COMFORTE SOLUTION BRIEF

COMFORTE - Discovery and Classification

Sustainable data discovery for privacy, security, and governance

With the growing complexity of modern networks and data storage environments, ensuring constant visibility of sensitive data across your entire organization is nearly impossible. Systems are more interconnected than they ever have been, and recent breaches show that implementing data security effectively requires a strong knowledge of your data landscape.

Manual snapshots and analytics won't get you there

Data usage is heavily dynamic and constantly evolving. At any given time, you really have no way of knowing all the places to look for sensitive data, especially if you depend on a reactive manual process and have an extensive data ecosystem.

Take complete control of sensitive data

Comforte's **Data Discovery and Classification** solution enables organizations to detect and analyze all usage of data and its lineage without relying upon organizational knowledge of the existence or location of that data. Better yet, the process is completely automated! This automation makes it much easier to get a clear picture of how your data is being stored, processed, and shared in real-time.

> Protecting data requires knowing where data is, and knowing what it is. You can't protect what you don't know exists.

Diverse regulatory requirements specify minimum standards of data protection and require complete compliance from organizations. Regulations make data protection an absolute necessity for your business. Keep in mind: the enterprise collecting, processing, and storing that data bears the brunt of responsibility for data protection in the eyes of regulators!

To mitigate risk most effectively, businesses need to **discover unknown sensitive data across the entire data landscape**.

- Understand near **real-time sensitive data lineage** and business context of any sensitive data element in your environment
- **Gain complete visibility of the usage of every data subject's information**
- Understand when sensitive production data is found outside of production environments
- Automatically generate a full master catalog of sensitive data in near-real-time
- Implement measurements, monitoring, and enforcement tools to govern the usage of sensitive data
- Understand how records from across disparate sources come together in unique data subject records.

96% of enterprises in the US see a risk in *unknown repositories and data flow*

Solution in a nutshell

Comforte's **Discovery and Classification** solution is a unique and proprietary passive network packet capture process to identify sensitive data (such as highly-regulated PII) flowing through the organization. This flow visibility enables our solution to identify repositories (databases, applications, file systems, and log files) where sensitive data resides. The solution then does a comprehensive scan of those repositories to get full visibility into the depth and breadth of the data environment. Finally, it analyzes and consolidates the data identified as a result of those scans in a structure that allows the user to see data lineage, respond to subject access requests, identify production data in non-production locations, and many other privacy, security, and data governance tasks.

> "Discover sensitive data and its usage, even sensitive data you didn't know existed in your environment.

Did you discover everything - and everything that is important?

Your data ecosystem is dynamic. Information constantly changes within it. Therefore, you need to discover these changes continuously and assess the amount of risk within it.

- Continously monitor your network and identify network elements
- Start with zero knowledge of what's in your data
- Discover both unknown and known data sources
- Leverage a unique and proprietary passive network packet capture process

> Update your inventory continuously with minimal manual resources.

Running use-case example

For example, a hospitality chain may have 25 million customers across 2,500 data stores. No organization can manage that amount of data manually, no matter how many people are thrown at the matter. And because new products and applications are introduced almost every week by their efficient DevOps team, these data stores are highly dynamic and ever-changing. Scanning these data stores once—or just once in a while—won't reveal all the potential risk. This data ecosystem must be continuously scanned in order to keep up-to-date on what's really in there.

Only scan what's important

But how do you decide what's important?

Our solution identifies sensitive data elements and repositories that manual processes or limited discovery solutions simply can't find. Why overlook unknown sensitive data that exposes you to risk?

> Address one of today's greatest cloud security challenges.

Comforte's **Discovery and Classification** solution connects to a multitude of data sources, central file systems, industry-standard databases, NoSQL and SaaS solutions, and even Amazon S3 buckets.

Running use-case example

Not all information is equally interesting or sensitive. In our running example, an enormous amount of data may be present across 2,500 data stores that is not sensitive at all. And yet, you need to know where all the PII-based information exists, like small pearls on a vast beach of sand.
You need a solution that can identify what PII is stored and processed as customer data (for example, perhaps there are 50-60 PII fields, including credit card details, loyalty data, historical booking data, preferences, passport, mobile device ID, travel insurance policy, flight details, and banking or financial information). This is where the real risk lies for the organization.

Understand sensitive data no matter what type

All data isn't alike.

