

Data Cosmos_SynthData

AI-Powered Synthetic Test Data Generation Engine

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

Coforge Limited | Data HBU

2026

Solution Overview

Data Cosmos_SynthData is an AI-powered synthetic test data generation accelerator designed to create production-realistic test data at scale while eliminating risks associated with using real or production data. It serves as an enterprise-grade engine that understands relational schemas, dependencies, and constraints to generate high-quality synthetic datasets with guaranteed referential integrity.

In modern enterprises, teams rely on production data copies or manual creation of test datasets, leading to compliance risks, privacy issues, and inefficiencies. SynthData eliminates this burden through AI-driven schema-aware generation and statistical pattern learning, enabling secure and scalable test data provisioning for development, testing, analytics, and CI/CD use cases.

Key Features

Feature	Description
Schema-Aware Generation	Understands primary keys, foreign keys, and constraints, generating data in dependency order with referential integrity.
CSV Pattern Learning	Learns distributions and correlations from CSVs without storing original data.
Deterministic Data	Ensures repeatable datasets for consistent testing and validation.
CI/CD Integration	Auto-generates scripts for pipeline-based test data provisioning.
Compliance-Safe Data	100% synthetic data eliminates regulatory and privacy risks.
Scalable Generation	Produces large volumes of test data rapidly for enterprise workloads.

4-Stage Intelligent Pipeline

Stage	Role	What It Does
Schema Ingestion	Schema Understanding	Ingest schemas and detect PK/FK relationships and dependencies.
Pattern Learning	Statistical Modeling	Analyze patterns and correlations for realistic data generation.
Synthetic Generation	Data Creation	Generate synthetic datasets with referential integrity.
Validation & Export	Governance	Validate and export datasets for testing pipelines.

How SynthData Works

- Discover — Connect schemas and test requirements
- Ingest — Load schema definitions or CSV patterns
- Analyze — Identify structure, constraints, distributions
- Learn — Apply statistical learning on patterns
- Generate — Produce synthetic datasets
- Validate — Ensure referential integrity
- Export — Deliver datasets to pipelines
- Integrate — Integrate with CI/CD workflows
- Execute — Use in testing and validation
- Govern — Maintain audit and compliance

Industry Trends and where it can be leveraged

Banking & Financial Services

Synthetic data for regulatory testing and data migration validation.

Insurance

Claims and policy testing with privacy-safe datasets.

Travel, Transportation & Hospitality

Booking system testing and simulation of peak loads.

Healthcare & Life Sciences

Secure synthetic patient data for HIPAA-compliant testing.

Target Audience

- Data Engineering Teams
- QA & Testing Teams
- DevOps Teams
- Security & Compliance Teams
- Analytics Teams

Azure Compatibility

SynthData integrates with Azure OpenAI, AKS, Azure Blob Storage, Azure Data Factory, and Azure DevOps for scalable synthetic data generation and governance.