Bloor
InBrief

Daniel Howard – Senior Analyst, Information Management

redgate

Redgate Test Data Management

The company

Redgate provides software that enables what it calls *'Compliant Database DevOps'*, meaning that it helps you to leverage your data inside a DevOps toolchain in a way that is compliant with all applicable regulations and policies. The majority of Redgate's products work exclusively with SQL Server, although several are also available for Oracle and MySQL databases. Moreover, many Redgate products support Azure and AWS environments in addition to on-premises SQL Server. In 2017, Redgate acquired Net2000, a data masking vendor, along with their products DataBee (now discontinued) and Data Masker.

The company was founded in 1999 and is based in Cambridge, UK. It has additional offices in the US, Germany and Australia. Redgate's customer base numbers over 200,000 and includes 91% of the Fortune 100. It is also a gold partner with both Microsoft and Oracle.
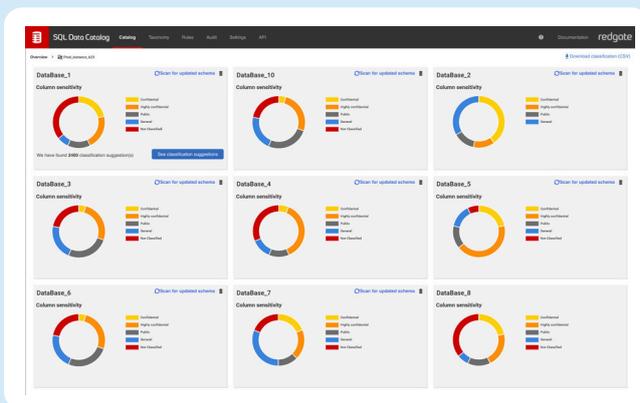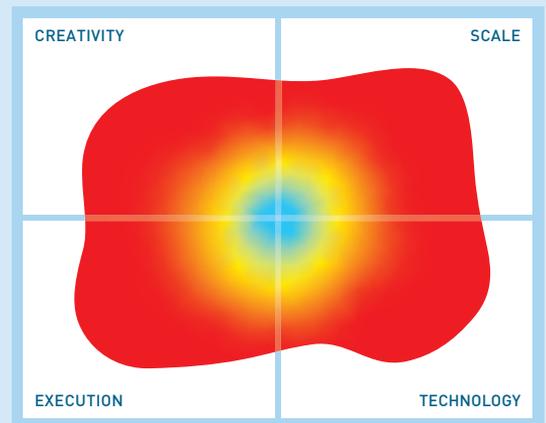

CREATIVITY	SCALE
EXECUTION	TECHNOLOGY

The image in this Mutable Quadrant is derived from 13 high level metrics, the more the image covers a section the better. Execution metrics relate to the company, Technology to the product, Creativity to both technical and business innovation and Scale covers the potential business and market impact.

Figure 1 – SQL Data Catalog landing page

What is it?

SQL Provision is a solution for (compliant) test data management that combines two Redgate products into a single offering: Data Masker for static data masking, and SQL Clone for database cloning and provisioning. SQL Data Catalog, shown in *Figure 1*, is also available as an additional offering that adds

data discovery and cataloguing. Taken together, Redgate allows you to create and distribute masked, production-like copies of your data wherever it is required within your existing processes and CI/CD pipelines. In other words, it provides a complete test data management solution.

SQL Clone and SQL Data Catalog run exclusively on SQL Server (although we are told that integration with other platforms is in development), while Data Masker will run on either SQL Server or Oracle.

What does it do?

Data Masker (as seen in *Figure 2*) is a rules-based static data masking product that performs on the level of millions of rows per hour. It maintains the credibility of masked data by ensuring that correlated values (for instance, age and date of birth) remain consistent using substitution rules and correlated data sets that contain potential values to substitute. These are available out of the box or can be user defined, and this process can be actioned between databases or even database instances. Data Masker also has the ability to generate synthetic data in a limited capacity, best utilised when production data is either unavailable or incomplete.