From highly structured databases to free-form documents like PDFs and TXT files, your business data is probably both structured and unstructured. You must discover all data, regardless of its type or format. Regulators won't draw a distinction just because a certain type of sensitive data is hard to find or work with.

- Find data on-premise or in the cloud, both structured and unstructured, and whether it is in motion or at rest
- Analyze databases, file systems, and other repository types
- Leverage machine learning capabilities that "read and understand" your data

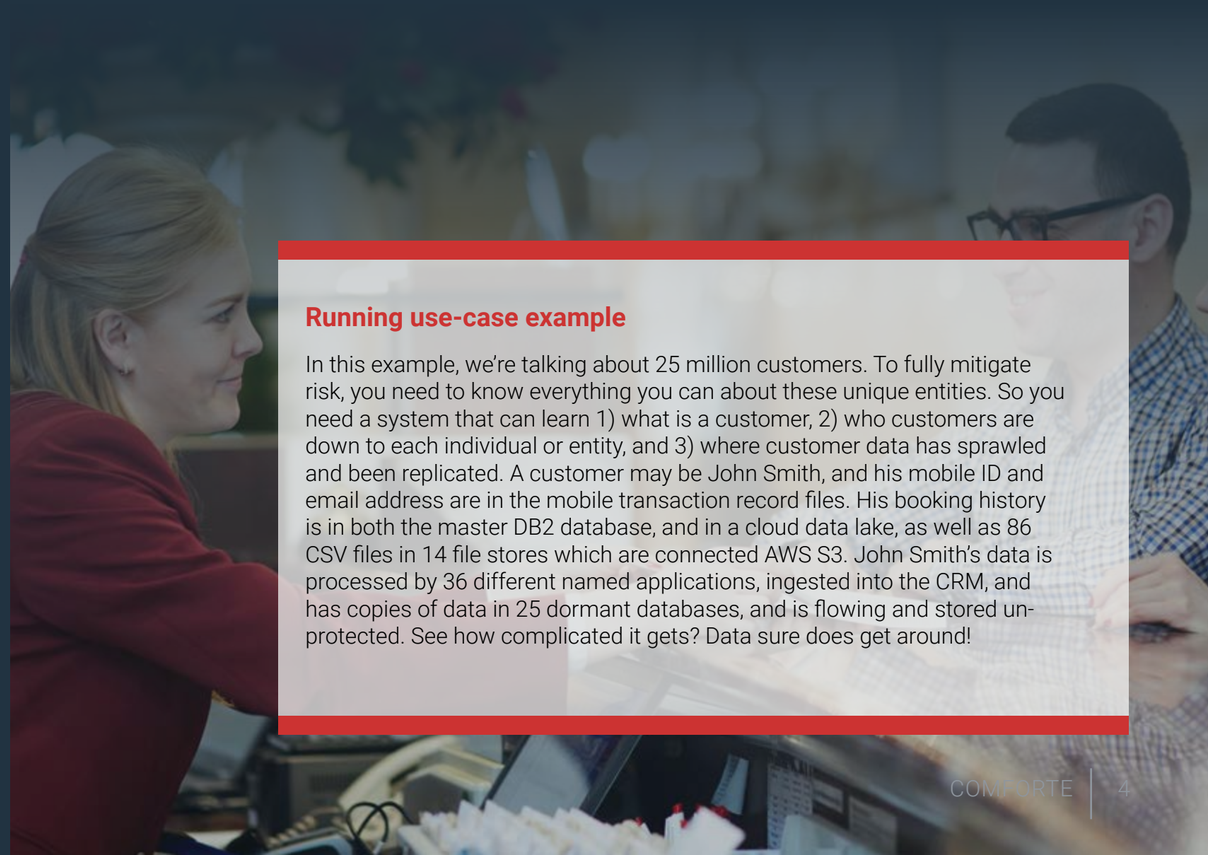
Running use-case example

In our running example, much of the information within the 2,500 repositories is structured within known databases. Information such as the 50-60 PII fields (credit card details, loyalty data, booking data) will definitely derive from known databases. Yet, some information related to customers might exist in unstructured data, such as correspondences from customers, data analytics information in a variety of report formats (Word documents, PowerPoint presentations), and other non-database-sourced information. You can't just assume that all PII is stored within known structured databases—that assumption leads to further risk and exposure.

