Data Masking 2017

MarketUpdate

The market

This is our third Market Update in this space and we already noted in the 2nd edition that the market had started to diverge. Data masking originally emerged as a complement to test data management, to protect sensitive data from unauthorised eyes (assuming that you were not using synthetic test data generation). However, it has always been, in essence, a security technology. What has been happening in the market is that some vendors – not all – have been transitioning their offerings from security for non-production to production data and this is taking them squarely into the security sector. With the advent of the EU's GDPR (general data protection regulation) and other legislation, this trend has been accelerating and we now see some vendors focused specifically on non-production data (mostly testing) while others are wholly devoted to security: so-called data-centric security. Suppliers in this market often have much broader security capabilities. Some companies, of course, are active in both markets.

It is worth discussing a number of points. To begin with it should be recognised that all the leading vendors can do more or less whatever you need when it comes to static data masking algorithms. To this extent, the market is commoditised. Some vendors may offer more masking algorithms than others but the difference between 54 algorithms and 84 is really very small, provided that you can do basic things like preserving referential integrity and consistent masking across multiple data sources.

One area where there are some significant differences between vendors is that there is an increasing recognition that conventional data profiling (primarily recognising that data matches a particular pattern such as a credit card number), while possibly adequate for non-production data, is insufficient for true security environments. For example, there is a 70% chance that any random collection of nine numbers will fit the format for SSN. Thus, there is a major emphasis on reducing false positives and, indeed, false negatives. A variety of techniques are being employed by the more advanced vendors for this purpose, including semantics and code (PL/SQL, for instance) introspection. Suppliers are starting to implement machine learning so that discovery techniques get more efficient over time. As an example of the sort of difference this can make we know of one user that replaced a data masking vendor that was limited to traditional profiling with an alternative that had more advanced capabilities: the result was that while the user had previously thought that he had 500 sensitive columns in his database he found that in reality there were 1,900.

Another trend we are seeing from both users and vendors is towards supporting the masking of unstructured data. This has two aspects. Firstly, the masking not just of things like pdf documents and text files but also formats such as SWIFT and EDI. CA has supported the masking of the latter for a long time but most other vendors have not. Secondly, there is masking for NoSQL databases such as Hadoop. Some suppliers support the masking of data in multiple NoSQL sources, others merely have this on their roadmaps.

Figure 1: The highest scoring companies are nearest the centre. The analyst then defines a benchmark score for a domain leading company from their overall ratings and all those above that are in the champions segment. Those that remain are placed in the Innovator segment if their innovation rating is over 2.5 and Challenger if it is less than 2.5. The exact position in each segment is calculated based on their combined innovation and overall score. It is important to note that colour coded products have been scored relative to other products with the same colour coding.

Two trends that we reported in our last Market Update are support for location-based capabilities and embedding within other products. In the first case, this would mean that you can see whatever you are entitled to see if you are in the office, but if you are working from home then sensitive data will be masked, and if you are in a different country then you will not be able to see any version of this data. Bearing in mind the different national laws regarding data security this makes a lot of sense and we expect other vendors to follow suit with similar capabilities. In the second case, this would mean embedding in data preparation tools, as an example. While there were masking vendors that had introduced these capabilities two years ago, we have only seen a slow introduction of comparable facilities from other providers.

Dynamic data masking

The previous discussions have revolved around static data masking. However, we should also consider dynamic data masking. This is mostly achieved by intercepting SQL calls to the database and then acting appropriately (blocking access, masking the data, and so on), most commonly (though not always) by using a database proxy. In general, this sort of masking is used in read-only mode and not for things like updates or deletes, because of the performance impairment it implies. As a result, vendors have been exploring less performance heavy technologies and we are seeing more and more suppliers adding format preserving encryption (FPE) to their offerings, especially now that there is a NIST standard for FPE. In addition to better performance, FPE is also useful when it is necessary to be able to reverse your masking. Conversely, some jurisdictions mandate that reversible methods are not permitted (in relevant circumstances), so that conventional masking is required in addition to FPE.

In practice, dynamic data masking is less widely used than static data masking. There are a number of reasons for this in addition to performance considerations. These include the workload involved in identifying and administering user rights, the risk of writing masked data back to the database, and the fact that data in the database is not masked, so you cannot completely nullify the possibility of an insider accessing that data. While dynamic data masking may seem simple, in reality that is not the case.

GDPR

The General Data Preparation Regulation is driving considerable interest amongst vendors. This applies to any company trading within the EU with hefty penalties (up to 4% of global revenues) for failure to comply in the event of a breach. While this is not the place to discuss the regulation in detail it is worth reflecting on the fact that the regulation allows for the pseudonymisation of personal data for, for example, analytic purposes. Such data is not usually considered to be within the domain of DevOps or test data management, so it is effectively production data: but production data to which static data masking should be applied rather than dynamic data masking. FPE may be appropriate in some instances – for example, fraud analytics, where you may need to know who you have been investigating – but often reversible approaches such as FPE will not be permitted (though sometimes it is mandated). All in all, GDPR is attracting considerable interest from vendors. Moreover, it is also a challenge for those companies that have previously focused only on test data, because the decision makers are likely to be different when it comes to the pseudonymisation of analytic data.

In so far as adoption of GDPR is concerned, we have seen the most interest from companies in the UK. Organisations in other countries seem to be taking a less urgent stance to what needs to be done, either because they have a relaxed attitude to such regulations or because they think (erroneously in our opinion) that they have done it all already. We suspect that the EU will apply some pretty swingeing fines as soon as its gets the opportunity, pour encourager les autres.

Vendors

With a few exceptions, vendors either tend to focus on test data, or security more generally. The former typically offer robust test data management capabilities while the latter provide a broader security portfolio without test data management. The major exceptions are Informatica and IBM, which are active in both markets, as are Mentis and Solix.

It is interesting to note that one of Bloor Research's competitors has ceased providing comparative analyses to data masking and now only provides a Market Report. The reason is ostensibly because it sees this market as mature. In so far as masking per se is concerned (especially masking of structured data), we agree but, as discussed, there are other surrounding elements which are not nearly as mature. Nevertheless, we are starting to see consolidation in this sector. Axis Technology was the first to go, acquired by Delphix prior to the last edition of this Market Update. Since then, Grid-Tools was acquired by CA, Privacy Analytics was acquired by IMS Health

(which subsequently merged to become Quintiles IMS), Hexatier (previously GreenSQL) was acquired by Huawei and, most recently, Camouflage Software has been acquired by Imperva. On the other hand, new entrants to the market – such as DatProf – are still emerging.

