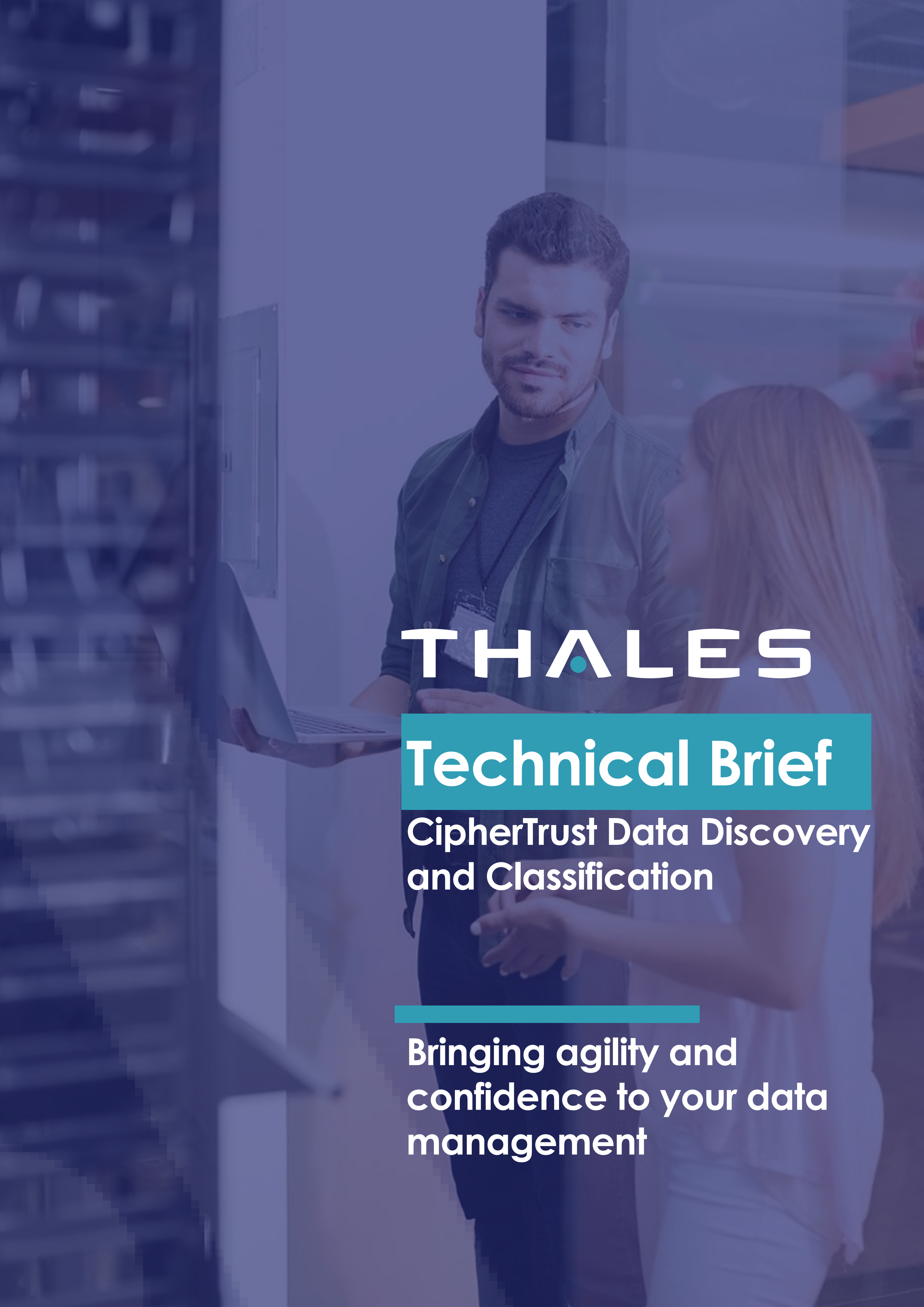


THALES

# Technical Brief

## CipherTrust Data Discovery and Classification

Bringing agility and confidence to your data management

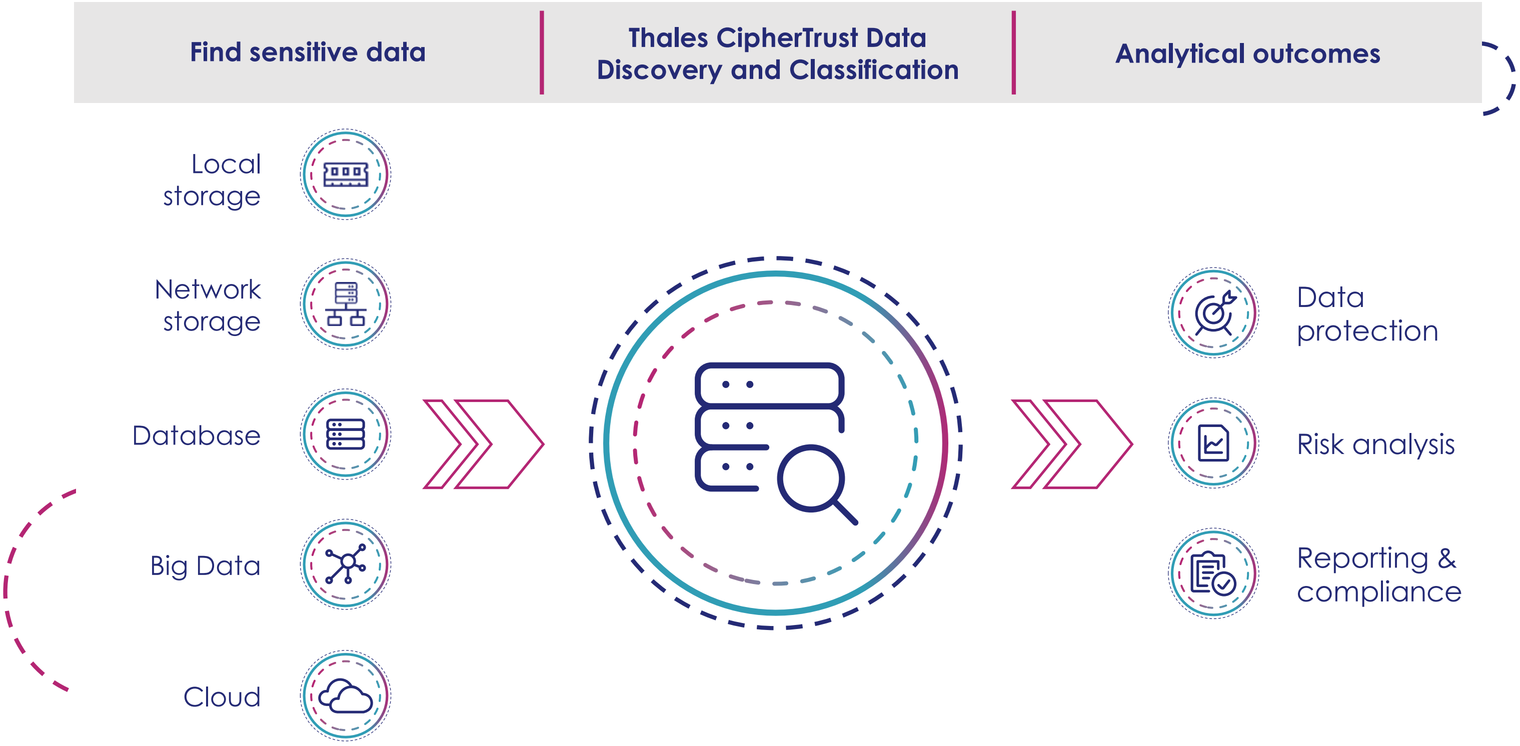
# Executive summary

The rapid, often exponential, growth of data year-on-year in organizations like yours makes effective data management an extremely challenging proposition. An increasing switch to remote working raises the bar even higher with many data storage areas or volumes coming into play which are outside your IT team's direct control.

As part of the broader group within your organization responsible for data privacy and security compliance, ensuring no vulnerable areas are overlooked when implementing your data protection strategy is of paramount importance. After all, a data breach will inevitably cause severe business disruption in addition to large fines incurred for non-compliance with the seemingly endless stream of new or enhanced data privacy laws and regulations. Failing to prepare properly is definitely not a viable option for you.

A significant weakness often seen in typical data management implementations is the lack of visibility into the precise types of data being held across various local servers, network drives and increasingly cloud storage locations. Many organizations have been breached already, some have encountered near misses and others are migrating large workloads to the cloud without fully understanding the fundamental nature of the data itself and the exposure risks involved – situations for you to avoid.

Historically staff knowledge was sufficient and simple 'off-the-shelf' encryption methods from database vendors would satisfy needs – no longer with the vast data footprint that is widely dispersed and growing by the second. You inevitably need assistance to take control and keep your organization safe.



