Lyrebird BY SUBSETRA

ENTERPRISE SYNTHETIC TEST DATA

Production-like test databases. Without copying production.

Lyrebird reads your database once, then generates as many synthetic test databases as you need: production-shaped, every foreign key intact, and people invented instead of copied. It runs fully offline, on your own machines.

100% offline | Every foreign key intact | PostgreSQL · MySQL · SQL Server · Oracle | Linux · macOS · Windows

0 / 75
broken foreign keys on four public sample databases

84%
of columns keep the source's distribution

< 1 min
for a 10 million row PostgreSQL test database

THE PROBLEM

Test data is stuck between realism and risk.

Realistic tests need data that looks like production. Privacy and compliance need customer rows to stay in production. Most teams end up with one of three bad compromises.

Production copies

Masked or not, real records end up in staging, CI and laptops, and every refresh is a new audit question.

Slow refreshes

Restores and masking jobs take hours or days, so tests run against stale data or wait in a ticket queue.

Hand-made seed data

Safe, but small and simple. It breaks on real schemas and misses the edge cases that fail in production.

WHAT CHANGES WITH LYREBIRD

COMPLIANCE

Keep customer data in production

Synthetic rows take PII out of dev, staging and CI, and shrink your KVKK and GDPR audit scope.

COST

Stop maintaining full replicas

Generate a test database sized to the job, use it, and throw it away. No clones to store, refresh and back up.

SPEED

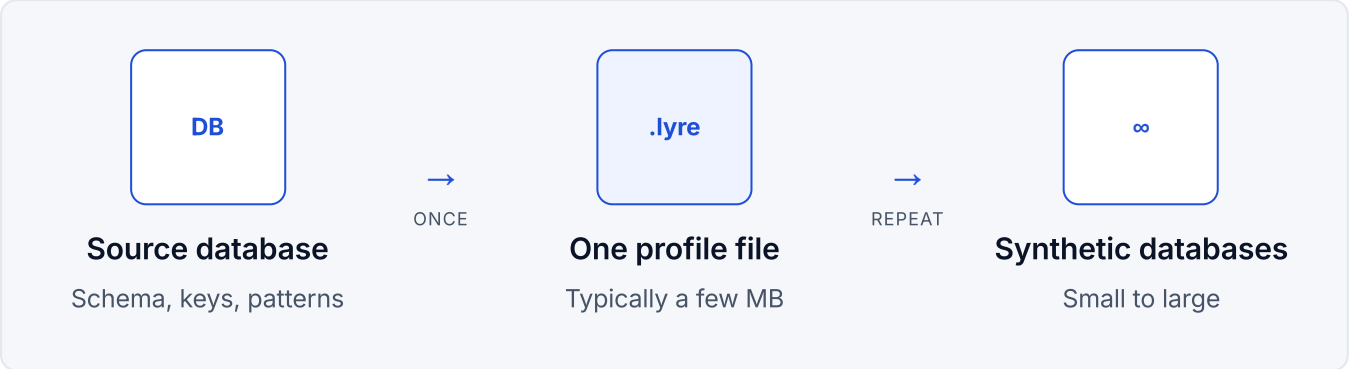
No ticket queue

Developers and CI pipelines create their own databases from a saved profile, in seconds to minutes.

HOW IT WORKS

Connect once. Generate independently.

Lyrebird reads your source database only to build a profile, read-only. Every generation runs from that compact file.



1

Analyze your database once

Connect read-only. Lyrebird captures the schema, primary and foreign keys, table ratios and value patterns in a reusable profile, using a local model.

2

Review and keep one small profile

The profile is a recipe, not a copy. A privacy audit lists what it keeps as written and scans it for personal data and credentials before it leaves your team.

3

Generate at the scale you need

Create a small dataset for unit tests or millions of rows for load tests, from the same profile, on a laptop, a CI runner or an air-gapped server.

4

Keep the profile current, incrementally

When the schema changes, build a new profile version from the previous one. Lyrebird re-reads only the added or changed tables and the tables that reference them; every other table keeps its model.

✓ **Generation never reconnects to the source database.**

CAPABILITIES

Built for real relational schemas.

Generating one realistic table is easy. Lyrebird is built for the hard part: hundreds of tables whose keys, ratios and values must stay consistent with each other.

01

Relational integrity

Parent rows are generated before their children, and every child references a valid parent. Composite keys and foreign-key constraints hold across the whole schema.