The one product that we have knowingly omitted from this paper is Oracle Data Masking and Subsetting, because this requires you to stage data into an Oracle database instance, in order to process the data. Even if you are an existing Oracle user there is a significant downside to this – performance, cost, administration, time – unless you use no other database apart from Oracle. We have also not included Hexatier (h) because the company has the following statement on its website: *"We changed ownership as of 27 Dec. 2016. We are working on the next generation. The existing product is not for sale any more."*

Summary

There is a clear split between companies focusing on DevOps and other non-production environments and those that are taking a broader security perspective. This actually makes direct comparisons between products and suppliers difficult: many of the solutions on offer will be selected as much for the complementary capabilities that are offered as for the product's pure masking capabilities. In the diagram on page 1, we have therefore colour-coded vendors with different emphases. Note that products focused on test data can be used with non-test data and vice versa, but the nature of the relevant company's other products, and salesforce orientations, means that they are likely to be focused in one area or the other. IMS Privacy Analytics and Delphix each offer (different) unique sets of capabilities so their scores are not easily comparable with that of any other vendors.

Bloor

Bloor Research International Ltd
20–22 Wenlock Road
LONDON N1 7GU
United Kingdom

Tel: **+44 (0)20 7043 9750**
Web: **www.Bloor.eu**
Email: **info@Bloor.eu**

CA

CA offers a data masking product, CA Test Data Manager (TDM), which is a comprehensive test data management solution that provides both data masking and synthetic data generation, along with many more test data management capabilities.

CA TDM provides native drivers (for DB2 z/OS, IMS, VSAM/ISAM, Oracle, SQL Server and Teradata) and makes use of native database utilities where possible. It can mask over 100 million rows of data per hour and can be used not only with data in a database, but also with Excel, XML, CSV, TXT, EDI, SWIFT and fixed definition files. It supports HIPAA 40-10, 50-10 and X12 formats.

CA TDM also has some, limited, dynamic masking capabilities based on the use of views. More generally, CA TDM provides a rich suite of data masking functionality, including over eighty distinct masking options. The software automatically discovers data that looks like it might be sensitive, keeps the data referentially intact, is fully auditable, can mask data either in place or in flight, and is demonstrably compliant with EU GDPR, GLBA, HIPAA, PCI DSS, PIPEDA and other regulations.

CA TDM also supports the generation of synthetic data (for which masking would not be required), and integrates with CA Agile Requirements Designer and other products within CA's DevOps product line. As this product range, might indicate, CA addresses masking specifically from the point of view of testers and developers in support of non-production data.

Strengths

- CA's data masking capabilities are exceptionally complete: you could hardly ask for a greater range of masking functions.
- CA's support for masking unstructured data formats such as EDI and SWIFT is well ahead of the majority of its competitors.

Threats

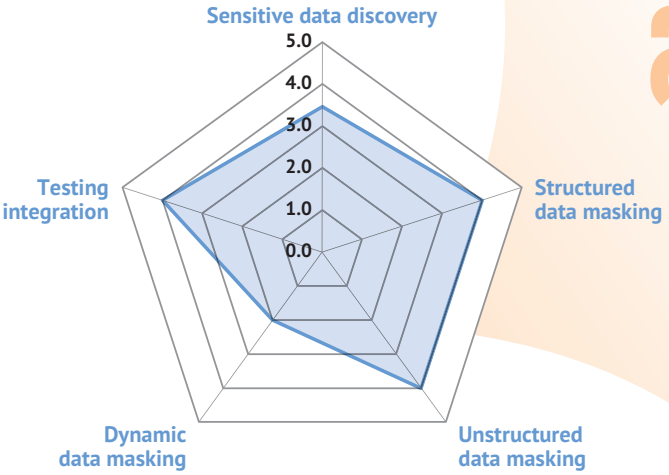
- While you can use CA Test Data Manager for production as well as non-production data, CA's focus on DevOps means that it may lose out with companies that want static data masking but for non-test data (for example, analytic data).
- Unlike a number of competitive products CA Test Data Manager lacks integration within a broader security environment.

CA Technologies
Islandia, USA

www.ca.com

Summary

CA is a leading provider of data masking solutions and its score reflects this. However, its strength is particularly within testing environments and there are growing demands for (static) data masking in other environments, especially analytics.



Compuware

Compuware's Test Data Privacy solution provides data masking capabilities. The solution leverages File-AID, its file and data management solution, and Topaz Workbench, their Eclipse-based IDE. Topaz Workbench provides a common framework and integrated user interface from which to initiate Compuware's array of mainframe development, testing and maintenance tools and is available at no additional charge to maintenance-paying Compuware customers.

Compuware Test Data Privacy provides the ability to identify and mask sensitive data. The interface within Topaz Workbench is used to create and manage privacy definitions. This tool allows you to abstract each type of data in your database (for instance, 'name' data) into a data element that can then have masking rules applied to it. Fields are matched to the data elements (by matching the field name) and the disguise rules defined for the data element are dynamically built and applied at execution time. A coverage view is available that will display which fields are mapped to which data elements, allowing a user to adjust them if appropriate. Relational integrity is always maintained and an audit log is generated upon execution. Compuware's patented Composite Processing finds data within a larger field and ensures that differently formatted values will be properly masked or privatised. For instance, it will recognise that "Mary Jane Smith" and "Smith, Mary Jane" are the same name and mask them appropriately (for instance, to "London Sara Klein" and "Klein, London Sara", respectively). Compuware's Test Data Privacy solution disguises data where it resides, either on the mainframe or in the distributed files or databases in which it exists.

Additional data management functionality is available from Compuware to browse and edit data in a variety of preset or custom layouts (including raw data, for those who need it); as well as comparing and subsetting data, visualizing data relationships, visualizing extract executions, and copying data from one mainframe LPAR to another.

