Fidelity	Validates correctness, completeness, and logic fidelity. Compares source and target outputs, checks semantic equivalence, and generates confidence scores per conversion unit.
AutoFix Loop	Automatic Retry & Patching	Automatically retries and patches validation failures, iterating through failed conversions with modified approaches until validation passes or escalates for manual review.

How Code Conversion Agent Works

- Ingest** — Ingest source code from SSIS, Informatica, Oracle, SQL Server, and other RDBMS/EDW platforms
- Map Lineage** — Lineage Agent generates data flow diagrams showing source-to-target transformations and dependencies
- Select Target** — Choose from 20+ modern targets including dbt, PySpark, and native SQL for Snowflake/BigQuery/Synapse
- Configure LLM** — Select LLM (Gemini, GPT-4o, Claude, Ollama) and customize prompts for naming and coding standards
- Convert** — Conversion Agent translates DDL, DML, stored procedures, and ETL logic to target-platform syntax
- Validate** — Validation Agent checks correctness, semantic equivalence, and generates confidence scores
- AutoFix** — AutoFix Loop retries and patches validation failures until they pass or escalate for review
- Batch Process** — Convert dozens of pipelines in parallel with cost transparency and token usage tracking
- Report** — Generate source vs. target code, validation reports, lineage diagrams, and confidence scores
- Deliver** — Package outputs into a downloadable ZIP with execution logs for enterprise auditability

Industry Trends and where it can be leveraged

Banking & Financial Services

Use Case	Example
Risk Procedure Modernization	A global bank uses Code Conversion Agent to convert Oracle PL/SQL risk procedures to PySpark on Databricks. The Lineage Agent maps hundreds of interdependent stored procedures for audit compliance, while the Validation Agent ensures risk-calculation logic is preserved with high-confidence scoring.
Regulatory Reporting Migration	A retail bank migrates legacy T-SQL regulatory reporting logic to native SQL on Azure Synapse. Auto-generated lineage diagrams simplify validation for regulators, while AutoFix handles complex aggregation logic conversions.

Insurance

Use Case	Example
Claims ETL Migration	A multi-line insurer migrates Informatica ETL to dbt on Snowflake for claims analytics. The Validation Agent ensures actuarial logic is preserved with 95%+ confidence, while batch conversion processes dozens of mappings in parallel.
Policy Admin Modernization	An insurer converts legacy policy administration stored procedures to PySpark, using customizable prompts to enforce enterprise naming standards and preserve business rules across auto, property, and liability lines.

Travel, Transportation & Hospitality

Use Case	Example
SSIS Package Conversion	A global airline converts hundreds of SSIS packages to PySpark on Databricks. Code Conversion Agent batch-processes packages in parallel, dramatically reducing a multi-month migration to weeks while preserving booking and revenue transformation logic.
Data Warehouse Modernization	A hotel chain converts legacy EDW transformation code to dbt on Snowflake, using auto-generated lineage diagrams to validate guest and loyalty data flows across fragmented source systems.

Healthcare & Life Sciences

Use Case	Example
Stored Procedure Migration	A healthcare provider migrates SQL Server stored procedures to Azure Synapse SQL. The AutoFix Loop handles HIPAA-specific data handling logic, ensuring sensitive data transformations are converted correctly with full validation reports.
Research Pipeline Conversion	A pharma company converts legacy clinical research ETL to PySpark, with the Validation Agent ensuring data transformation fidelity across diverse trial data formats and the Lineage Agent documenting provenance.

Cross-Industry Use Cases

Use Case	Description
Legacy ETL Modernization	Convert SSIS and Informatica ETL mappings to dbt or PySpark on modern cloud platforms, eliminating manual rewrite effort.
Stored Procedure Conversion	Translate Oracle PL/SQL and SQL Server T-SQL stored procedures to PySpark or native cloud SQL with preserved business logic.

Use Case	Description
Data Warehouse Migration	Convert legacy EDW transformation code to modern lakehouse and warehouse targets with auto-generated lineage and validation.
Audit-Ready Modernization	Deliver source vs. target code, confidence scores, and lineage diagrams for enterprise-grade modernization audits and governance.
Accelerated Cloud Migration	Batch-convert dozens of pipelines in parallel to compress multi-month migration programs into weeks with cost transparency.

Target Audience

- Data Engineering & Modernization Teams
- Cloud Migration Leads
- Data Architects
- ETL / Data Platform Engineers
- Analytics & Governance Leaders

Azure Compatibility

Data Cosmos_Code_Conversion_Agent integrates natively with Azure OpenAI Service for multi-LLM reasoning, Azure Kubernetes Service (AKS) for scalable deployment, and Azure Blob Storage / Azure Data Lake Storage for source code and outputs. Conversion targets include Azure Databricks (PySpark, dbt), Azure Synapse Analytics (native SQL), and Snowflake on Azure. It integrates with Microsoft Purview for governance and Azure Monitor for observability. It is part of the Data Cosmos Supernova portfolio for data modernization, fully designed to leverage Azure-native services for scalable, secure, and governed enterprise legacy code conversion and data platform modernization.