# Data Set Utility - Key Features v25.0

## Data Sub-setting and Transportation

- Easy identification and extraction of consistent data subsets (master data and/or transactional data with their details, and referential data when needed).
- Smart extraction and import strategies (no need to disable referential constraints before importing).
- Several data extraction, storage, and transportation methods (direct DML operations through DB links, scripts, RPC, XML, import/export, security policies).
- Preview of data sets via the creation of ad-hoc views.
- Visualisation of the flow of extracted data using a Data Extraction Flowchart (based on Graphviz) to better understand the extraction process.
- Configuration using API's or the DEGPL language.
- Typical use cases: transfer your data between environments; back-up, delete and restore your data; restore your data after a destructive test; replicate your data; etc.

## Change Data Capture (CDC)

- Capture all DML operations performed on a set of tables (as XML, via triggers).
- Rollback (undo) or roll forward (replay) captured operations.
- Replication of captured changes synchronously or asynchronously via a job.
- Generation of redo and undo scripts.
- Typical use case: rollback committed changes made by a destructive test; keep two environments synchronized; feed a data warehouse with changes made to an operational database.

## Sensitive Data Discovery

- Library of pre-defined sensitive data types and their search pattern.
- Multi-criteria search on column types, names, comments, and data.
- Discovery based on regular expressions and data set look-up.
- Detailed execution report of the sensitive data discovery process.
- Default data masking provided for all sensitive data types.
- Pre-defined look-up data sets.

## Sensitive Data Masking

- On-the-fly data masking (no need to make a copy of your production data).
- SQL-based masking with a library of pre-defined functions.
- Large choice of masking functions: randomise, obfuscate, mask, encrypt, and decrypt.
- Shuffling of column (or group of columns) values, with or without partitioning.
- Tokenization with secured access to tokenized values (encryption).
- Generation of surrogate keys based on in-memory sequences.
- Masking of unique identifiers and auto-propagation to foreign key columns.
- Support of format preserved encryption (FPE).
- Preview of data before and/or after masking.

# Synthetic Data Generation

- Smart generation strategy with no need to disable referential constraints.
- Visualisation of the flow of generated data using a Data Generation Flowchart (based on Graphviz) to better understand the generation process.
- Generation of reference, master/detail, hierarchical and historical tables.
- Generation of unique identifiers using Oracle sequences.
- Generation of lookup foreign keys (with possible filtering) to ensure integrity.
- Generation of column values via user-defined SQL expressions.
- Rich library of random functions (with seed to make them deterministic).
- Ability to generate random values compliant with regular expressions.
- Ability to generate random values from weighted lists of values or CSV data sets.
- Ability to generate complex use cases (e.g., historical tables) via generation views.
- Ability to compute fields denormalized between tables via post-generation code.
- Verification of generated data via the creation of ad-hoc views.
- Configuration using API's or the DEGPL language.