Compuware
1 Campus Martius, Detroit, Michigan 48226
www.compuware.com

Strengths

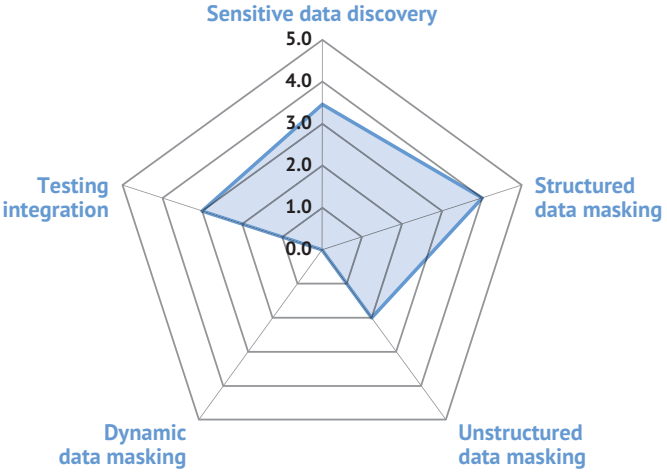
- Test Data Privacy leverages Topaz Workbench, an Eclipse-based IDE, which makes it very easy to use. This is especially impressive given that the product can be used to mask both mainframe and distributed data types. This multiple-environment support also means that masking can be deployed in a consistent fashion across data sources.
- The abstract nature of data elements for defining and applying privacy rules makes it easy to mask several different but conceptually similar fields using the same set of rules. More to the point, it minimises maintenance issues if those rules are ever changed.
- The fact that Topaz Workbench is free of charge to existing, maintenance paying File-AID users is a major plus point.

Threats

- Matching fields to data elements based on their names is not always ideal. In practice, fields tend to be named arbitrarily and not necessarily with forethought. This is mitigated somewhat by the presence of the coverage view.

Summary

If you are a File-AID user, then employing Topaz and the Compuware Data Privacy solution should be a no-brainer. Conversely, if you do not have a mainframe, or are unconcerned about masking mainframe data, then Compuware will not be for you. If you are a mainframe user that does not employ File-AID, you would probably be best to consider Topaz first and data masking second.



Dataguise

Dataguise, which was originally a spin-out from Oracle, started life with a conventional static data masking tool with sensitive data discovery capabilities for conventional data sources. It has subsequently extended its capabilities to NoSQL database environments supporting not just Hadoop but also Cassandra, MongoDB, Redis, Apache Hive (a first, as far as know) and others. This range of support is relatively rare within the data masking market. The company supports both on-premises and cloud-based deployment. The company's overall product is DgSecure and it has four elements: Detect, Protect, Audit and Monitor.

Discovery (Dataguise Detect) is based on natural language processing in conjunction with pattern recognition and it can discover sensitive data both in on-premises and cloud-based data stores. Notably, in the latter category, the company has just announced sensitive data discovery for Google Cloud Storage. A major problem with discovering sensitive data is that you can get a lot of false positive and negatives. Dataguise has addressed the former issue by building machine learning into its product, learning initially from sample data or from examples of false positives. The company also provides features to reduce the number of false negatives. Facilities include support for industry or customer specific ontologies.

As far as Dataguise Protect is concerned both static and dynamic masking are provided as well as (format preserving) encryption. Both full and partial redaction is possible and the product supports masking for both structured and unstructured data. In the case of dynamic masking the company leverages native capabilities for this purpose (for example, Cassandra's API). The Audit and Monitor capabilities provide policy-driven monitoring (in real-time) and recording of who accessed data, when, where and what they did with the data. Display is via a persona-based dashboard.

To support GDPR, Dataguise is extending its software to support more European languages.

Strengths

- The facilities provided to reduce the number of false positives and negatives – a major bugbear when it comes to sensitive data – is market leading.
- The implementation of machine learning by Dataguise is ahead of most of its competitors.

Dataguise
2201 Walnut Ave. #260
Fremont, CA 94538, USA

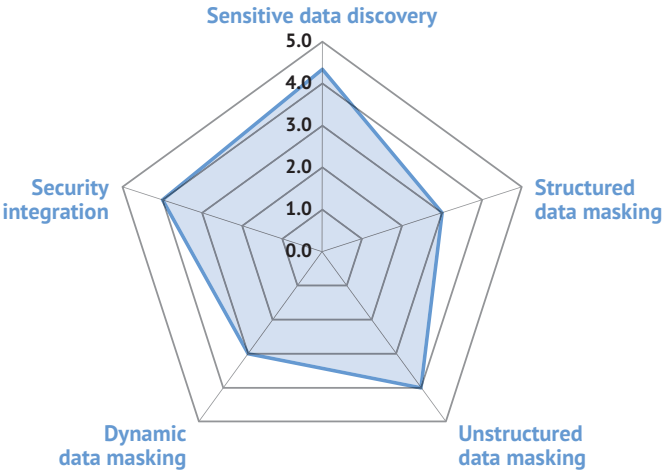
www.dataguise.com

Threats

- While the range of support for both SQL and NoSQL databases, and for both on premises and cloud-based data stores is extensive, there are unstructured sources in particular (for example, EDI and SWIFT) that are not supported.
- The company does not offer test data management (TDM). While Dataguise has clients using DgSecure in conjunction with test data, the lack of TDM will leave it at a disadvantage when competing with suppliers that offer both TDM and data masking.

Summary

Dataguise, in terms of its features and capabilities, is one of the leaders of this market: a significantly improved position compared to our last Market Update on Data Masking. We have scored it as equal first in this Market Update. The company's support for NoSQL is exceptional.



DATPROF

DATPROF's principle data masking product is DATPROF Privacy. In addition, DATPROF Analyze (which is currently in beta) is used in a supporting role to discover and profile sensitive data. Both of these tools can work with (in theory) any source database but require an Oracle, SQL Server or IBM DB2 database to process test data, although only metadata is ever actually extracted from your system. The area where DATPROF most excels is ease of use: both DATPROF Privacy and DATPROF Analyze are exceptionally easy and intuitive to work with, and don't require any significant training. This means that you can discover and mask your data easily and – more importantly – quickly.

From a functional point of view, we would say that DATPROF's masking capabilities are good without being exceptional. As we have stated, ease of use is the primary selling point, and in this respect DATPROF's products are exceptional. We also particularly like the custom masking rules, constraints on existing rules and rule dependencies offered by the Privacy product. It also maintains its own audit log. Lastly, Privacy completes the masking procedure by generating and running a SQL script, ensuring that it remains performant.

Strengths

- Ease of use and user experience are exceptional, and a cut above competing products in the space.
- An important feature to emphasise is that DATPROF never physically extracts data, only metadata.

