



## Fast Data Masking™ Technical Product Sheet

## The Struggle with Sensitive Data

The growing number of high-profile incidents, in which sensitive and confidential data has been lost, stolen or leaked, has created concern within the industry. As strict regulations and guidelines, such as the HIPAA and GLBA (USA), PCI DSS and EU Data Protection Directive, continue to put pressure upon organizations, greater consideration is being given to managing patient health records, HR data and financial data. Data breaches can no longer be swept under the rug, and organizations are looking for data masking solutions to help them protect their sensitive records for use in non-production.

Data masking is used to de-identify specific data within a database table or flat file, ensuring that data security is maintained and sensitive customer information is not accessible to non-authorized individuals. By applying masking algorithms across all data tables, sources and executions, referential and business integrity can be maintained. This means you can ensure regulatory compliance, while still provisioning usable, meaningful sets of data.

## The Fast and Easy Solution

Fast Data Masking™ allows you to mask, reduce, compress and reorganize your test data in one operation. Powered by our award-winning Datamaker™ solution, it uses a powerful array of industrial-strength data masking functions, seed tables and high performance techniques to provision fully compliant, high-quality, meaningful test data for use in non-production.

Fast Data Masking™ uses native database utilities to generate masking scripts, allowing users to mask large volumes of data quickly and easily, and ensure full regulatory compliance.

### *Data Masking*

*You can assign a rich variety of out-of-the-box data masking functions to columns within a database and either extract clean data sets for external usage OR update the data in situ.*

### *Data Compression*

*As the masked database is built the data can be compressed, reducing your data footprint by up to 80%.*