**Find sensitive data**

**Thales CipherTrust Data Discovery and Classification**

**Analytical outcomes**

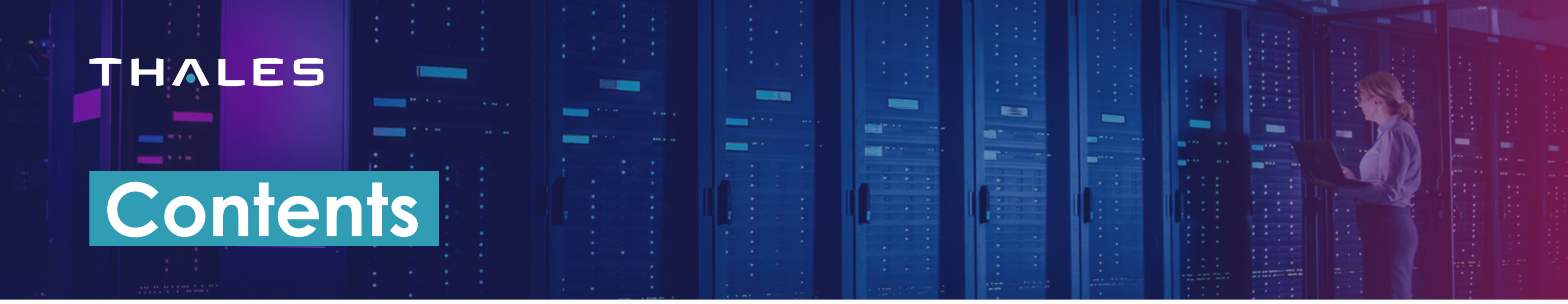
## Thales CipherTrust Data Discovery and Classification is…

A tool that is significantly more effective and efficient than manual methods for data discovery. It helps classify all your data, while supporting every mainstream operating system, data storage type, structured and unstructured data you are likely to possess. Existing processes and tools are supplemented, rather than replaced, providing high levels of automation in your search for sensitive data, wherever it resides. Ultimately, it can help you become more agile and support better data management decisions:

- It finds sensitive data that you may not realize even exists in your organization, helping you eliminate threats to your business continuity and reduce your data footprint

- It enables you quickly to assess your risks of non-compliance and determine the appropriate protection actions to take

- It provides confidence that you are protecting the right assets, including proprietary data or intellectual property, across all your data storage locations including the cloud



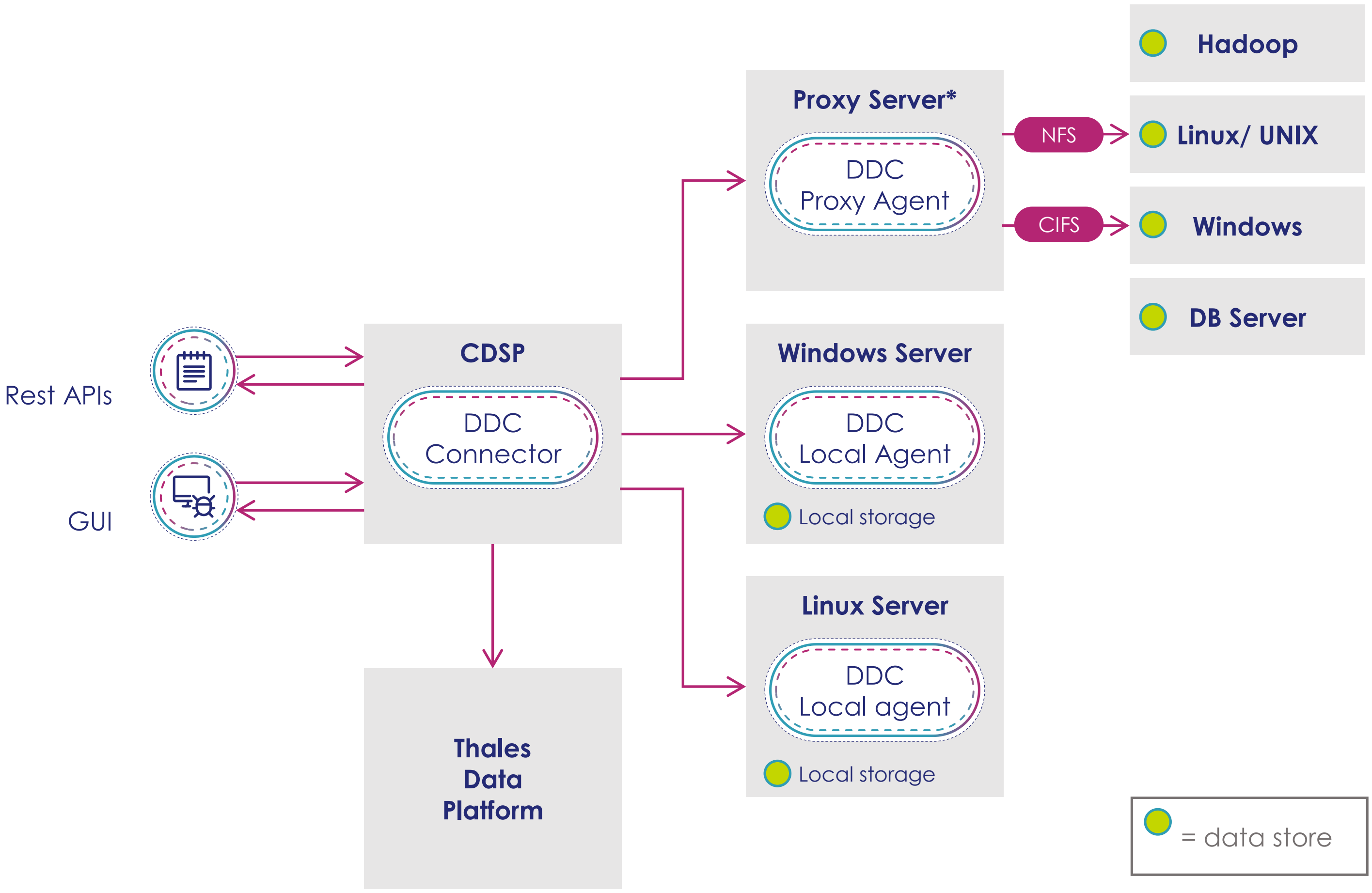
THALES

# Contents

THALES

# The role of CipherTrust Data Discovery and Classification (DDC)

DDC offers a more streamlined, accurate and unattended (or automated) approach in comparison to sensitive data searches using manual methods. If you are using a data discovery tool from another vendor you should still take a closer look at DDC - it likely covers additional data storage types, extending the range of structured and unstructured data that can be analyzed while offering higher levels of accuracy by finding sensitive data other tools often miss.



* A Windows proxy is needed to connect to databases.

Sensitive items found by DDC are not just restricted to data privacy laws and regulations. Fine tuning using custom infotypes (which you define) facilitates visibility on where data sensitive to your organization is present. By enabling critical assets (such as internal financial data, trade secrets, intellectual property and confidential business plans for example) to be located, their current levels of protection and associated risks of unintended exposure can be better understood. Considerable flexibility comes from running multiple scans simultaneously, each looking for different types of information (if desired) to create a faster and more efficient way to analyze your complete data footprint across a diverse range of storage locations. It is the ability to create multiple classification profiles (customized where needed) that underpins such performance and efficiency benefits that never could be achieved in a comparable timeframe using manual methods or less capable tools.

Discovering the data is only part of the task in hand – **classification is equally important**, providing deeper understanding of the types of data and their footprint in your organization. DDC does not move or modify your data, merely summarizing what it uncovers during scans in a non-intrusive manner.

You can **see exactly how your data is split** between the personal, financial, medical and national ID categories – you may be surprised that some of this even exists. At least now you will have insight into precisely where it resides to manage the business risk appropriately.

The insights available after scans have been performed provide **important guidance** on what you may need to do next, especially related to protection. You can choose how many scans you wish to incorporate in any given report. Each report improves data visibility, providing detailed insight into:

The names of the infotypes found together with the number of sensitive data matches

The risk associated with each of the data objects analyzed during the scan in question

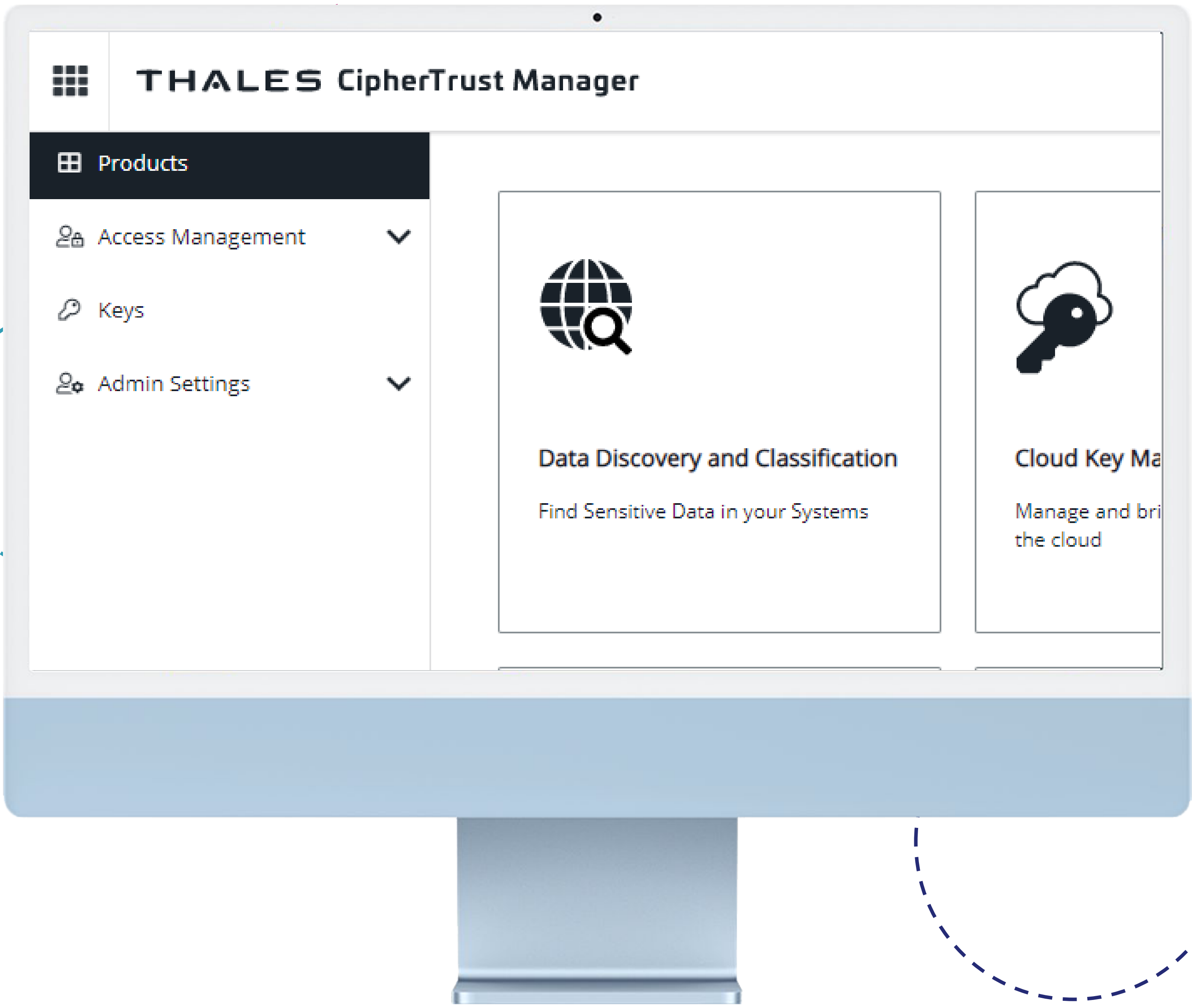
The protection status so that you can see what it protected and what is not