Threats

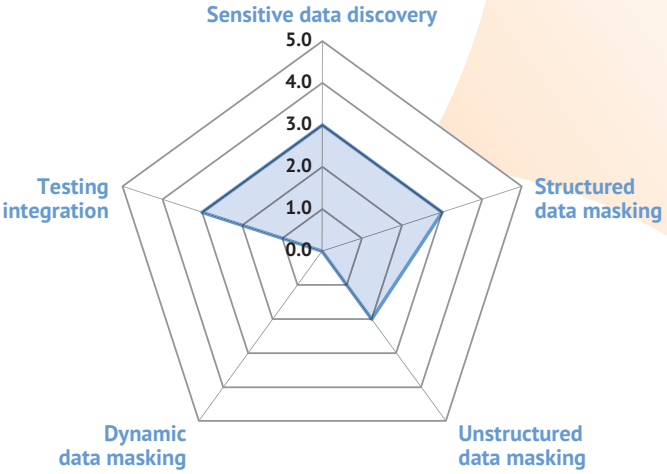
- DATPROF does not have the bells and whistles that some other vendors offer. For example, it doesn't support 80 different masking methods. On the other hand, do you really need such complexity?
- DATPROF is a relatively small company and not well-known outside its native environment (the Netherlands).

DATPROF
Friesestraatweg 211, 9743 AD Groningen
The Netherlands

www.datprof.com

Summary

The biggest selling point for DATPROF is its ease of use. It is also likely to be considerably less expensive than competitive offerings that include more features. This tends to suggest to us that the company's probable target customers would be in the mid-market. However, its customer base to-date suggests that it is winning enterprise clients, especially in conjunction with the company's test data management offering DATPROF Subset.



Delphix

The main focus of Delphix is enabling on-demand access to data. It does this by means of its dynamic data platform. Delphix holds a single, continuously updated, copy of your production databases (including all binaries, configuration files and so forth). It can then provision complete virtual copies of this data, as required. The automation and self-service that is provided fits well as part of agile and DevOps initiatives, with features that include the ability to refresh, reset and rewind data, as well as the ability to bookmark, branch and share data. Different databases can even be provisioned in a synchronised fashion. Delphix software can be hosted on-premises or in the cloud (AWS or, shortly, Microsoft Azure) and leverages your existing storage. Supported data sources include Oracle, MS SQL, Sybase, PostgreSQL, MySQL, DB2 as well as file system data, and the company is adding support for other Amazon supported databases such as MariaDB. Support for other data sources is under consideration.

If your test data is potentially sensitive (discovered through using the Delphix Data Profiler) then it will need to be masked, and in 2015 Delphix acquired one of its partners: Axis Technologies, which specialised in data masking. How this works is that you take your copy of the production database and then you create a masked virtualised version of that and then provision virtual copies from that source. As far as features are concerned, whilst Axis was targeted at the mid-market, Delphix has now successfully deployed their masking in their enterprise customer base. Facilities include a large range of out-of-the-box algorithms, including deterministic masking to ensure that common fields in different data sources are masked to the same value. Stand-alone (without data virtualisation) masking support is also available for the aforementioned databases and additionally mainframe (iSeries, VSAM and z/OS) systems. Delphix also supports a tokenisation capability. Audit and compliance reporting is provided out of the box.

Delphix
1400A Seaport Blvd, Suite 200,
Redwood City, CA 94063

www.delphix.com

Strengths

- The nice thing about Delphix's approach is that you can mask once and then deploy multiple copies of that masked data to whoever needs it.
- Useful for moving dev/test to the cloud. Production data can remain on-premises with Delphix masking and then replicate only the masked data into the cloud.

Threats

- The data masking provided is good without being outstanding, though it is improving.
- Whilst support for structured databases is strong, Delphix doesn't currently support unstructured data.
- The real value of Delphix masking is combining it with virtualisation but the company may miss opportunities for masking when there is no requirement for Delphix's data virtualisation. Interestingly the platform does allow you to combine its data virtualisation with other 3rd party masking tools.

Summary

Unlike almost all other vendors in this Market Update we are not providing a spider diagram to illustrate the strengths and weaknesses of Delphix's offering. This is because, in our opinion, Delphix exists within a category of one, and it would be unfair to compare it with other products that are essentially different. While it is true that you can use Delphix for stand-alone data masking purposes the whole raison d'être for licensing Delphix is to make use of its virtualisation, which no other company offers.

HPE

HPE's SecureData product suite provides a comprehensive, end-to-end approach to enterprise data protection. In particular, it offers HPE Format-Preserving Encryption (FPE), a capability that focuses on filling gaps in security using a dynamic, data-centric approach to encryption and data masking. Although FPE is more oriented towards a dynamic data masking approach, it additionally offers static data masking if it is required. FPE uses the NIST standard AES FF1 encryption method that HPE helped pioneer (See NIST Special Publication SP 800-38G) and the AES-256 encryption algorithm. SecureData supports a wide range of both relational and NoSQL environments, over twenty core platforms including databases, data warehouse, cloud platforms, and mobile devices, with additional platforms supported on request. Alternately, you can integrate with it yourself using the API provided. The company provides tools for discovering sensitive data in both structured and unstructured data sources.

The European Union (EU) General Data Protection Regulation (GDPR) recommends pseudonymisation and encryption as two mechanisms that can be used to protect personally identifiable information (PII). The GDPR is careful not to prescribe specific forms of encryption or pseudonymisation. However, GDPR calls out two important encryption features: the ability to decrypt the data when necessary and the ability to continue to run business processes on the encrypted data. HPE FPE meets these guidelines.

Dynamic data masking through FPE masks the data at its source via encryption. The masked data is then provided to applications at run-time, decrypted appropriately for the authorisation provided by the calling application and the current user. Notably, this is not binary: it is possible to be partially authorised, in which case only the portion of the data that the user or application is authorised to see will be decrypted. This is useful if, for instance, a user group needs access to only the last four digits of a credit card number. FPE can provide those four digits to the appropriate users without exposing the entire number at any time. In the case of dynamic data masking, the masking is obviously reversible; if static data masking is used, the masking can be either reversible or non-reversible, as required. In either case, the data masking retains referential integrity between databases. Masking in FPE is also (as the name suggests) format-preserving: it retains the original format of the data. This minimises the changes required to existing programs in order to handle masked data. HPE FPE preserves referential integrity across disparate data sets so protected data can be consistently referenced and joined, provides multi-language support with Unicode Latin-1 and preserves context and relationships such as date range preservation.

FPE is configured through the SecureData Management Console, which provides central control over masking policy, including which authentication and authorisation methods are in use. Note that this can include using a third-party plugin. Additionally, the management console provides real-time event auditing, monitoring and reporting, including a granular view of events.

