IRI RowGen
Smart Test Data Generation

- Prototype DB and ETL Ops
- Improve Application Quality
- Comply with Data Privacy Laws
- Benchmark Hardware and Software

Test Data Generation and Management

Product Summary

What Does RowGen Do?

[IRI RowGen](#) synthesizes safe, realistic test data and populates one or more targets at once, including:

- Referentially correct database (DB) tables
- EDW ETL and Data Vault DB prototypes
- Structured and semi-structured files
- Custom detail and summary reports
- Virtual views and DevOps copies

RowGen reads the metadata of production databases, applications, and file layouts to guarantee realistic data structures, values, and primary/foreign key relationships. While creating that data, RowGen can also transform and reformat it. RowGen uses a powerful Eclipse GUI and self-documenting job scripts to support the re-use and sharing of test data layouts and jobs.

What are the Benefits of RowGen?

RowGen reduces the risk, effort, guesswork, and costs involved in complex test data generation. Designed with developers, DBAs, ETL architects, and compliance teams in mind, RowGen provides an easier way to generate test data sets that:

- Are safe to outsource
- Comply with privacy laws
- Preserve referential integrity
- Conform to your business rules
- Do not tax database resources
- Do not require complex programs
- Cut development and testing time/costs
- Undergo detail and summary transforms
- Scale across an enterprise data warehouse
- Leverage existing data models and file layouts
- Improve quality control via data realism and ranging
- Populate multiple targets simultaneously, including reports
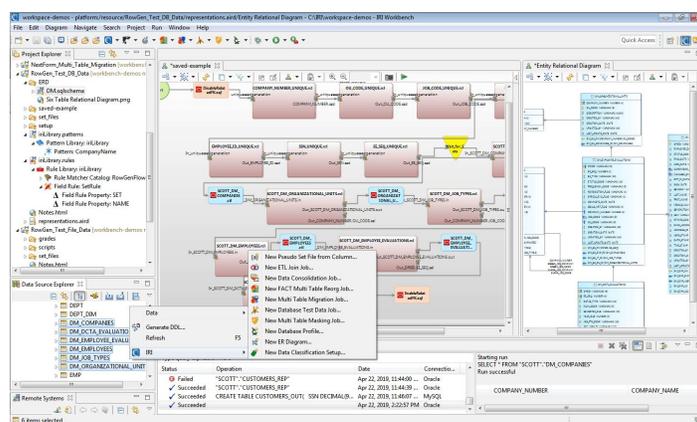
Because RowGen job scripts are based on, and compatible with, CoSort SortCL program metadata, they use the same data manipulation and formatting parameters to customize the look and feel of the test targets. So naturally, RowGen jobs that generate test data can easily switch to CoSort for manipulating, or FieldShield for masking, real data in the same format once testing shifts to production. RowGen also uses the same IRI Workbench GUI, built on Eclipse™, which supports data discovery.

Selected RowGen Uses

Database Population
Rapidly generate and load millions of rows of ordered test data into multiple tables at once, while adhering to their DDL-defined structures and relationships. Populating tables with pre-sorted data also speeds very large load, unload, and query prototypes.

Application Development
Synthesize, transform, and format data in the GUI to meet the volume, range, and layout requirements of any program or prototype. Or, use the Java software development kit (SDK) to dynamically generate test data in Hadoop. Simulate, stress test, and be less dependent on other developers or confidential data.

Format Sharing
Build test targets ranging from simple, single-field flat files to custom-formatted, detail and summary reports. Outsource development and demonstrate programs with realistic data while keeping private data safe. Use computationally valid and realistic format generators for multiple national ID (NID) formats, credit cards, email and street address, phone numbers, etc.

Benchmarking
Leverage fast I/O techniques in RowGen to create massive test tables and files at once. Specify any desired layouts, value ranges, and distributions. Rapidly build scalable data sets for complete and consistent testing of different software and hardware products or configurations.

RowGen Operations

RowGen jobs can rapidly build multiple test data targets in different formats ad the same time. Use RowGen to create safe, realistic data to test relational and NoSQL databases, ETL tools, and applications, or to benchmark or demonstrate systems. Write RowGen Control Language (RCL) job scripts by hand, or create them automatically in IRI Workbench ...

RowGen in IRI Workbench

RowGen is front-ended in a free graphical IDE, built on Eclipse™, called IRI Workbench. In Workbench, you can connect to, search, and classify data in multiple sources, and leverage that metadata in the creation and modification of your test data jobs. Additional plug-ins for Eclipse like BitBucket can pull requests from RowGen project users in Git to manage on-demand needs for test data in the same pane-of glass. Workbench also plays host to every other data management feature in the IRI Voracity platform for ETL, PII masking, data migration, and reporting, plus KNIME analytics, Java development tooling, and much more.