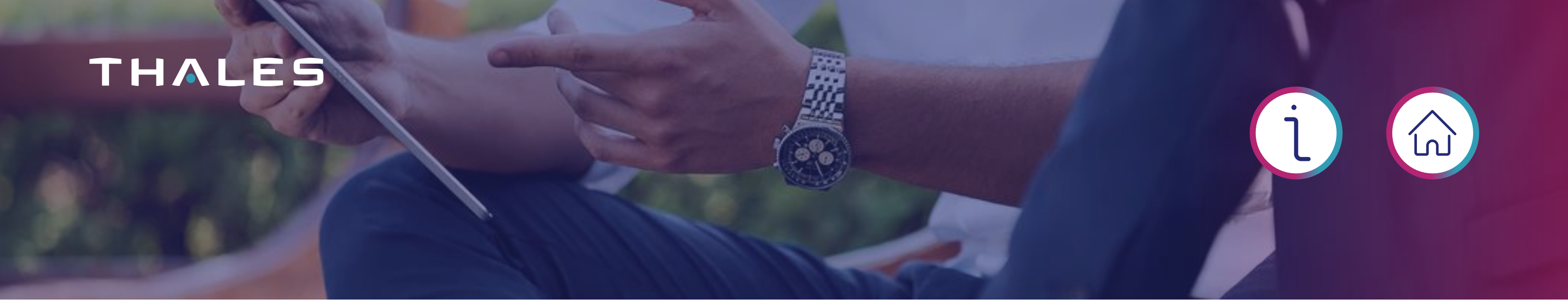
DDC provides options to run scans at any time, over and over again if desired. The tool is intended to be used regularly to keep your insight up to date rather than just being a one-off activity. Since data is always in a state of flux, DDC is ready to convey the latest true representation of what you have and what you might need to do to keep your business safe.

# How it all fits together

DDC is configured using a graphical user interface (launched from within the CipherTrust Manager console) or via the REST API. It offers a simple and intuitive workflow, enabling you to define data locations for interrogation, specify the types of data you wish to discover and decide when and how often you wish to repeat the scanning process. Ultimately the rich views provide a deep dive into the sensitive data matches, the associated protection status and risk assessment of the data locations in question. The main features are summarized in the sections below, providing a high level insight into the power, flexibility and control available to you.

The workflow for DDC is easy to follow via the user interface. **Three main ingredients** help define your environment – **data stores, classification profiles** and **infotypes.**


THALES CipherTrust Manager
Products
Access Management
Keys
Admin Settings
Data Discovery and Classification
Find Sensitive Data in your Systems
Cloud Key Ma
Manage and bri
the cloud

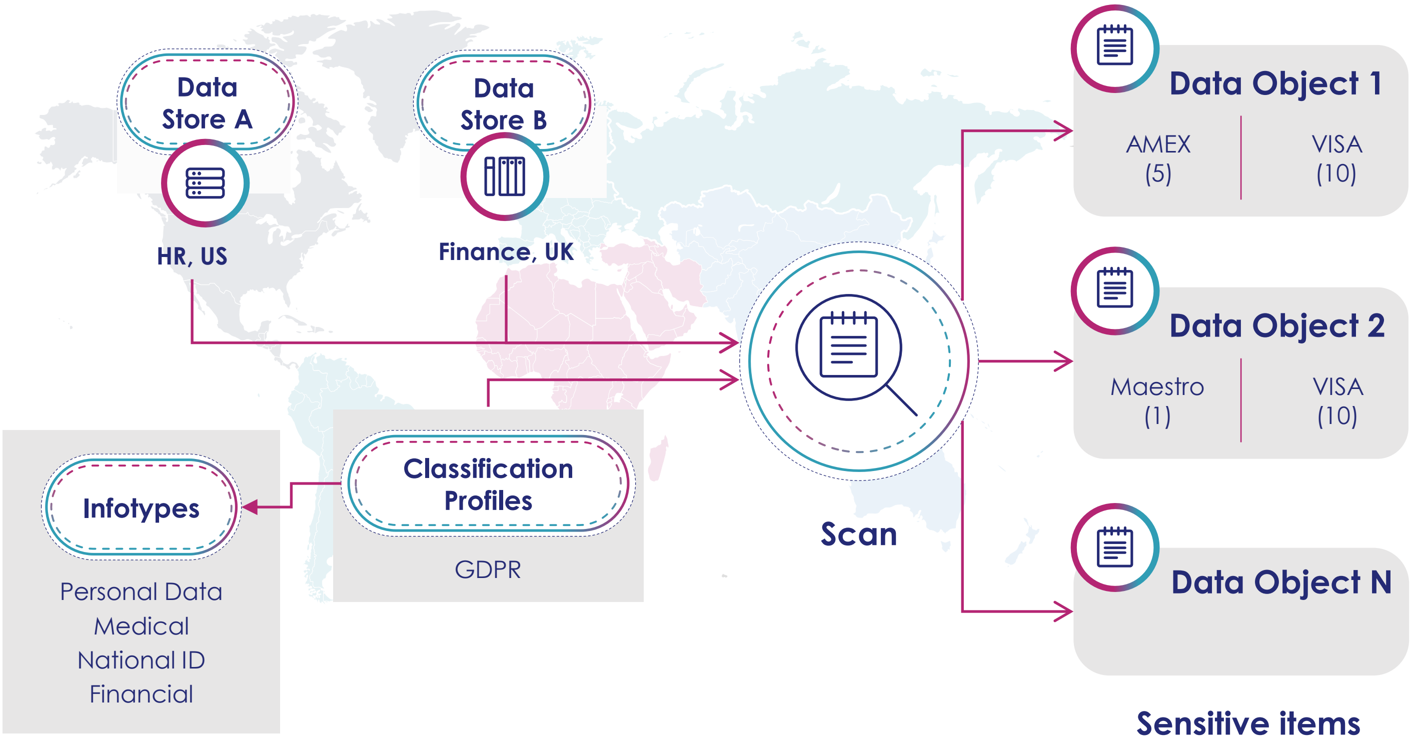


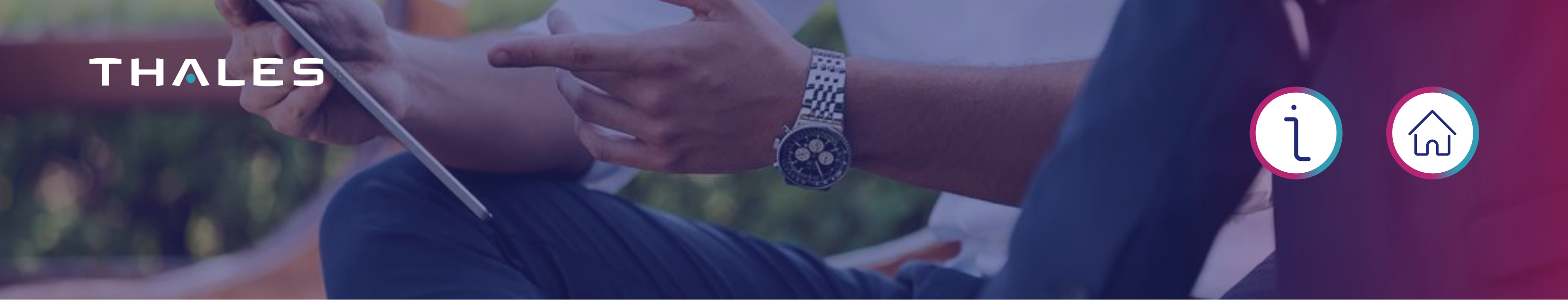
## Typically, you might adopt the following sequence when using DDC for the first time:

Select and configure **data stores** of interest – we recommend concentrating on the most frequently accessed data stores first and a selected path rather than the whole store.

Choose one or more **classification profiles** – create custom classification profiles if necessary. During this process, select one or more **infotypes** of interest - use the custom infotype capability if the prebuilt list is insufficient.