Hewlett Packard Enterprise

HPE
Cain Road, Amen Corner
Bracknell, Berks RG12 1HN, UK

www.hpe.com

Strengths

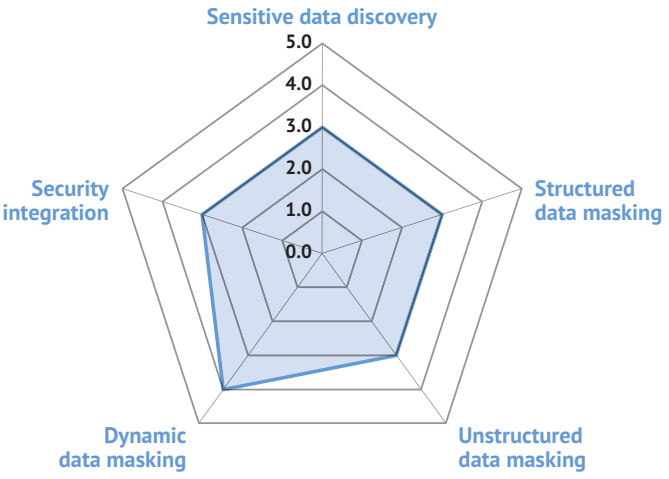
- HPE's dynamic approach to data masking protects data at rest, in motion and in use. In particular, the ability to have partial authorisation helps to minimise the possible exposure of sensitive data.
- The fact that HPE's masking is format-preserving means that little or no application code needs to be changed to accommodate the newly masked data.

Threats

- We see HPE's static data masking capability as being strictly average. This is not necessarily a bad thing, as HPE has a very clear focus on format preserving encryption rather than masking per se, but it is a point of distinction between the company and its competitors.

Summary

HPE is an outlier in this market and this report in that its focus is on FPE rather than data masking per se. Other companies often have the emphasis the other way around. Our view tends to concur with those others, but there is certainly a case to be made for FPE and, if that is your preference, then HPE is certainly a company you should be talking to.



IBM

IBM provides both static and dynamic masking alongside a variety of other products that fit into security, archival and test data management environments. For discovery purposes, you use Information Analyzer and this provides not just pattern matching and profiling but also facilities such as application and source code scanning and, for dynamic masking, semantic capabilities relevant to particular verticals. Products work with both structured and unstructured data (including data in Hadoop) and extensive encryption capabilities are also offered. One unique feature is the ability to apply dynamic data masking for web traffic, with software sitting between the browser and the application server. Significant auditing facilities and real-time monitoring of data activity are also provided as to who is accessing data, when, and for what purpose.

IBM offers Data Masking as a product within its Optim suite of tools. Historically, it was bundled with IBM Test Data Management, but this is no longer the case. This makes sense since TDM is essentially about improving efficiency and productivity, while data masking is about reducing risk and ensuring compliance. You can use Optim, which supports both mainframe and distributed environments, for dynamic data masking but this is more usually the domain of IBM Guardium.

Masking itself is comprehensive. Notable features include affinity masking (for example, maintaining case), consistent masking across multiple platforms, and semantic masking, though the latter is not easy to use. The company has also implemented in-database masking in a number of its environments (for instance, Netezza) and is actively extending this to others (for data preparation, for example). This is supported via user defined functions. IBM is also actively integrating masking into other environments (for example, there is a Data Masking Stage in DataStage).

Strengths

- Data masking capabilities are extensive. It is interesting to note that IBM is seeing more customers interested in masking unstructured as well as structured data, and it is pleasing that the company is developing capabilities to support these environments.
- IBM is implementing the ability to call masking functions from Java. This will be released during the course of 2017 and will support the Hadoop "menagerie".

IBM
Armonk, USA

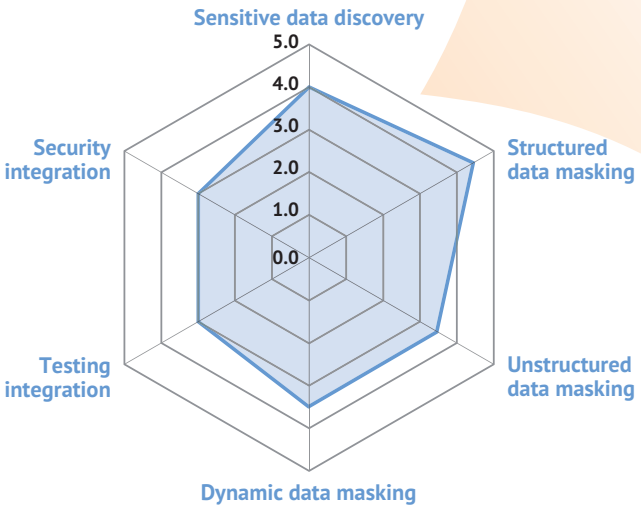
www.ibm.com

Threats

- IBM's offering is not as easy to use as we would like.
- We get the impression that the Rational/Optim interface is not as tight as it might be, and that the company is losing opportunities to leverage synergistic capabilities.

Summary

In terms of features and capabilities IBM is clearly a leading vendor. However, its products are not as easy to use as they might be and integration with non-core capabilities are not as close as we would like them to be.



Imperva Camouflage

Camouflage Software, which has recently been acquired by Imperva, was one of the early innovators in the data masking space. The company has, for example, two patents on data masking including one ensuring that the same data in multiple sources is masked consistently. Imperva, of course, has a broad portfolio of security products and in time we can expect the Camouflage products to be integrated tightly within the Imperva product suite.

Over the last couple of years, Camouflage has been investing both in the performance of its masking tools and in its discovery product. To reduce the incidence of false positive and negatives when discovering sensitive data, the company has been focusing on the classification of data. This is a pre-cursor to introducing machine learning that will incrementally improve accurate sensitive data discovery. In effect, the company has been building the platform that will support a machine-learning based approach.

While the company's core strength lies with its static masking solution, dynamic masking is available, and is supported via two mechanisms: either via a database proxy or using a message (XML) based methodology that intercepts messages, parses them and then applies relevant rules. The broader Imperva product suite also offers additional vehicles within which dynamic use cases can be achieved (for example, in conjunction with Database Activity Monitoring). Sensitive data discovery and classification, along with risk assessment and threat modelling, are helped by being able to view information graphically and this is becoming increasingly important at the executive level, as a result of industry requirements around data discovery and classification such as GDPR. Both structured and unstructured data (including Hadoop) are supported but the company does not currently support any other NoSQL data sources.