02

Fully offline execution

The local model and the generator run on your laptop, CI runner or private servers. No cloud service or external API is involved.

03

Profile privacy audit

Before a profile leaves your team, Lyrebird lists what it keeps as written and scans it for personal data and credentials.

04

Data sized to the job

Scale every table by a factor, from a small 1% copy to a much larger load-test database, or set total and per-table row counts. Table ratios and foreign-key degrees are preserved.

05

Cross-engine generation

Profile PostgreSQL, MySQL, SQL Server or Oracle, then generate for the same engine or another one, for example an Oracle profile into PostgreSQL (SQL Server from SQL Server profiles).

06

Reproducible, ready-to-load output

Same profile, settings and seed give the same data. SQL files with a manifest, native bulk-loader files for SQL Server and Oracle, or a database loaded straight into Docker.

ENTERPRISE SECURITY & DEPLOYMENT

Built to pass your security review.

Lyrebird is designed for regulated teams that cannot move production data into test environments, and cannot send it to a third party either.

Air-gapped by design

Profiling and generation run on your laptops, CI runners or private servers. No SaaS dependency, and no internet connection required.

No external AI API

The model that reads value patterns runs locally. Rows, samples and schemas are never sent to a cloud model.

One read-only connection

The source is read only to build a profile version. Every generation runs from the file, never from production.

Profile privacy audit

Before a profile leaves your team, Lyrebird lists what it keeps as written and scans it for personal data and credentials.

Smaller compliance scope

Test environments hold synthetic rows, not customer records, which narrows the systems in KVKK and GDPR audit scope.

Built for pipelines

Headless CLI for Linux, macOS and Windows, reproducible seeds, and offline licence validation for locked-down runners.

MEASURED, NOT ESTIMATED

Fast enough for every CI job.

A fresh relational test database per job: an empty PostgreSQL in RAM, then schema, rows, constraints and indexes, and a row count check. Public Pagila profile, 22 tables, 36 foreign keys, one laptop, CPU only.

~8 s	< 1 min	~300M	0
1 million rows, ready and checked	10 million rows, ready and checked	rows from five real-world databases	broken foreign keys, nearly 300 checked

FULL-SIZE RUNS ON REAL-WORLD SCHEMAS

DATABASE	TABLES	PROFILE	ROWS	GENERATED IN
Database A	50–100	~2 MB	200M+	~25 min
Database B	200+	~4 MB	50–100M	~6 min
Database C	10–25	< 0.5 MB	5–10M	< 10 s
Database D	10–25	< 0.5 MB	5–10M	~1.5 min
Database E	< 10	< 0.1 MB	~1M	< 5 s

Measured in November 2026 on one laptop (Intel Core i7, 16 threads), CPU only, with a development build made before release 0.5.4. The databases are unnamed and figures are rounded or shown as ranges to protect the source. Foreign keys were checked after loading every table at a smaller size into MySQL in Docker. A shared CI runner has fewer cores, so expect longer times there.

See it before you install it

Four public databases (Pagila, Chinook, AdventureWorksLT and Oracle HR) are compared with their synthetic copies made with Lyrebird 0.5.5, column by column: 0 of 75 foreign keys broken, 102 of 122 columns with the source's distribution, cities, states and countries that agree. **subsetra.com/lyrebird/samples**.

EDITIONS

Evaluate for free. Scale with Enterprise.

Start with Community to evaluate Lyrebird against your own databases. Move to Enterprise when you need larger volumes and pipeline automation.

FREE EVALUATION

Community

For teams evaluating realism, relational consistency and local workflows on developer workstations.

- ✓ Local profiling and generation
- ✓ Up to 5,000 synthetic rows per table
- ✓ Desktop app for Linux, macOS and Windows
- ✓ Profiling, inspection and JSON export
- ✓ 100% offline, air-gapped execution

SCALE & AUTOMATION

Enterprise

For organizations putting high-volume synthetic data into CI/CD pipelines, staging and performance testing.

- ✓ No 5,000-row limit per table
- ✓ Headless CLI for pipelines and CI/CD runners
- ✓ Scaling factors, row counts and seeds
- ✓ Offline enterprise licence validation
- ✓ Direct contact with the engineers who build Lyrebird

See Lyrebird on your own schema.

A 30-minute call with the engineers who build it. We generate from your schema, not a canned demo, and your data never leaves your network.

subsetra.com/request-demo

info@subsetra.com

Download Community:

subsetra.com/lyrebird/samples