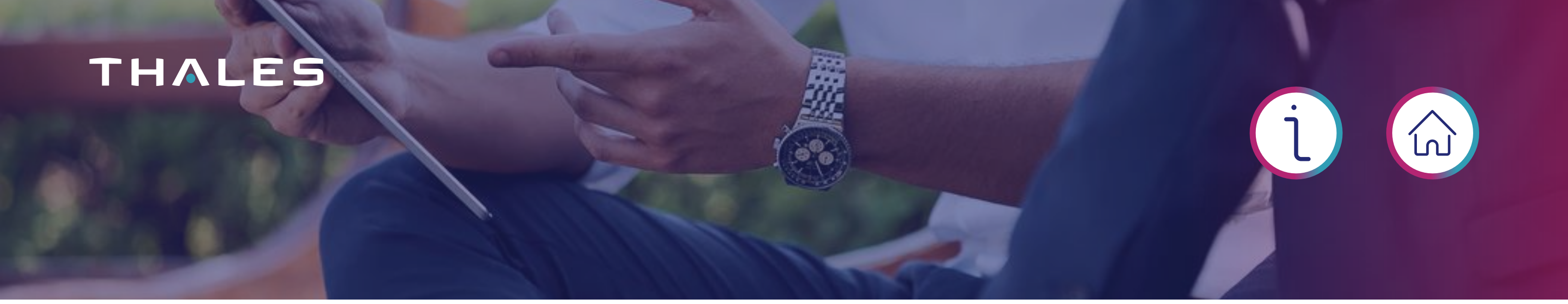
Configure and start the **scan** – consider just selecting one data store for your first scan to test your configuration.





| Data stores | | |
|---|---|---|
| Local storage | Local storage | |
| Network Storage | Windows share (CIS/SMB) | Unix file system (NFS) |
| Database | IBM DB2<br>Microsoft SQL<br>MongoDS<br>MySQL | Oracle DB<br>PostgreSQL<br>SAP HANA |
| Big Data | Hadoop clusters | Teradata |
| Cloud | AWS S3 Buckets<br>Azure Blob | Azure Table<br>Office 365 (Exchange & SharePoint) |

| Types of file supported | | | | | |
|---|---|---|---|---|---|
| **Databases** | Access<br>Dbase | SQLite<br>MSSQL MDF & LDF | **Microsoft Office** | v5<br>6<br>95<br>97<br>2000 | XP<br>2003 onwards<br>Office Files: *Word, Excel, PowerPoint. Access, OutlooK, Other (.pub & .xps)* |
| **Images** | BMP<br>FAX<br>GIF<br>JPG | PDF (embedded)<br>PNG<br>TIF | **Open Source** | Star Office / Open Office / Libre Office | |
| **Compressed** | Bzip2<br>Gzip (all types) | TAR<br>ZIP (all types) | **Open Standards** | PDF<br>RTF<br>HTML | XML<br>CSV<br>TXT |
| **Microsoft Backup Archive** | Microsoft Binary / BKF | | | | |



| Classification profile templates (prebuilt) | | | |
|---|---|---|---|
| **APA** | Australia Privacy Amendment | **HIPAA** | Health Insurance Portability and Accountability Act |
| **APPI** | Act on Protection of Personal Information | **KVKK** | Turkish Personal Data Protection Law |
| **CCPA** | California Consumer Privacy Act | **LGPD** | General Data Protection Law (Brazil) |
| **GDPR** | Financial | **NYDFS** | New York State Department of Financial Services |
| **GDPR** | General Data Protection Regulation | **PCI DSS** | Payment Card Industry Data Security Standard |
| **GDPR** | Healthcare | **SHIELD** | Privacy Shield Framework |
| **GDPR** | National ID | **UK-GDPR** | General Data Protection Regulation (UK) |
| **GDPR** | Personal Details | | |

The structured and unstructured data coverage provided by DDC is frequently being updated with new capabilities – check out the latest specifications here.

Core classification profiles are prebuilt and aligned with major data laws and regulations, both regional and global. You can make changes to these or add totally new profiles if required. Various shortcuts are provided in terms of editing and copying so that you can quickly create and test your own profiles. There is full flexibility in specifying various parameters including sensitivity levels and infotypes.

| Sensitivity levels | | |
|---|---|---|
| **None** | **Public** | **Internal** |
| **Private** | **Restricted** | |

# Categorizing sensitive data

DDC is prebuilt with various infotypes, over 200 and counting, covering the vast majority of regional and global data privacy laws and regulations. When a prebuilt classification profile template is selected, the appropriate subset of infotypes is added automatically. You can modify or extend this list by creating your own custom infotypes, allowing rules to be defined which describe precisely how the scanning engine should look for the data strings in question. Full access to the groups and categories is available as part of the definition so that data matches can be displayed in the appropriate sections of the scan results.

| Infotype categories (prebuilt) | | |
|---|---|---|
| **Financial** | Credit/ debit cards | Bank account info |
| **Personal data** | Email addresses<br>Login credentials<br>Card number<br>Ethnicity<br>License number<br>Roll number<br>Passport number<br>Date of birth<br>MAC address | Mailing address<br>Telephone number<br>Gender<br>Religion<br>IP address<br>Phone number<br>Name<br>Profanity |
| **Medical** | Patient health data | |
| **National ID** | Personal identification | |

# Using agents for discovery

Detailed data analysis is carried out by discovery agents which perform the scanning and report the results back to the DDC Connector for analysis and processing. Two types of agents exist – local and proxy. You would normally install local agents on data store locations (for which you have the appropriate access rights) to ensure the data never leaves the server for security or performance reasons. In contrast, proxy agents are used for network or remote data stores where a separate proxy server hosts the agent.

| Type of agent | Pros | Cons |
|---|---|---|
| **Local** | Faster scan<br>Data remains local<br>No need for credentials | Longer deployment time<br>More complex management<br>Only supports local data stores |
| **Proxy** | Faster to deploy and scale<br>Ability to scan multiple data stores<br>Supports multiple data store types<br>No resources consumed on target host | Data sent over network to agent prior to scanning<br>Increased network load and data footprint<br>Ideally should be located on same virtual LAN (VLAN) |

## Common benefits apply to both types of agent:

An **unlimited number** of agents is available for deployment at no additional cost since licensing is based on the volume of new data scanned

**Multiple simultaneous sub-scans** can be launched to reduce the time required for the overall scan

**TLS encryption** is available for all communications between data stores and agents to protect sensitive data from eavesdropping

**Source data is untouched** as agents only access a temporary data copy in internal memory during the scan

**No sensitive data is stored** during the process as only a summary of the scan results is sent to the DDC Connector

Scans can **still proceed** and record results even when connection to the DDC Connector is lost

With your agents installed, you can then start to think about what and how you wish to scan. We provide a great deal of flexibility in how scans are configured, managed and operated. Important parameters configurable for each scan independently include:

- Name (up to 64 characters)
- Description (up to 250 characters)
- One or more data stores (where to scan)
- One or more classification profiles (what to look for)
- Manual or scheduled operation (when to activate)

On completion of any given scan, it is added to a list of all scans that have been performed. Next, you would use the reporting tool to view detailed results from the scan. Depending on what is found you may make changes to the scan configuration and rerun the scan which is beneficial when working on different data stores or new types of compliance regulations for the first time.

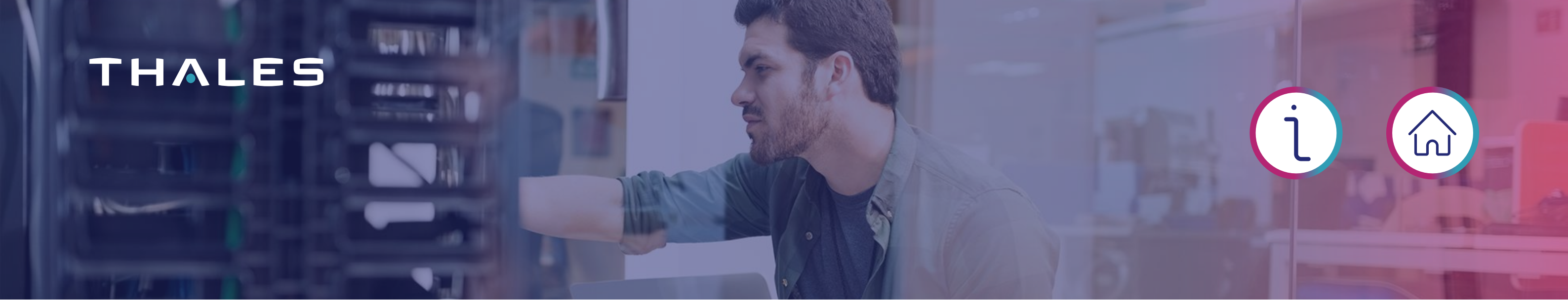
# Analyzing scan results

The results from historical and new scans are available for use with the built-in reporting tool. Three main views for scan reports are offered:

- **Scans**
- **Data stores**
- **Data objects**

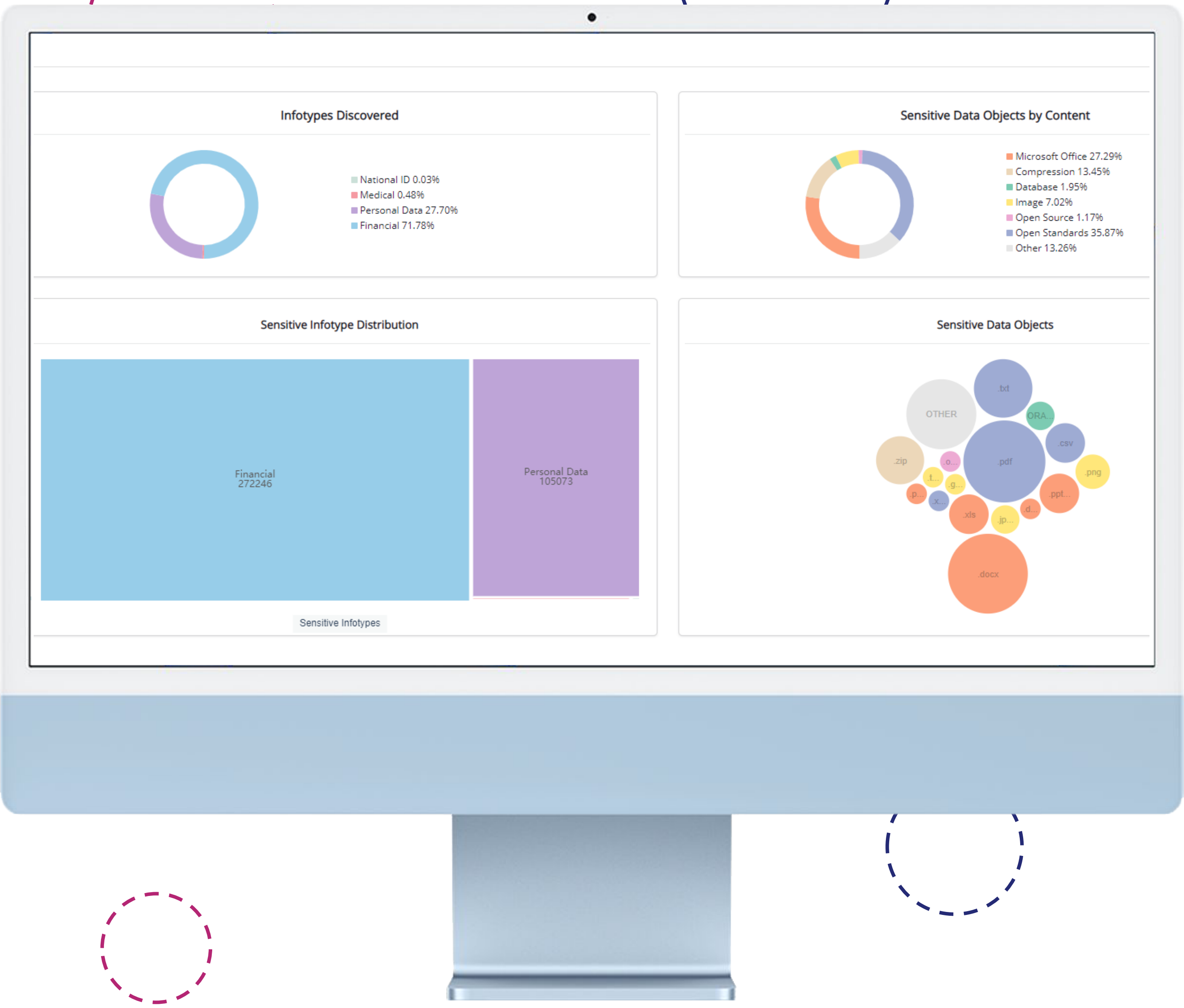
Before detailed information is available, a report is generated which aggregates the results from one or more successful scans. Often for logistical, security or performance reasons you may wish to use multiple scans (each looking for different types of regulated data) to analyze your entire data footprint. With our flexible approach you can still do this and fine-tune each scan as required using multiple iterations, before generating a final report which can represent the current status of all your data. It is also easy to understand your organizational posture from a single regulation point of view.

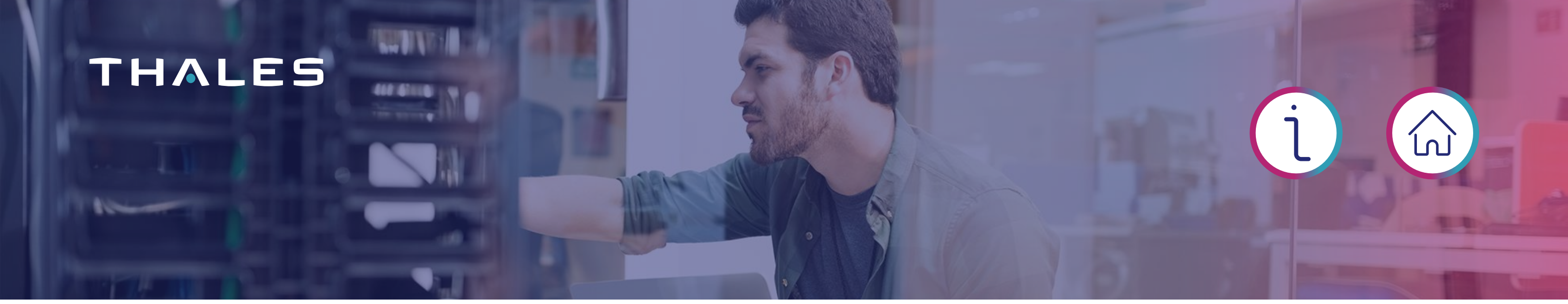
Each tab within the report module offers an insight into slightly different aspects of your data – all however have a common goal in helping you identify quickly any sensitive data at risk.



## Scans view

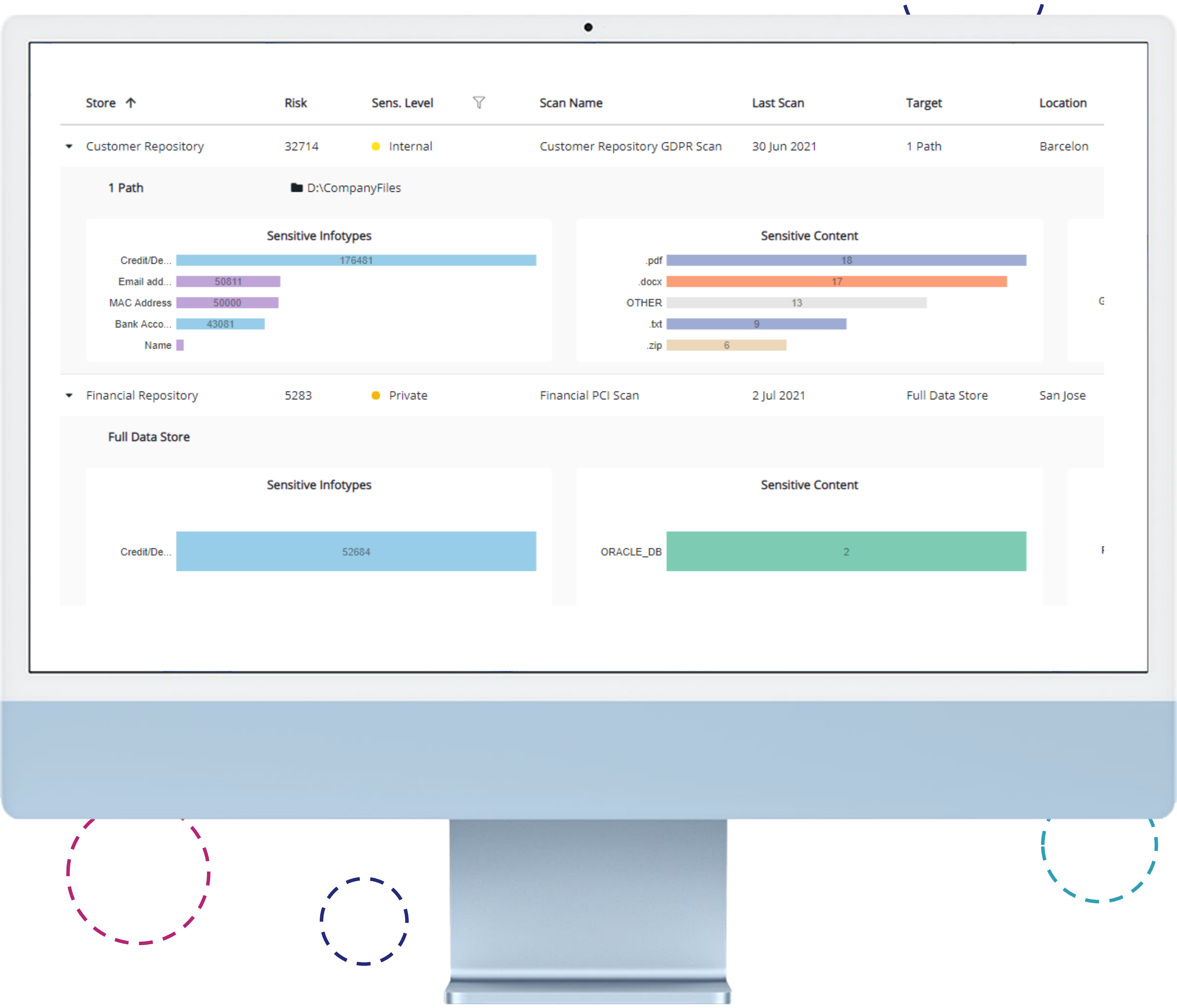
Provides a high level, aggregated scan view of the infotypes found (broken down by percentage in each category present) together with three other charts representing the presence of sensitive data objects in terms of category distribution, content types and file types in question.