Strengths

- The risk assessment and threat modelling is not a usual feature associated with data masking and is a definite plus. This will no doubt dovetail well with Imperva's security offerings.
- Imperva is in the advantageous position of delivering a broad spectrum of data security technologies for both production and non-production environments, including offering alternative approaches for dynamic data masking.

IMPERVA®

Imperva Inc
3400 Bridge Parkway, Suite 200
Redwood Shores, CA 94065, USA

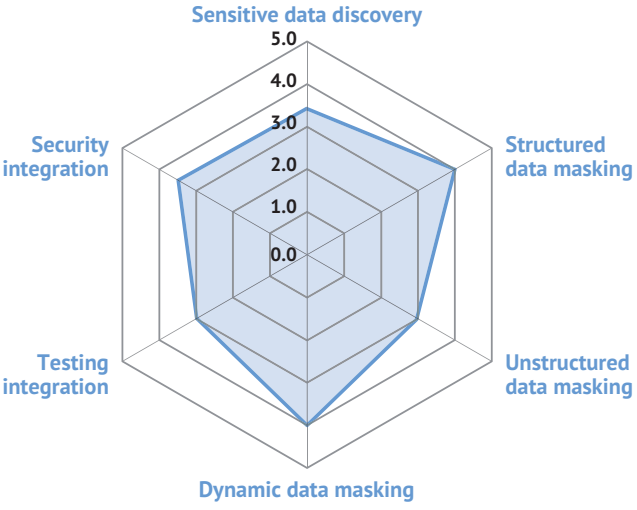
www.imperva.com

Threats

- While we are pleased to hear that the company is planning to offer machine learning for sensitive data discovery, there are several vendors that have already done this.
- The company does have a database sub-setting product for test data management but we have the impression that this is not a major focus. However, Imperva has a partnership with Actifio, which provides data virtualisation capabilities that support test data management capabilities, so we expect Imperva Camouflage to remain proactive in this area.
- The NoSQL support is limited and behind several of the company's rivals.

Summary

Camouflage was an innovative vendor in the data masking space while Imperva is well-known within the security arena. We have scored the product highly within this Market Update but, going forward, much will depend on how the integration between the two organisations works out.



Informatica

Informatica provides both static and dynamic data masking. For discovery of sensitive data and relationships in the data stores to be masked, Informatica provides the Discovery Option based on pattern matching, dictionaries, algorithmic, and other techniques – including machine learning – to reduce false positives. Relevant curation facilities are also provided.

In addition, Informatica Secure@Source provides enterprise-wide discovery of sensitive data, continuous multi-factor sensitive data risk monitoring, the ability to detect anomalous user activities on sensitive data, and facilities to automate the orchestration of data protection. Informatica Secure@Source extends the Data Masking offerings into a data-centric security solution to provide visibility and control of sensitive data. Secure@Source not only scans data stores across the enterprise to identify where sensitive data is located but also analyses where data proliferates within and across organisations based on data flows, such as Informatica PowerCenter, Informatica Cloud, B2B, Big Data Management, Cloudera Navigator, and other 3rd party data management solutions. APIs are available to leverage an organisation's existing data security solutions, such as discovery results from DLP, asset inventory from CMS systems, and the orchestration of 3rd party data encryption and tokenisation solutions. Full integration for Secure@Source to automate the orchestration of data protection including Informatica Data Masking and other 3rd party data protection solutions is planned for later during 2017, when Secure@Source and Persistent Data Masking will be packaged as a single product. The latter will still be available as a stand-alone product.

From a masking perspective, the company offers extensive options, including support for masking unstructured data and consistent masking across multiple data stores. Dynamic masking – usually used with production data – is available but requires a separate license. Compliance is also an important driver and Informatica ships with out of the box policies to support PCI, PII and PHI compliance. GDPR is planned.

The company also offers synthetic data generation as a part of its test data management solution and use of this means that data masking would not be required.

Informatica
2100 Seaport Blvd, Redwood City
California, USA, 94063

www.informatica.com

Strengths

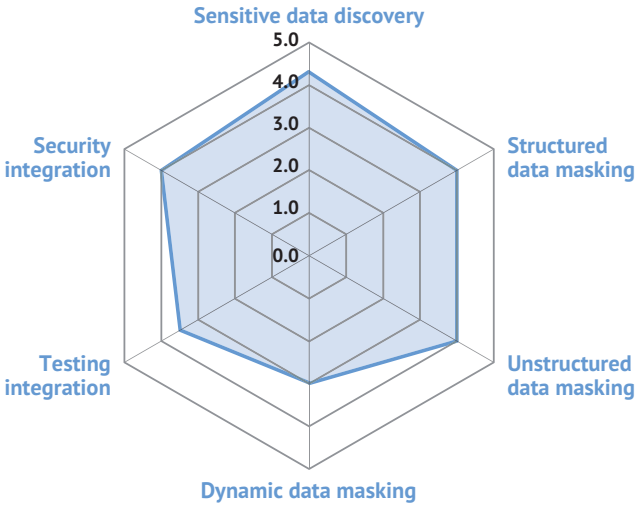
- Informatica's offering is impressive overall, with a broad capability and no obvious weaknesses.
- The integration (when it is complete) with Secure@Source is a major plus point.

Threats

- Despite its breadth of capabilities Informatica may fall victim to competitive pressures where other suppliers can extend discussions into areas where Informatica is less strong.

Summary

Informatica is a market leader in this space and, not surprisingly, scores very highly in this Market Update. Now that it is in private hands the company continues to innovate and we expect it to retain this leadership position going forward.



Mentis

Mentis focuses on "sensitive data lifecycle management" and it offers a number of relevant and complementary products for this purpose. These include iDiscover, which is used to discover sensitive data that needs to be masked, using either iScramble for static data masking or iMask for dynamic data masking. Complementary products include iRetire, for archived data, which can be used to support the "right to be forgotten"; and iSubset, which is used for test data management. iMonitor is used to monitor user activity for connection, statements or programs that access sensitive data.