Masking in Data Masker is irreversible, always retains relational integrity, and can mask primary

Data subsetting	N/A	Data masking	★★★★
Synthetic data generation	★★★	Automation	★★★★
Database virtualisation	★★★★	Ease of use	★★★★
Data discovery	★★★★	Integration	★★★

or foreign keys without a join operation. It automatically generates reports whenever a masking rule is run, making the masking process fully auditable, and allows you to either apply your masking rules immediately or export them for use elsewhere (in SQL Clone or the SQL Data Catalog, for example).

Data Masker provides some basic sensitive data discovery functions, but in practice the meat of this is in SQL Data Catalog, which allows you to define your own taxonomy of classifications that your data can be matched against using an extensible library of built-in pattern matching rules. You can then attach actions to these tags in order to, for instance, automatically mask data that is classified as sensitive. To that end, compatibility with HIPAA, GDPR, and other regulations is provided out of the box, as is integration with various test data management processes.
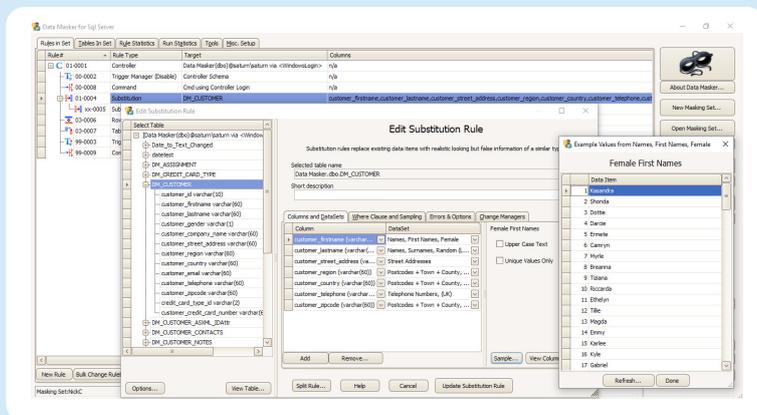

Figure 2 – Editing a masking rule in Data Masker

SQL Clone allows you to create and centrally manage images and virtualised clones of your production data. Images are complete point-in-time copies of a database, taken from either a live server or a backup. As they are often quite large (since they are complete copies) they are usually stored centrally. You can modify your image during its creation using either SQL scripts or sets of masking rules exported from Data Masker.

Clones, on the other hand, are derived from an image and only store the differences between themselves and the image they were derived from. Due to this, they are small in size (usually less than a hundred megabytes) and can be created

very quickly and wherever they are needed. For test data management, this means that your testers can provision a masked clone to their local machines whenever they need test data without having to wait on an administrator. Team management, permissions, and self-service features are provided to this end, as is integration with Git. PowerShell-driven automation is also available for image creation and management. By leveraging SQL Data Catalog, you can even distribute clones (and thus test data) as an automated part of your business processes.

Why should you care?

SQL Provision and SQL Data Catalog together provide a complete test data management solution that includes test data discovery, masking, and provisioning (via database clones) and can readily enmesh itself into your existing processes. Cloning in particular offers significant advantages over both data subsetting and synthetic data generation: for example, it effectively guarantees that your test data will be representative.

Redgate also makes distributing that test data fast and easy. This is always a boon, but it is particularly important for enabling high-speed testing that can easily fit into (and keep pace with) a DevOps pipeline. Moreover, access to self-service and team management features help to make life easier for your developers, testers and DBAs alike, and integration with Git allows you to automatically build, validate and deploy an appropriate database clone whenever a pull request is made. In turn, this means that you can review the request against a live environment that implements it, instead of just raw code. This all works together to help facilitate a *"shift left"* approach to testing, allowing you to test earlier, faster, and altogether more smoothly.

The Bottom Line

SQL Provision and SQL Data Catalog form a capable, comprehensive, and cost-effective test data management solution. If it weren't for its limited database support – its full range of functionality is only compatible with SQL Server – we would be singing its praises unreservedly. If that's not a problem for you, Redgate should definitely be on your shortlist.

FOR FURTHER INFORMATION AND RESEARCH CLICK <u>HERE</u>