## Data stores view

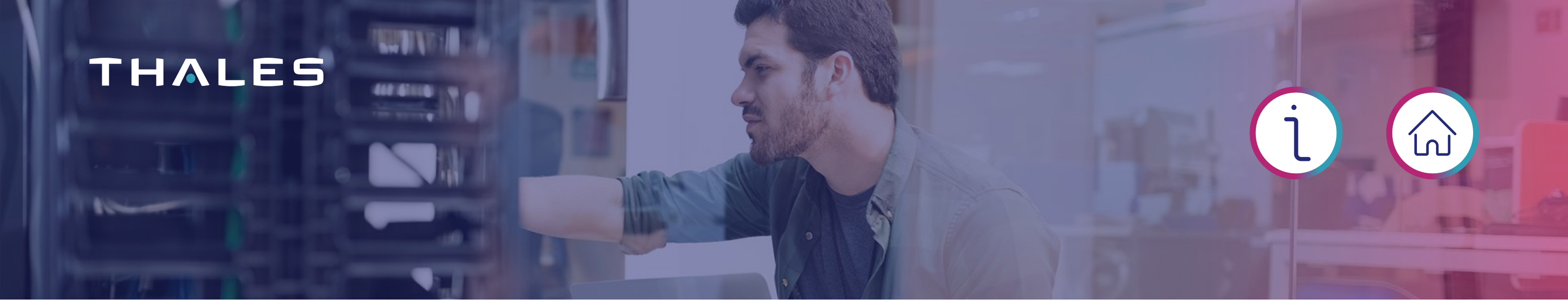
Provides a list of the data stores with details including the risk, sensitivity level, infotypes found and the number of sensitive data objects present for each data store – this provides a rapid means of assessing at a high level whether or not sensitive data is at risk and which data stores are involved.



| Store ↑ | Risk | Sens. Level | Scan Name | Last Scan | Target | Location |
|---|---|---|---|---|---|---|
| ▾ Customer Repository | 32714 | ● Internal | Customer Repository GDPR Scan | 30 Jun 2021 | 1 Path | Barcelon |
| ▾ Financial Repository | 5283 | ● Private | Financial PCI Scan | 2 Jul 2021 | Full Data Store | San Jose |

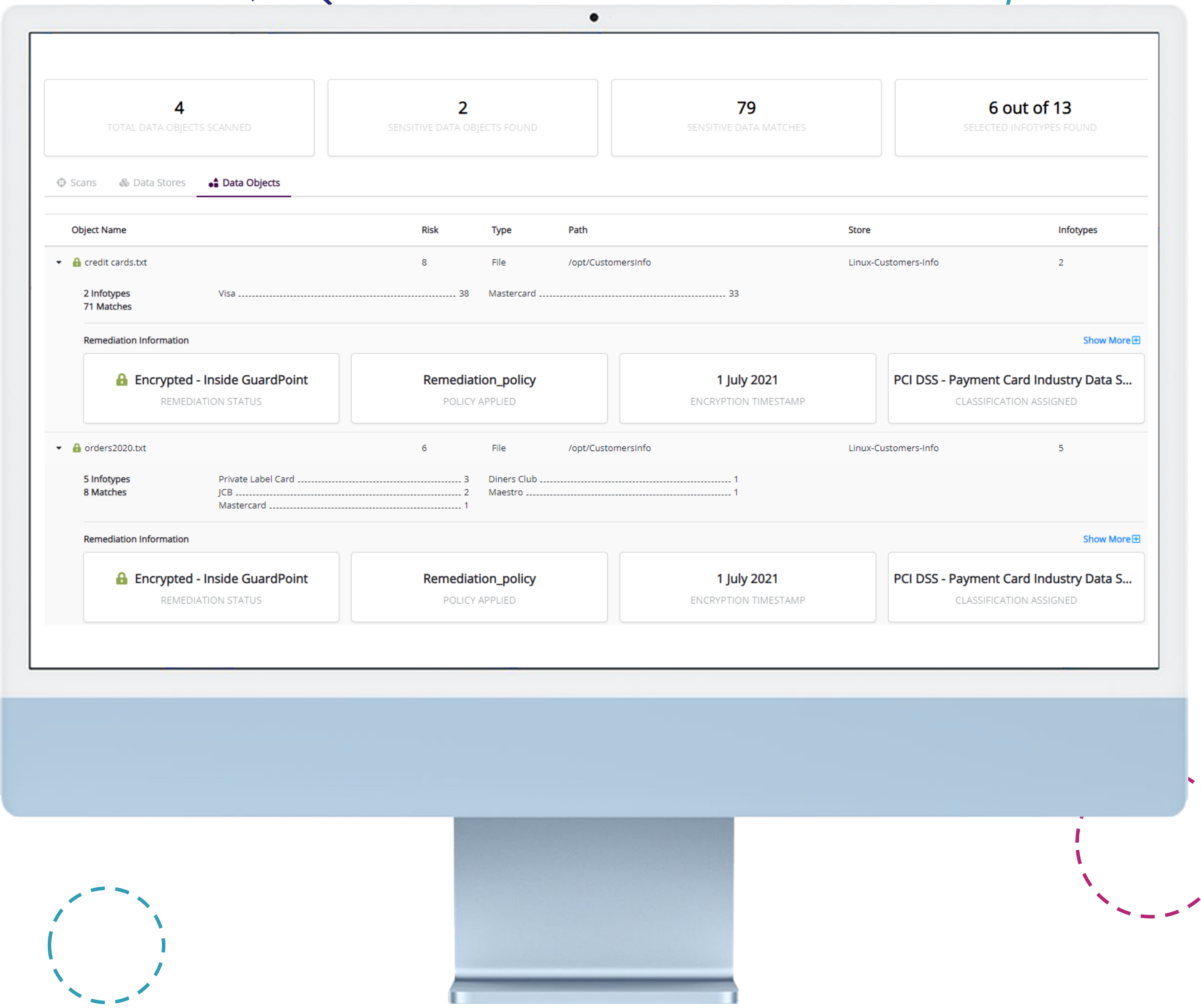
**1 Path** D:\CompanyFiles

**Full Data Store**



# Data objects view

Provides a detailed list of all data objects scanned containing sensitive data, sorted in descending order of risk of exposure. For each sensitive data object in question, you can see where it is located, the data store in which it resides and the individual names and data matches for all of the associated infotypes found. Remediation status information is also present which provides a deeper insight into your sensitive data, confirming whether or not an encryption policy is active. No sensitive data is stored as part of the reporting process.



4 TOTAL DATA OBJECTS SCANNED

2 SENSITIVE DATA OBJECTS FOUND

79 SENSITIVE DATA MATCHES

6 out of 13 SELECTED INFOTYPES FOUND

| Object Name | Risk | Type | Path | Store | Infotypes |
|---|---|---|---|---|---|
| credit cards.txt | 8 | File | /opt/CustomersInfo | Linux-Customers-Info | 2 |

2 Infotypes
71 Matches

| | | | |
|---|---|---|---|
| Visa | 38 | Mastercard | 33 |

Remediation Information

| Encrypted - Inside GuardPoint | Remediation_policy | 1 July 2021 | PCI DSS - Payment Card Industry Data S... |
|---|---|---|---|
| REMEDIATION STATUS | POLICY APPLIED | ENCRYPTION TIMESTAMP | CLASSIFICATION ASSIGNED |

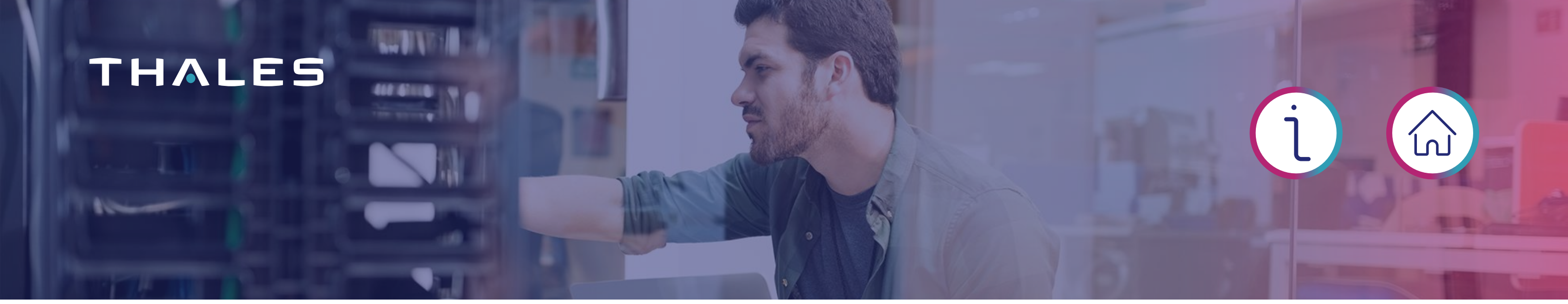
| Object Name | Risk | Type | Path | Store | Infotypes |
|---|---|---|---|---|---|
| orders2020.txt | 6 | File | /opt/CustomersInfo | Linux-Customers-Info | 5 |

5 Infotypes
8 Matches

| | | | |
|---|---|---|---|
| Private Label Card | 3 | Diners Club | 1 |
| JCB | 2 | Maestro | 1 |
| Mastercard | 1 | | |