Organize the data you find

Correlate your data with people or entities that use that data. Determine what your data really is and then create a unified view:

- Combine sensitive data from disparate sources into one data subject record
- Compare the data subject record with known business usage to confirm known, managed PII



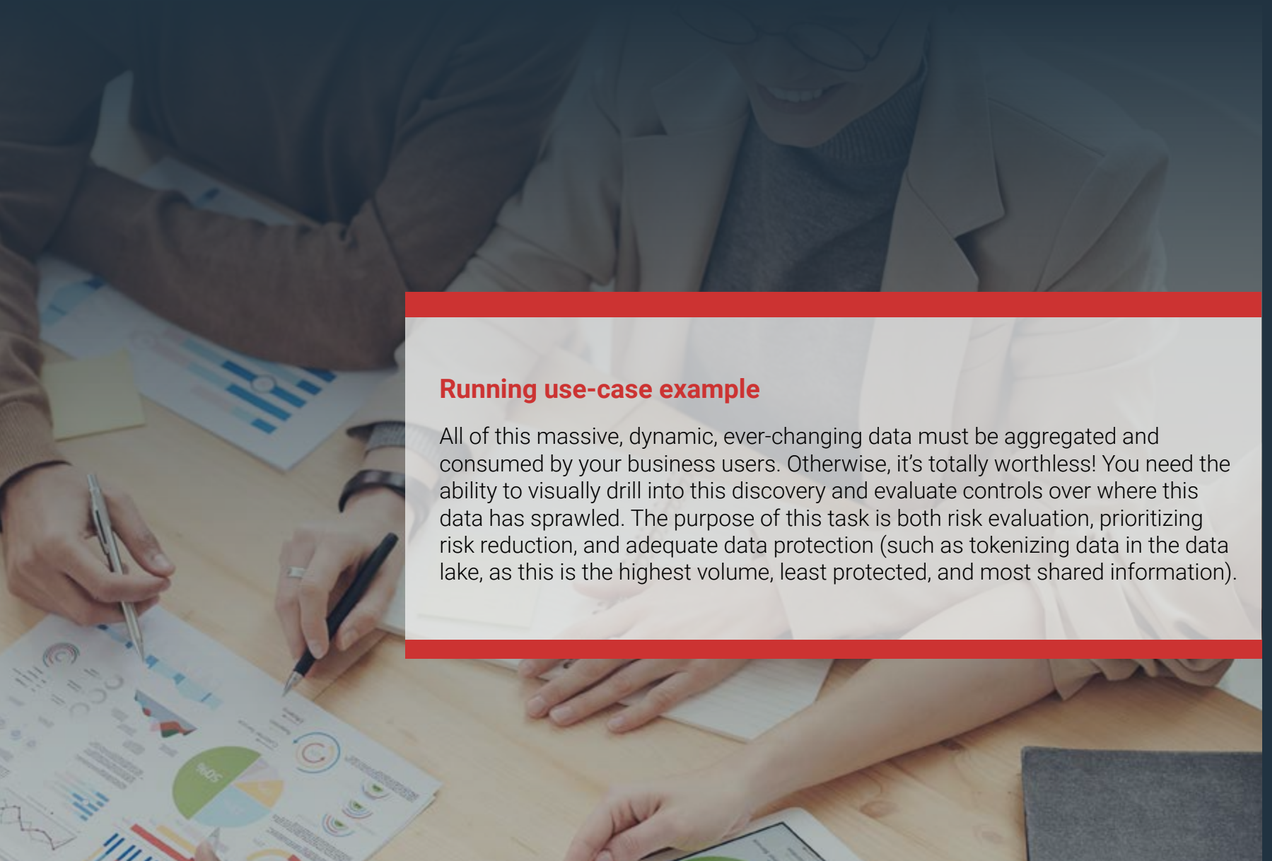
Running use-case example

In this example, we're talking about 25 million customers. To fully mitigate risk, you need to know everything you can about these unique entities. So you need a system that can learn 1) what is a customer, 2) who customers are down to each individual or entity, and 3) where customer data has sprawled and been replicated. A customer may be John Smith, and his mobile ID and email address are in the mobile transaction record files. His booking history is in both the master DB2 database, and in a cloud data lake, as well as 86 CSV files in 14 file stores which are connected AWS S3. John Smith's data is processed by 36 different named applications, ingested into the CRM, and has copies of data in 25 dormant databases, and is flowing and stored unprotected. See how complicated it gets? Data sure does get around!

