

Data Masking with Capgemini using Informatica



With Capgemini's metadata framework, your data masking implementation is scalable and repeatable across the enterprise – making it completely safe and highly cost-effective.



Data masking is now an important requirement across all industries. It is acknowledged as an invaluable method for transforming production data to provide realistic data for non-production purposes such as development and testing, without any risk to security or privacy. It also enables support to be carried out safely in the production environment with users seeing only the data they need for their role – sensitive elements are invisible.

Data protection legislation and tight regulation of industries such as financial services and healthcare have increased the pressure for effective data masking – particularly for organizations that want to carry out development or testing offshore. Yet many still tackle masking in an ad hoc manner that is not always reliable and involves substantial manual effort.

Capgemini offers a unique metadata framework that remedies this situation by enabling repeatable, consistent data masking at the enterprise level. Our accelerators can jumpstart new engagements, reducing cost and effort by around 40%. We partner with Informatica to implement this framework using its leading data integration tools like PowerCenter, PowerExchange and ILM. We can deliver our solution in a number of ways, from a one-off “quick fix” or implementation for a single business unit to an enterprise-wide implementation of data masking as a managed service.

Organizations need a more coherent approach to data masking

Data protection laws – and relevant regulations in industries like financial services – are becoming more stringent, making effective data masking imperative. For organizations that want to take advantage of cost-effective delivery models involving offshore or near-shore resources, data masking is a vital enabler.

Capgemini offers the following key data masking techniques:

- **Persistent data masking**, which permanently removes sensitive elements before production data is moved to a non-production environment.
- **Dynamic data masking**, which enables data residing in a production database to be masked in different ways based on entitlements, so that production sees the full record while support (for example) sees a masked version.

How our approach works

Capgemini's framework tackles data masking at the enterprise level, as a shared service. The framework provides a common infrastructure for development, testing and support, enabling scalability and repeatability across applications.

We achieve reuse and scalability via the use of process-oriented metadata that defines the way masking is to be carried out for each item of data in terms of subsetting, encryption, manipulation and so on. Used in conjunction with Extract Transform Load (ETL) tools and operational scripts, the metadata results in a completely standard masking process whenever anyone in the organization needs to mask a given piece of data for a given purpose.

Tools we typically use are Informatica PowerCenter with PowerExchange for Extract Transform Load (ETL) and ILM for masking: products that embody best practice for handling large datasets across multiple technologies. We have built accelerators for every phase of the masking process, to enable and automate the entire lifecycle. These include a metadata profiler, code generator, data comparer and automated refresh report. They are all tightly integrated with the metadata solution to enable seamless execution and save cost and effort.

Benefits

It is our use of metadata, together with our adherence to the six principles, which makes our data masking approach unique in terms of reusability and scalability.

With it, you can obtain quality test data to facilitate seamless and thorough testing across all applications while complying with information security guidelines. Testing can safely be outsourced since there is no need for access to production data – and it is seamless because everyone uses a single version of the test data across all applications. The framework also enables secure access to production databases for operational and production management teams.

This approach, with its proven tools, accelerators and processes, improves efficiency and cost-effectiveness while minimizing the risk of reputational or financial losses from data breaches.

Six guiding principles for our approach

- Data masking rules must be managed and applied by a centralized rules engine. They must be maintained at a business unit level and governed at an enterprise level.
- Masking must not be reversible: it must never be possible to use masked data to retrieve the original sensitive data.
- Masked data must be representative of the source data and must resemble it, otherwise there is no advantage over randomly generated data.
- Referential integrity must be maintained, for the same reason.
- Non-sensitive data should be masked only if it can be used to recreate sensitive data, or to link to or identify it.
- Masking must be a repeatable, automated process.