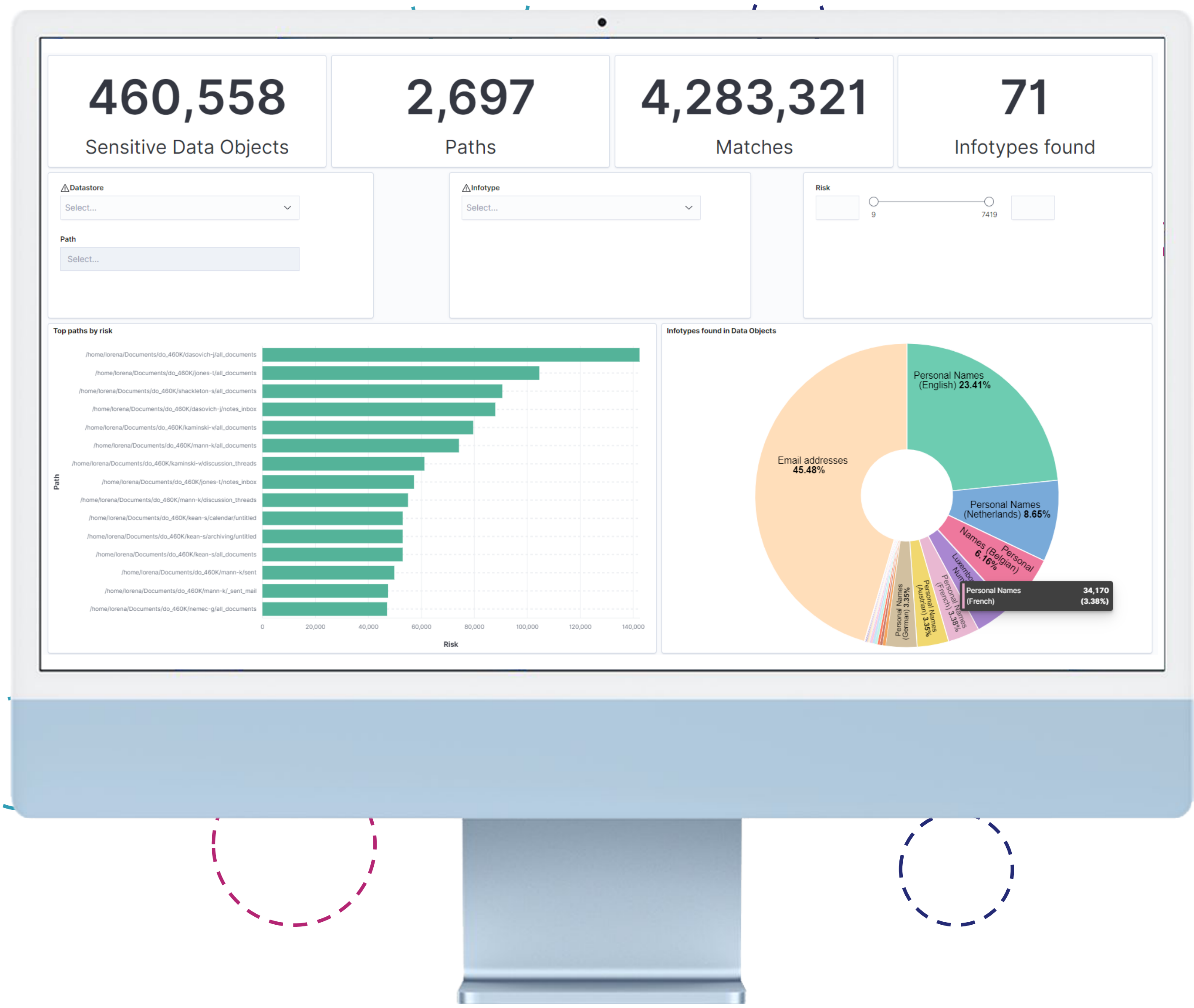
Remediation Information

| Encrypted - Inside GuardPoint | Remediation_policy | 1 July 2021 | PCI DSS - Payment Card Industry Data S... |
|---|---|---|---|
| REMEDIATION STATUS | POLICY APPLIED | ENCRYPTION TIMESTAMP | CLASSIFICATION ASSIGNED |



## Exporting data

In addition to viewing objects using the built-in reporting tool, you can also export data in a NDJSON format for analysis by an external reporting tool such as one from Elastic used in Thales demonstrations.



460,558 Sensitive Data Objects

2,697 Paths

4,283,321 Matches

71 Infotypes found

Datastore: Select...

Path: Select...

Infotype: Select...

Risk: 9 – 7419

Top paths by risk
/home/lorena/Documents/do_460K/dasovich-j/all_documents
/home/lorena/Documents/do_460K/jones-t/all_documents
/home/lorena/Documents/do_460K/shackleton-s/all_documents
/home/lorena/Documents/do_460K/dasovich-j/notes_inbox
/home/lorena/Documents/do_460K/kaminski-v/all_documents
/home/lorena/Documents/do_460K/mann-k/all_documents
/home/lorena/Documents/do_460K/kaminski-v/discussion_threads
/home/lorena/Documents/do_460K/jones-t/notes_inbox
/home/lorena/Documents/do_460K/mann-k/discussion_threads
/home/lorena/Documents/do_460K/kean-s/calendar/untitled
/home/lorena/Documents/do_460K/kean-s/archiving/untitled
/home/lorena/Documents/do_460K/kean-s/all_documents
/home/lorena/Documents/do_460K/mann-k/sent
/home/lorena/Documents/do_460K/mann-k/_sent_mail
/home/lorena/Documents/do_460K/nemec-g/all_documents
Path
0
20,000
40,000
60,000
80,000
100,000
120,000
140,000
Risk
Infotypes found in Data Objects
Personal Names (English) 23.41%
Email addresses 45.48%
Personal Names (Netherlands) 8.65%
Personal Names (Belgian) 6.16%
Luxembo… Num…
Personal Names (French) 3.38%
Personal Names (Austrian) 3.35%
Personal Names (German) 3.35%
Personal Names (French) 34,170 (3.38%)

# Getting answers to key questions quickly

## We make it easy to get up and running, so you can experience the wealth of insight obtainable from DDC

### *What types of regulated data exist in my organization?*

Create a series of scans to search for different types of regulated data

Use one classification profile per scan for optimum performance

Narrow initial scanning to paths within data stores you use the most

Use the reporting tool to collate the scan results

**The extent and categories of sensitive data found are presented clearly**

### *Have I any sensitive data that is not protected?*

Run one or more scans involving the data stores, classification profiles and infotypes of interest

Select the appropriate sensitivity levels

Consider using custom infotypes to help find data proprietary to you

Generate a report incorporating results from one or more previous scans

**Use the subsequent report view to determine if you have any vulnerabilities**

### *What are the priority items to address?*

Use the 'Data Stores' tab view to sort data stores in descending order of risk

Click on the first data store to retrieve the next level of detail

**The breakdown of the sensitive infotypes split by family is displayed, sorted by descending order of occurrence**

Alternatively look at sensitive content by data object by using the 'Data Object' tab where the data object list is sorted by descending risk of exposure

### *Where is the vulnerable data located?*

The 'Data Stores' tab provides information on the precise location of any target paths (if applicable) within the data store in question where sensitive data has been found

**It always provides the name of the location where the data store is present**

### How can I use encryption to protect sensitive data?

Use the CipherTrust Intelligent Protection option to perform discovery and protection in a single step with no manual intervention

An alternative manual approach is to use DDC to find your sensitive data, then use CipherTrust Transparent Encryption (CTE) as a separate manual step to encrypt the data

You may also consider other CipherTrust platform protection solutions such as tokenization, data masking or ACLs as an alternative to encryption

### What happens if some of my data is already encrypted?

Data with no encryption applied can be scanned and analyzed providing the discovery agent has been granted access to the location or path in question

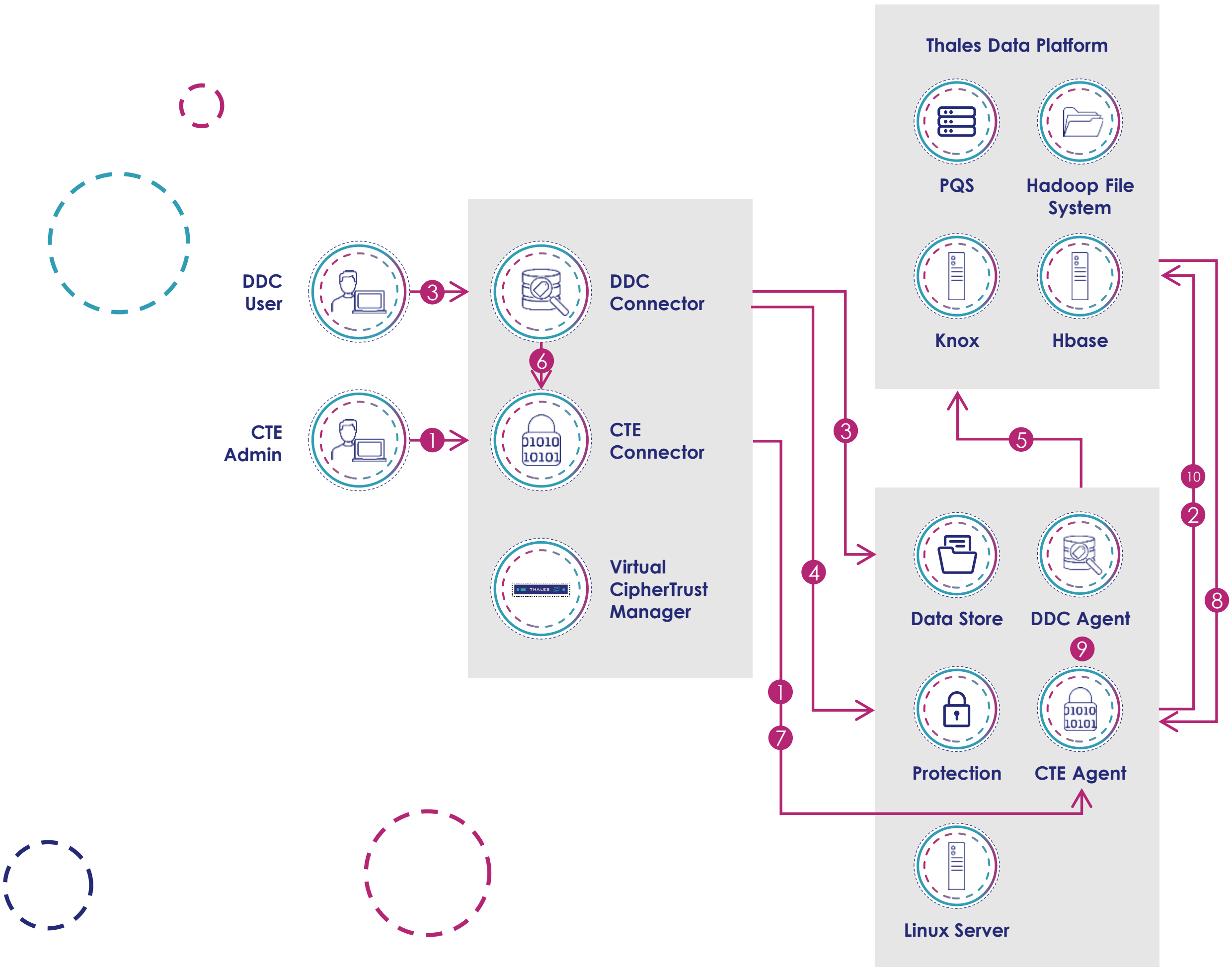
Data encrypted using CTE can be scanned and analyzed providing the discovery agent has been granted access to the location or path in question within the data store