Understand the journey of your data

Link all the pieces into an informational picture of your data – its source, its usage and its transformation over time

- Understand *data lineage* and *data flow*
- Update data subject records and references to discovered PII
- Determine data subject relationship(s) to your company
- Identify sensitive data shared with 3rd parties



Running use-case example

All of this massive, dynamic, ever-changing data must be aggregated and consumed by your business users. Otherwise, it's totally worthless! You need the ability to visually drill into this discovery and evaluate controls over where this data has sprawled. The purpose of this task is both risk evaluation, prioritizing risk reduction, and adequate data protection (such as tokenizing data in the data lake, as this is the highest volume, least protected, and most shared information).

Use discovered risks to drive the most appropriate controls

Create a complete virtual catalog of sensitive data, including to whom it belongs and where it is processed, stored, and used.

These data elements can be associated with tags (such as *compliance*, *customer*, *line of business*, and *process*). This allows the data to be queried, analyzed, ranked, sorted, and processed for risk assessments against compliance gaps, data volumes, and data types.

In addition, our solution detects processes where sensitive data is flowing that may represent risks of leakage:

- Reporting data in previously unknown dormant databases, files, stores or shares,
- Processes sharing data with a cloud,
- Unexpected file store, such as extracts from a database going into SharePoint in Office 365
- Processes that are not managed by the data owners (like copies in data lakes)

The visibility and insights enable you to decide which data to protect or which to delete, all in fulfillment of regulations and mandates. With comforte's platform, your business is able to protect sensitive data from risk of accidental loss, exposure, or accidental leakage from misconfigurations, all with a multitude of protection methods so that you can match the right method for the discovered data.

> Hardware or not hardware – that's the real question

Comforte's scale-out, distributed architecture allows you to deploy as many discovery instances as needed to aggregate the data. Analytic Cores can be installed as a physical or virtual instance (which can be deployed in AWS within your VPC).

Visibility into threats or risks to the data

Obtain a clear picture of how your data is being stored, processed, and shared in near real-time.

Automatically discover and analyze all usage of data and its lineage without relying upon your organization's pre-existing knowledge of the presence or location of data.

Reduction of risk

With this discovery knowledge, you can create effective protection policies and implement security controls that fit to your business use cases. You can identify sensitive data, protect it properly, then monitor ongoing changes in your data ecosystem.

Compliance with privacy regulations

Comforte's **Discovery and Classification** solution enforces better privacy, security, and governance measures by creating a *Master Data Catalog* inventory. Linking all the pieces into a comprehensive informational picture of your data allows you to identify compliance risk and manage data subject access requests—including the right to erasure, update, or share data changes.

Manage sensitive data across your entire organization.

Comforte's **Data Discovery and Classification** solution offers a complete platform for data lineage detection for all PII across a scaled enterprise, with rapid mapping of data storage, processing, and use. Our solution is ideally suited for detecting data, purpose, and use, and it is aligned with modern privacy compliance requirements for data-intensive industries.

The desired outcome for our customers is automated data risk discovery as a continuous process, versus a one-off process or personnel-heavy manual process plagued by false positives. Our solution provides rapid and effortless discovery of unknown risks that could cause serious harm to your business, brand, and reputation.

In combination with comforte's data protection capabilities, our solution converts discovered data risks into prevented data breaches and true regulatory compliance.

Recap

Data discovery isn't something to take for granted. Manual discovery of potentially sensitive information simply doesn't work for sophisticated and dynamic data ecosystems, especially in highly regulated industries such as financial services, healthcare, and insurance. Why risk the risk?

A comprehensive solution

Implementing data-centric security requires a platform that allows you to discover, protect, and manage sensitive data. It must enable you to integrate these capabilities quickly and easily into your enterprise applications and existing cyber security infrastructure.

Comforte's data security platform enables a comprehensive end-to-end data security strategy. Our customers are protecting hundreds of millions of payment transactions, sensitive healthcare records, insurance records, and more, all reliably running in business-critical environments.

DISCOVER & CLASSIFY
Instrument Sensitive Data Discovery as a Continuous Process

INVENTORY
Identify Data, Ownership, Lineage & Flows

POLICY
Enable Data Security as a Service from the CI/CD

PROTECT
Instrument Data Security in applications

DEPLOY
Reduce Implementation cost and effort. No code to low-code

Comforte Data Security Platform

Your next steps

We're able to show you our discovery and classification capabilities in action. Contact us to get a demo or to organize a 30min technical discussion.

www.comforte.com

comforte

www.comforte.com