In so far as Data Masking is concerned the main products that will be used are iDiscover, iScramble and iMask. Mentis goes further, in our opinion, than any other supplier in its facilities for discovering sensitive data. In particular, in addition to pattern recognition and similar profiling capabilities the software has the ability to introspect business rules written in SQL (for example, PL/SQL) that may identify sensitive data, thus enabling the discovery of sensitive data that might otherwise be missed. As far as masking is concerned, perhaps the greatest strength of Mentis' products are that that they provide consistent discovery (iDiscover) and pseudonymisation across both cloud and on-premises environments and across both structured and unstructured data sources. In addition, Mentis was the first company, as far as we know, to introduce conditional masking based on location, whereby you can apply different masking or access rules, depending on the location of the user (in or out of the office, in another country and so on). This applies to iMask. In addition to conventional databases Mentis also supports masking in Hadoop environments and it has integrated solutions that work with both the Oracle eBusiness Suite and PeopleSoft.

Format preserving encryption and tokenisation are both provided as options in addition to masking.

Strengths

- The discovery capabilities offered by Mentis are excellent, and market leading.
- The company offers a number of unique or near unique features such as location-based, conditional masking. Its iRetire product will be especially beneficial in regulated environments such as the EU's GDPR (general data protection regulation).
- The company's offering (which also includes iProtect, for intrusion prevention) is much broader in scope than most of its competitors.

Mentis
3 Columbus Circle, 15th Floor
New York, NY 10019

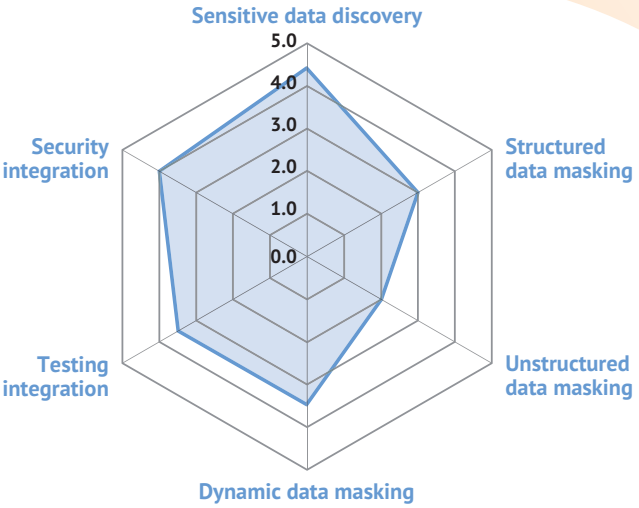
www.mentisoftware.com

Threats

- While Mentis has good support for structured database and other products (in both mainframe and distributed environments) it is less strong when it comes to unstructured data. While it supports a number of text based formats, its NoSQL support is on the company's roadmap rather than actually available, at this time.

Summary

While there are a lot of things to like about the Mentis product suite – not least its discovery capabilities – it is let down by its lack of support for NoSQL data sources. The company remains a leading contender for data masking solutions if this is not an issue.



Net2000

Data Masker is a static data masking product offered by Net2000 product that supports SQL Server and Oracle. Data Masker is fast (processes are run in parallel) and can mask millions of rows an hour. It has a dual focus on compliance (for instance, with HIPAA) and maintaining the credibility of masked data. The latter is done by ensuring that correlated values (for instance, age and date of birth) remain consistent after masking. This is done via substitution rules using a correlated data set that contains potential values to substitute. For instance, your data set could contain random surnames, or it could contain random US zip codes with accompanying state, county and town names appropriate to that zip code. These data sets can be user defined, either manually by the client or, as has happened often in the past, by Net2000 on request. In addition, this can be done between databases or even database instances.

Notably, masking in Data Masker always retains relational integrity and includes the capability to mask primary or foreign keys without a join operation. Data Masker also provides a column finder that allows you to search your database based on column name. Lastly, it automatically generates reports whenever a masking rule is run, making the masking process fully auditable.

Strengths

- Data Masker is fast, able to process millions of rows an hour. Its support for parallel processing is particularly impressive.
- Data Masker is lightweight and has an intuitive interface. In particular, it has no extraneous functionality that would make it difficult or overcomplicated to work with.

Threats

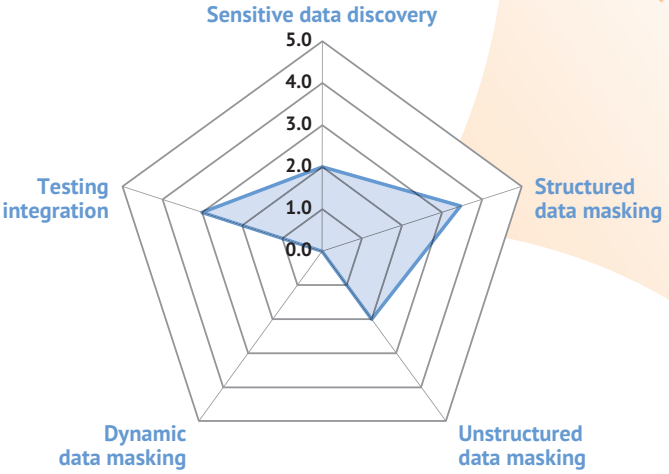
- The column finder is a useful feature, but only searching by column name is limited. We would like to see something more sophisticated.

Net2000
Net 2000 Ltd. , Llangunllo, Knighton
Powys LD7 1SP, UK

www.net2000ltd.com

Summary

Data Masker is, arguably, aimed at the mid-market rather than large enterprises. It is fast and easy to use but lacks some of the sophistication of more expensive products. On the other hand, do you really need to pay for all those bells and whistles? If not, then Data Masker may be an appropriate choice.



Privacy Analytics

Privacy Analytics was a Canadian company associated with (rather than in) the data masking space. It was acquired by IMS Health during the course of 2016, which merged with Quintiles in the same year, so that Privacy Analytics is now a part of QuintilesIMS.

Privacy Analytics provides risk-based de-identification (or pseudonymisation) particularly (but not only) for the healthcare and life sciences industries. It addresses the issue – which is particular to this industry – that you not only want to hide personally identifiable information but you also want and need medical researchers to be able to analyse that data. The latter may well require some data that might otherwise be regarded as sensitive such as patient's age, medical condition and so on. There is thus a balancing act between security on the one hand and research on the other, and Privacy Analytics aims to provide this. The product uses risk analysis to review the sensitivity of the data, then applies techniques to minimise risk of re-identification while maintaining the granularity in the data. Once de-identified, the risks are reviewed again so that an organisation can balance privacy with utility, so that you can ensure relative security while still supporting the work of researchers.