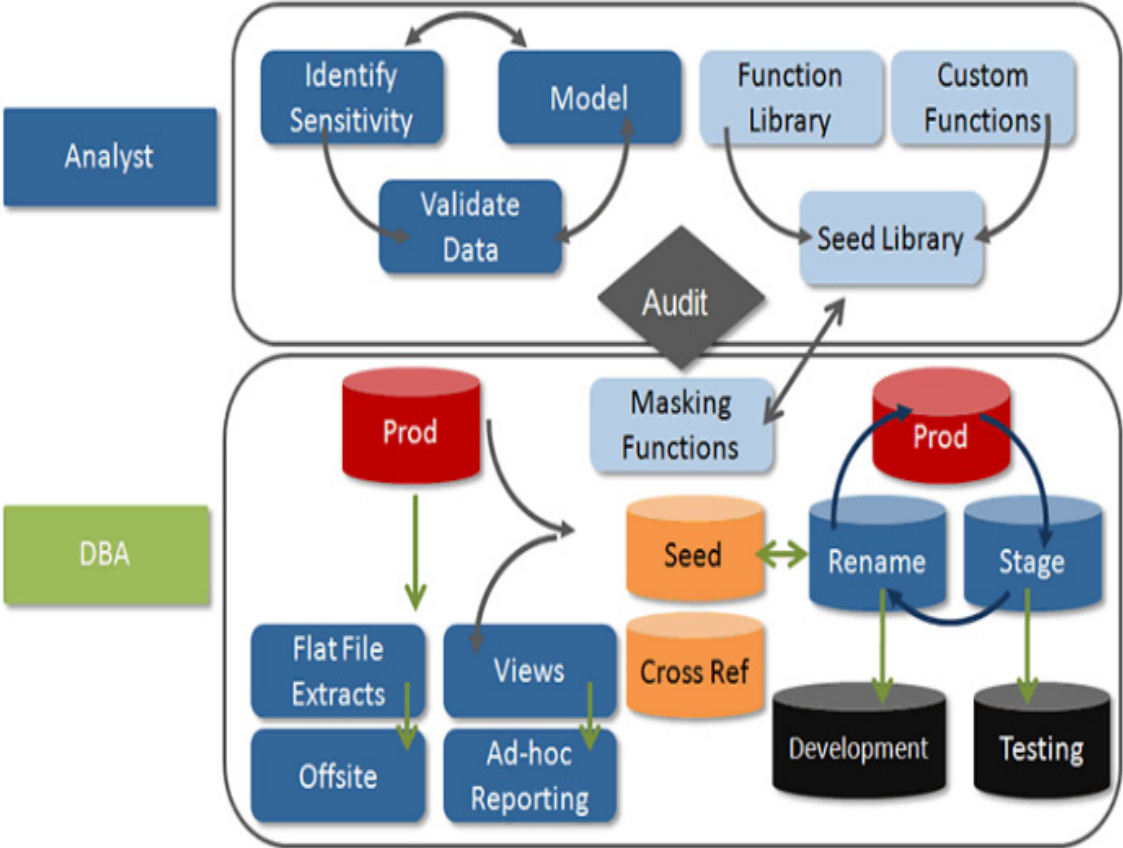
### *Data Reorganization*

*As the masked database is being built, reassign tables and indexes to new tablespaces, allowing DBAs to balance their database disk access.*

### *Data Reduction*

*A powerful database subsetting tool allows you to selectively extract sets of rows and their related tables. For example, only use a 5% sample of active customers and their associated tables.*

# Fast Data Masking™ Architecture



Data can be extracted to flat file for use offsite, or it can be masked, reorganized and compressed in situ, using a combination of high performance techniques.

Fast Data Masking™ allows systems analysts to choose from a rich array of standard functions and standard seed tables, complemented by their own custom masking functions. Once these rules have been defined, a powerful set of scripts are generated to quickly extract the data, or mask it in situ. All Fast Data Masking™ scripts use standard, native database utilities and techniques to maximise performance,

## High performance techniques include:

**Database subsetting** - reduce your data footprint and time to mask by ignoring old and redundant data.

**Unix parallelism** - Use the power of Unix to submit multiple processes at the same time.

**Oracle parallelism** - Use Oracle's parallel load facility to maximize data loads.

**Staging tables** - Performing large updates in situ can be very slow. Using shadow tables and renaming them back to the original name is much faster.

**Logging management** - Reducing logging during the masking process will reduce the run times.

**Optimized masking functions** - Ensuring that the minimum amount of code to mask the data is executed.

## Management features include:

Managing constraints and foreign keys

Recompiling affected objects

Reorganizing data into new tablespaces

Applying compression to reduce space and improve performance

Rebuilding Indexes

Rebuilding Catalog statistics

Securing the database in exclusive (secured) mode during masking

Preview data before masking

Design and build once – run many times

No license restrictions on the number of target databases and copies of the masked data

Strong version and audit control

Source control is managed by your existing source management systems

The same rule set can be used to generate synthetic data directly into testing and development databases

One tool can be used for all your other data sources: DB2, SQLServer, Flat files, etc.

# Performance: Example Runtimes

## FOR ORACLE

Linux x86 2x4 CPU Poweredge T605 Opteron 2.1 GHz
Memory 6 Gig
632 Columns masked across 180 tables
26 Different masking types, 20 seed tables
60 Gig database 300 million rows
<b>Total Time: 45 minutes</b>

## FOR SQL SERVER

Task	Time
Generated 100,000 de-identified insert statements	40 secs
To load insert data	3m 40secs
To update 100,000 rows with scrambling	80 secs

100 Million row table
6 columns masked
Random, Number and Text
<b>Total Time: 58 minutes</b>

## About Grid-Tools

Grid-Tools are the leading test data management vendor internationally, specializing in data generation, test data management and provisioning of agile data for Agile environments.

We have revolutionized the way test data is being provisioned by developing innovative software solutions for synthetic test data creation, data masking, subsetting, design and archiving.

Our innovative solutions, Datamaker™ and Enterprise Data Masking™, offer companies of any size and market sector the ability to provision high quality test data for their testing, development and training environments, as well as outsourcing both off-shore and near-shore.

Grid-Tools have been working with some of the largest government agencies, financial institutions, telecoms, insurance providers and media and communications corporations around the globe. We strive to promote better processes and best practice methods for provisioning compliant test data that is fit for purpose.