For RowGen specifically, the GUI contains several job wizards and functional dialogs, plus a syntax-aware script editor and dynamic outline for RCL jobs and metadata. For test database creation, a fit-for-purpose RowGen job wizard automates:

- **Parsing** - Select the schema and tables you want to populate. RowGen parses the DDL descriptions and integrity constraints and translates them into RCL scripts that specify the source structure, dependent sets, and data creation.
- **Generation** - RCL scripts are built and run to create one test file per table that can be bulk loaded, or used offline.
- **Population** - Target tables are rapidly bulk-loaded *pre-sorted* in the order necessary to maintain referential integrity.

Another RowGen wizard creates ad hoc test data targets in standalone flat-file, report, or table formats. RowGen also supports: all and valid (joined) pairs, composite data types and values, standard and weighted distributions, and the inclusion of randomized real data to enhance the appearance of test data without compromising security.
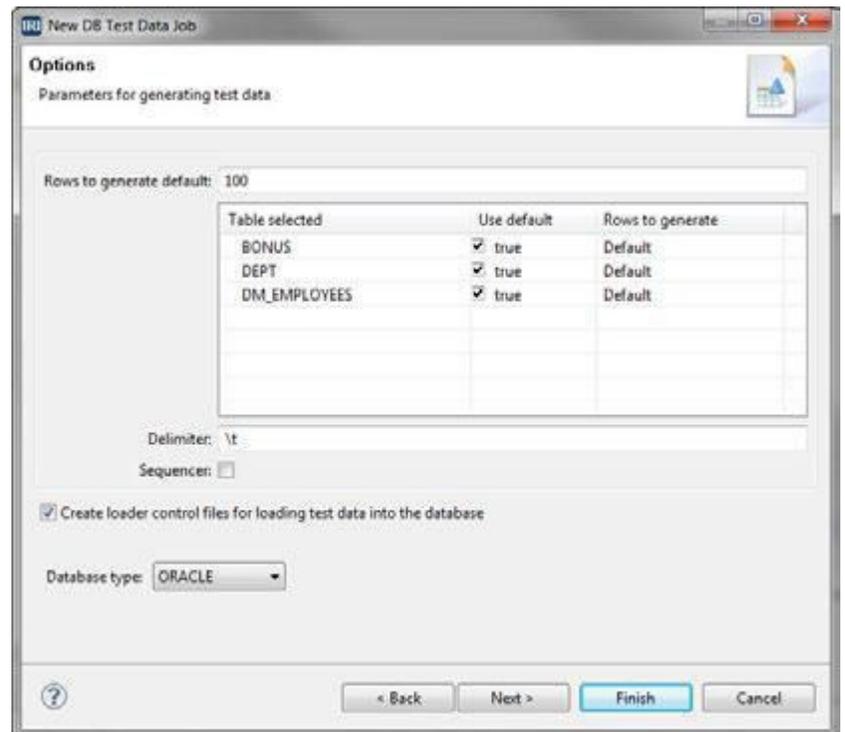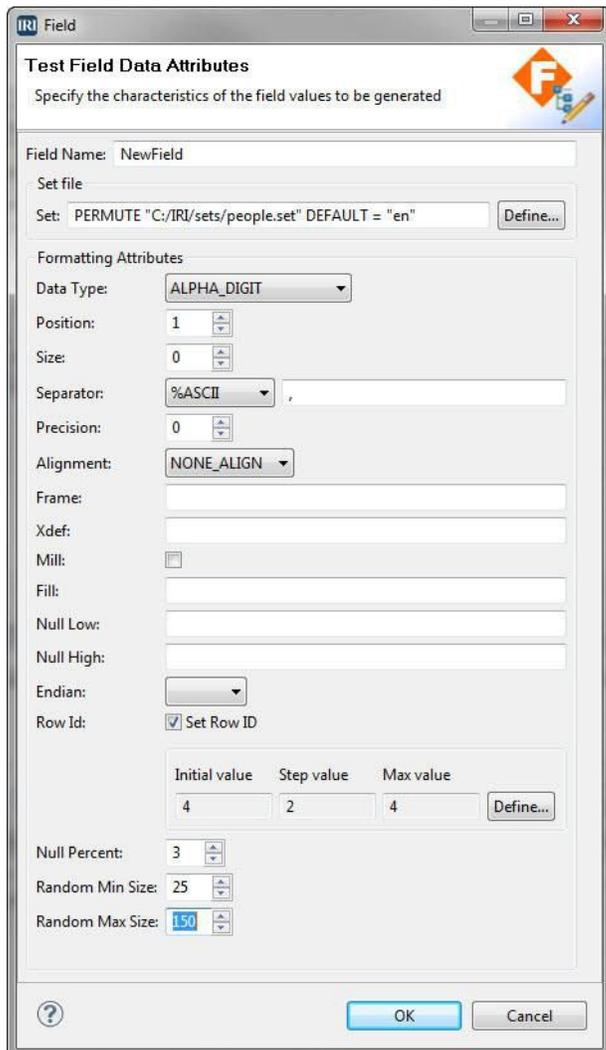
The screenshots left and above illustrate some of the ways RowGen users in the IRI Workbench can specify and modify test data generation. The easy-to-use test DB wizard automates RowGen script creation by directly reding DDL information from any JDBC-connected database.