In practice, Privacy Analytics provides two de-identification tools that incorporate masking techniques, and a Risk Monitor where the last of these provides a risk score (based on HIPAA statistical methods) that quantifies the likelihood of sensitive data being de-identified based on both direct and quasi identifiers. As far as we are aware, this is unique in the marketplace.

The company's other products are also, uniquely, risk-based. Apart from this the base product is otherwise unexceptional in terms of how it masks data, but the recently introduced Privacy Analytics Eclipse (which includes Risk Monitor) is a Spark-based, in-memory offering deigned to de-identify sensitive data within stream processing environments. Again, as far we know this is unique in the industry and it is the primary reason for the company attracting attention outside its traditional market in, for example, IoT environments and financial services (hedge funds).

PRIVACY ANALYTICS
NOW PART OF IMS HEALTH

Privacy Analytics Inc
251 Laurier Avenue W, Suite 200
Ottawa, Ontario, Canada K1P 5J6

www.privacy-analytics.com

Strengths

- The Risk Monitor is unique and is worth considering even in conjunction with third party data masking products.
- The Eclipse product is compelling: we know of no other vendor with this capability.

Threats

- The company does not offer a specific product for discovering sensitive data. Instead, the company relies on its team of data analysts to assess the risk of re-identification. They look at the data schema for non-identifying information, quasi identifiers and direct identifiers. The company has developed capabilities dealing with financial and geospatial data as well as health related information. Privacy Analytics is not well known outside the Healthcare market and this may make life difficult in new markets.

Summary

Unlike almost all other vendors in this Market Update we are not providing a spider diagram to illustrate the strengths and weaknesses of the IMS Privacy Analytics product. This is because, in our opinion, it exists within a category of one, and it would be unfair to compare it with other products that are essentially different.

Protegrity

Protegrity is a provider of data security solutions across both traditional and big data environments, working both on-premises and in the cloud. It is a long-time partner of Teradata in providing data-centric security.

Protegrity takes an interesting approach to data masking in that it provides static data masking (which they call tokenisation), presentation masking on display and format preserving encryption. While provision of both of the first two is not unusual, the emphasis is more on the tokenisation than the masking, which is. The big difference, of course, is that tokenisation is reversible while data masking (typically) is not. The basic premise adopted by Protegrity is that pseudonymisation is policy-based. That is, you define who, based on their roles, can see what part of the data (for example, the last four digits of a credit card number). If you do not define a policy, then irreversible masking routines are used but otherwise you will be deploying tokenisation. You can define policies that combine different approaches so that, for example, a policy might include both static and presentation masking.

Protegrity does not yet offer automated discovery capabilities for finding sensitive data, instead identifying it as part of the implementation and engagement process. However, the company is working on a suitable offering. This will be based upon classifiers, each of which may be weighted, and will provide confidence levels to help to reduce false positives and negatives. We understand that machine learning capabilities will be built-in so that these falsities can be further reduced over time.

Strengths

- Protegrity provides a comprehensive range of options for securing sensitive data.
- The company provides significant support across both database and deployment platforms.

Threats

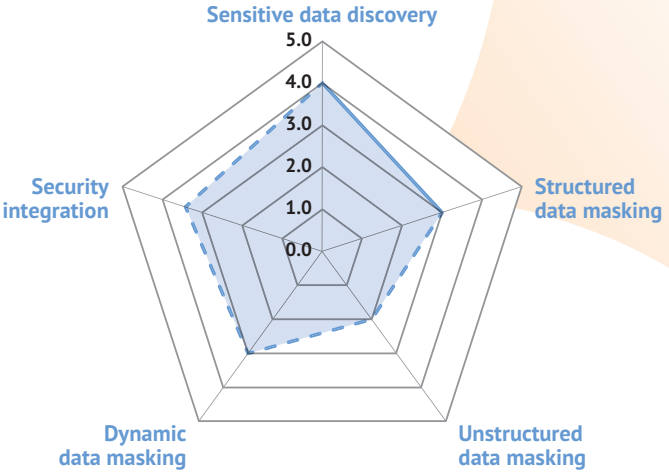
- The current lack of automated discovery capabilities is a drawback though the planned product capabilities look encouraging (and beyond what many of its competitors currently offer).
- The company does not provide test data management, which will leave it at a disadvantage for those users focused on non-production DevOps environments.

Protegrity
5 High Ridge Park
Stamford, CT 06905 USA

www.protegrity.com

Summary

Protegrity has a solid offering for data masking without being outstanding. The big hole in the product set is the lack of discovery capabilities and without that, frankly, we would not choose this product. In the spider diagram we have put in a putative (dotted line) score based on what we know about the forthcoming release of this capability.



Solix

Solix is a provider of information lifecycle management solutions for infrastructure optimisation, data security, and analytics. It provides enterprise archiving and application retirement solutions for both structured and unstructured data. Solix also provides data masking and test data management as a part of its solution. From an archiving standpoint Solix's main differentiator is that it offers multiple tiers of archiving – partitioning, database archiving, and archival on Apache Hadoop. From a data masking viewpoint, Solix provides static rather than dynamic data masking but it includes on-the-fly capabilities. Solix also supports encryption and security analytics where this may be a better solution than masking: for example, where protecting intellectual property such as network or security logs. Discovery capabilities include automated recognition of PII, PCI and PHI data elements.

Solix has two solution suites: the Big Data Suite and the Enterprise Data Management Suite (EDMS) with the latter incorporating the company's data masking offering. Solix offers the normal sorts of masking algorithms you would expect. These are complemented by a discovery tool that both looks at metadata (column names) and actual data (looking for patterns) to discover sensitive data. To reduce false positives it uses sampling to help users decide what is and is not sensitive. A notable differentiator is that Solix offers pre-packaged masking capabilities for Oracle and PeopleSoft application environments. While the Solix products run generically across most leading relational databases using their java based masking algorithms, the company has also implemented its masking algorithms natively in Oracle PL/SQL.

Solix
4701 Patrick Henry Dr., Bldg 20
Santa Clara, CA 95054

www.solix.com

Strengths

- The pre-packaged capabilities and specific support for Oracle environments will be beneficial when it is Oracle-based data that needs to be masked.

Threats

- Solix does not currently support data masking for NoSQL environments. However, the fact that the company's Big Data Suite includes capabilities for data lake management mean that this is likely to appear sooner rather than later.

Summary

For Oracle-based environments and masking in conjunction with archival Solix is clearly a go-to vendor, but otherwise we regard it as solid without being outstanding. We would like to see NoSQL support as soon as possible.