Success stories

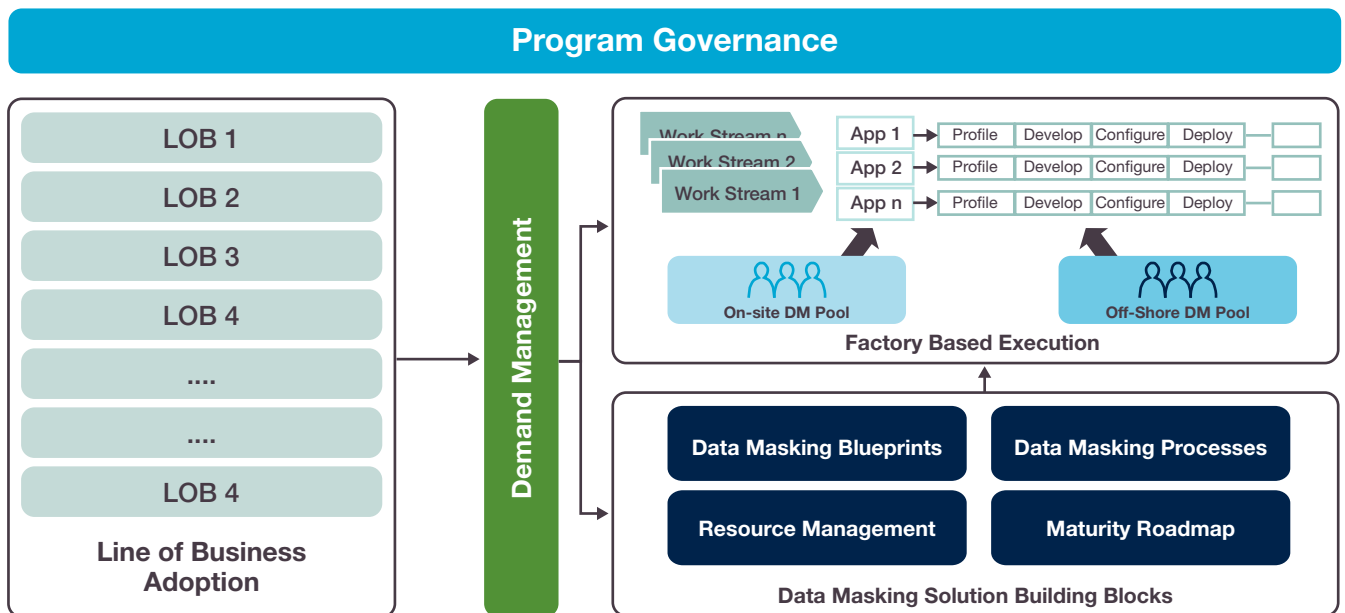
For a global investment bank, we used an existing Informatica platform in conjunction with our framework to implement enterprise-wide data masking as a managed service. We used accelerators to automate data masking processes, resulting in 30% cost and effort efficiencies. We completed the program 2 weeks ahead of time that saved up to 5% of the total program cost.

For a leading international bank, we set up an Informatica-based Data Masking Center of Excellence to provide a shared service to all business units, using consistent masked data. This initiative achieved significant cost savings on existing masking requirements by reducing costs up to 20% to mask applications for a new business unit.

Deployment options

We adopt a phased approach to deployment so that our framework is progressively rolled out across all applications and business units with a large number of applications. A factory model is used so that several applications can be tackled in parallel. Use of our accelerators and best practices from other implementations makes deployment straightforward in either case.

Figure 1: Our operating model



We offer a range of implementation options. We can pilot just a few applications first to confirm the approach, or can quickly tackle a particular application or activity that has issues. We can also provide data masking as a managed service. Please contact us to discuss what would suit your organization best.



About Capgemini

Now with 180,000 people in over 40 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2014 global revenues of EUR 10.573 billion. Together with its clients, Capgemini creates and delivers business, technology and digital solutions that fit their needs, enabling them to achieve innovation and competitiveness. A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.

For further information visit

www.capgemini.com/insights-data

or contact us at

insights@capgemini.com

About Informatica

Informatica Corporation (Nasdaq:INFA) is the world's number one independent provider of data integration software. Organizations around the world rely on Informatica to realize their information potential and drive top business imperatives. Informatica Vibe, the industry's first and only embeddable virtual data machine (VDM), powers the unique "Map Once. Deploy Anywhere." capabilities of the Informatica Platform. Worldwide, over 5,000 enterprises depend on Informatica to fully leverage their information assets from devices to mobile to social to big data residing on premise, in the Cloud and across social networks.

For more information, visit

www.informatica.com