RowGen also allows you to modify the parameter of generated data using a point-and-click interface, or through a syntax-aware job script editor.
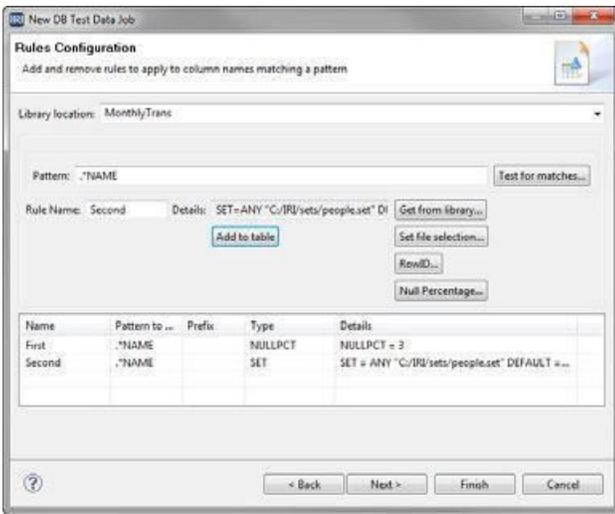
Applying a column selection rule in the IRI Workbench GUI for RowGen.
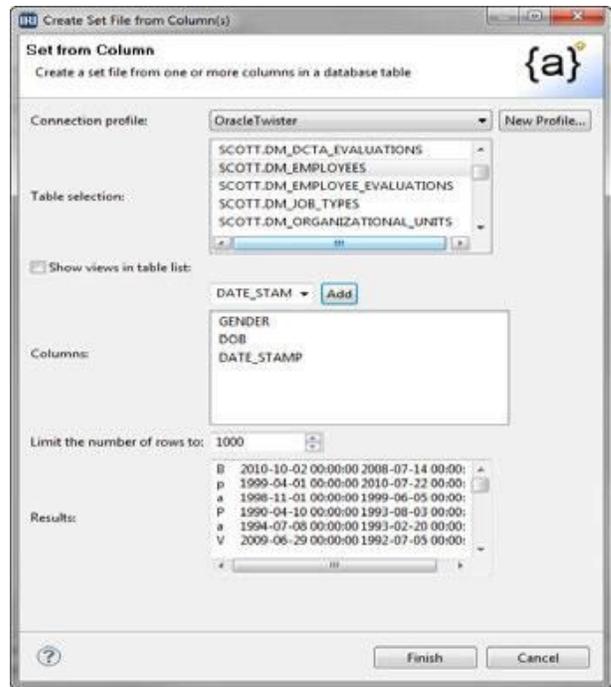
Test-Related Functions

- Random Data Generation and Selection
- Parallel Sorting (speeds loads & queries)
- Conditional Filtering (for value ranging)
- Multi-Targeting (esp. for table relations)
- Calculations, Aggregations & Sequencing
- Frequency Distribution
- Custom Flat-File & Report Formatting
- Logging (in XML format, for compliance)

Test File Formats

- ACUCOBOL-GT Vision
- Fixed- or Variable-Position Flat Text
- FHIR, HL7 & X12 EDI Formats
- Line, Record, or Variable Sequential
- JSON and LDIF
- Micro Focus Variable Length & ISAM
- Mainframe Variable Blocked Format
- MS Comma-Separated Values (CSV)
- W3C Extended (Web) Logs & XML

Test Data Types

- ASCII & EBCDIC Characters
- Digits and Binary Numerics
- Numeric, Whole, Currency, IP Address
- RM and MF COBOL Numerics
- US, European, ISO, Japanese Timestamps
- Unicode & Multibyte Characters

Supported Metadata

- ASN.1 CDRs
- 3NF SQL DDL & Data Vault 2.0
- MF COBOL copybooks
- CoSort (SortCL) data definition files (.ddf)
- LDIF and flat JSON and XML files
- Microsoft Excel, CSV & W3C ELF headers
- Oracle SQL*Loader control files

Supported Platforms

- Unix (AIX, HP-UX, MacOS, Solaris)
- Linux (x86, Itanium, IBM i, p & z Series)
- Windows (7-11, Server 2000-2022)

Compatible Products

- ERWin - Mapping Manager & EDGE
- IRI CoSort - Data Transformation
- IRI FieldShield - PII Classification & Masking
- IRI NextForm - Data & DB Migration
- IRI Voracity - Data Management
- MIMB - Meta Integration Model Bridge
- AWS Pipeline, Azure DevOps, GitLab, Jenkins
- Windocks - Virtual DB Cloning

IRI

Total Data Management

2194 Highway A1A
Melbourne, FL 32937 USA
1.321.777.8889 * 1.800.333.SORT

iri.com/rowgen