Data that has been encrypted using a third party tool can be scanned and analyzed properly only if DDC has access to the decrypted data – otherwise DDC will scan the encrypted data and not find any match

THALES

# Discovery and protection in a single step

This feature creates an automatic workflow stage that links CipherTrust Data Discovery and Classification (DDC) with CipherTrust Transparent Encryption (CTE) – we call this CipherTrust Intelligent Protection. This is an optional feature that you can switch on or off as required for any given data store or file location which is linked to a GuardPoint accessible to both CTE and DDC.



1. CTE Admin creates a policy (using a classification-based resource set) on the CTE Connector and configures the GuardPoint on the CTE Agent

2. CTE Agent scans the GuardPoint and uploads the list of files to the Thales Data Platform

3. DDC User adds a data store located on the server where the CTE Agent resides and configures the classification profile

4. DDC User creates and initiates a scan on the data store linked to the GuardPoint path and enables protection

5. DDC Agent stores the scan results in the Thales Data Platform on completion of the scan, including the classification for each of the files linked to the GuardPoint

6. DDC Connector notifies the CTE Connector that new scan results are ready

7. CTE Connector notifies the CTE Agent that new classifications are ready

8. CTE Agent downloads the classification information from the Thales Data Platform

9. CTE Agent applies the security rules, to enforce access control, based on policy and classifications

10. CTE Agent encrypts each file according to the policy key rule, updating the file status information in the Thales Data Platform at the same time

THALES

# Reasons for our choice of architecture

## Single integrated platform

We decided to make DDC an integral part of the Thales CipherTrust Data Security Platform because, in our experience, data discovery is the foundation of any effective data security strategy. This integrated platform approach enables CipherTrust Manager to act as the central management console for various connectors including data discovery, transparent encryption and tokenization. If you are using CTE already, you will be familiar with the console and will benefit from consistency in how common management tasks such as user groups, data policies, access control and logging are implemented.

## Accurate pattern matching

Thales performed an extensive investigation to determine the best data discovery technology in the market. After much research and bench testing, the team experienced a significant difference using GLASS Technology$^{\text{TM}}$ over regex. GLASS is a proprietary market-leading technology from Ground Labs (a Thales Technology Partner) for defining and matching simple and complex patterns of data at scale as opposed to a developer code language not built specifically for purpose. It was designed from the ground up with the express purpose of finding patterns of data in modern data sets across both structured and unstructured data storage scenarios.

## High performance discovery

An important goal was to provide a highly efficient and accurate discovery engine. Most data discovery solutions available today use the regular expression (regex) language (originating from the 1950’s) to search for patterns inside text strings. This approach requires extensive developer knowledge to implement and the syntax is not based on English grammar. The biggest enemy of regex is performance – as pattern concurrency and complexity increases, the slower the performance achieved. Regex has limits to the patterns it can match accurately, making it far from ideal in the modern world where there are hundreds of different privacy regulations and a broad range of data management issues to address.

## Easy customization capabilities

Customization support was another important criteria to supplement the comprehensive prebuilt templates. To address proprietary needs, GLASS data patterns can be written by non-developers using English-style grammar to express how a piece of data or pattern exists. Detailed analysis shows it has overcome all important regex limitations for pattern matching techniques, especially its ability to operate across multiple platform and cloud environments, while being very frugal in terms of CPU overhead. This made GLASS an ideal core component for embedding inside DDC, working efficiently and transparently as part of the CipherTrust platform in identifying and securing critical data.

## Flexible licensing model

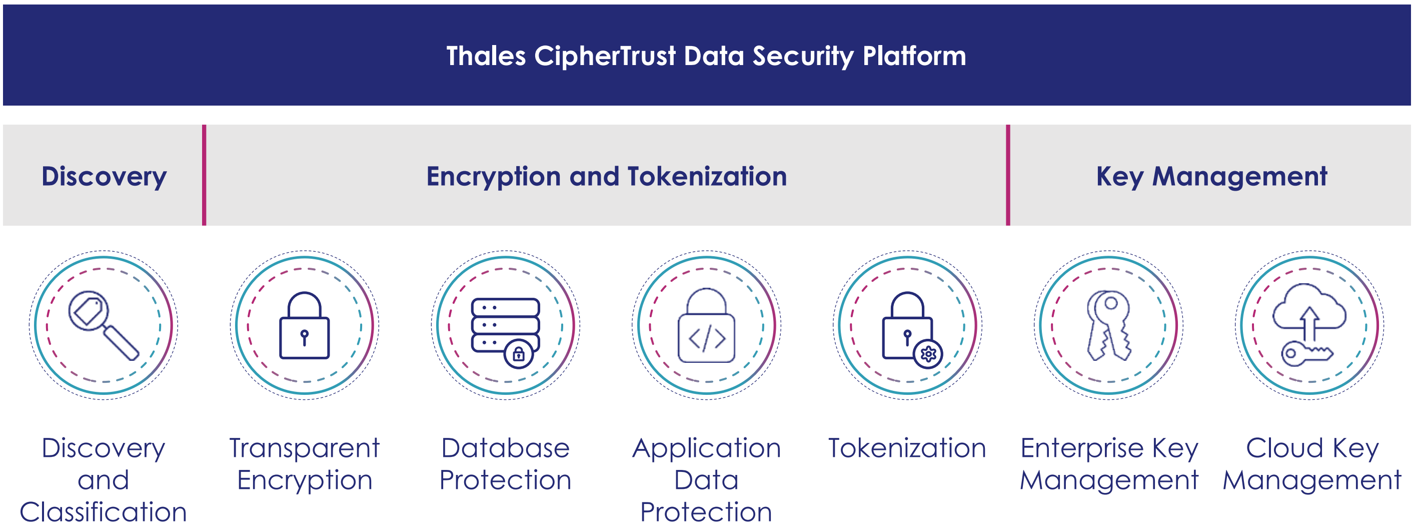
A consumption model based on total data scanned is used to license DDC, providing close alignment with many cloud subscription models and offering an unlimited number of discovery agents without incurring any additional fees. This makes it much easier for all your teams to have access to the data discovery capabilities, with the only real decision to make is how much overall data discovery capacity is required for your organization. Re-scanning existing data does not subtract from your remaining allowance, only new data scanned is counted.

## Security first

Naturally as a market leading security company we design and implement secure solutions. DDC is no exception with strong emphasis on ensuring no sensitive data is stored or vulnerable to exposure:

- Ensuring that all data flowing between your data stores and the agents is protected using strong encryption
- Providing granular access controls so that you can tailor the solution to your exact needs
- Enabling you to secure any sensitive data discovered quickly using the automated workflow associated with CipherTrust Intelligent Protection

# Integrating with other solutions



Thales CipherTrust Data Security Platform

| Discovery | Encryption and Tokenization | | | | Key Management | |
|---|---|---|---|---|---|---|
| Discovery and Classification | Transparent Encryption | Database Protection | Application Data Protection | Tokenization | Enterprise Key Management | Cloud Key Management |

## CipherTrust Data Security Platform

DDC is optimized for use as part of the CipherTrust Data Security Platform, an integrated suite of data-centric security products and solutions that unify data discovery, protection and control in one platform. Consequently you can cover all data protection needs using a single platform from one vendor, in this case Thales. This provides you with a full suite of data protection capabilities including key management, user access control, file-level encryption, application-layer encryption, database encryption, format-preserving encryption, tokenization and data masking.

## CipherTrust Manager

All components utilize CipherTrust Manager as their management console, delivering central management of encryption keys, granular access control and security policy configuration. One key advantage is that our solution is designed to protect your data wherever it resides, on-premises or in the cloud and whatever storage mechanism is in use, such as files, databases, big data or containers. Ultimately you will end up requiring fewer resources dedicated to data security operations, meet your compliance obligations with greater confidence and reduce the business risk to your organization.

It is also possible to deploy DDC alongside other third party data protection tools, such as those supporting data encryption or tokenization for example. It is non-intrusive, complementing tools you have already, rather than replacing them. However, CipherTrust Manager is still necessary since it provides the sole method to launch the data discovery capability and to manage the licenses required for its operation together with keys required for encrypting sensitive data.

Unlike deploying CipherTrust Intelligent Protection, if DDC is used to find the sensitive data and another non-CipherTrust tool from a third party is used to encrypt the data, two distinct steps would be necessary (potentially with a long lapse in-between).

Some of the main reasons Thales customers have been using DDC in conjunction with third party products include:

Trying to fill gaps in coverage since their existing discovery tool cannot scan all their structured and unstructured data

Seeking higher accuracy with data matches to eliminate the limitations they are experiencing with regex technology

Supporting a wider range of operating systems and platforms to interrogate more of their data footprint

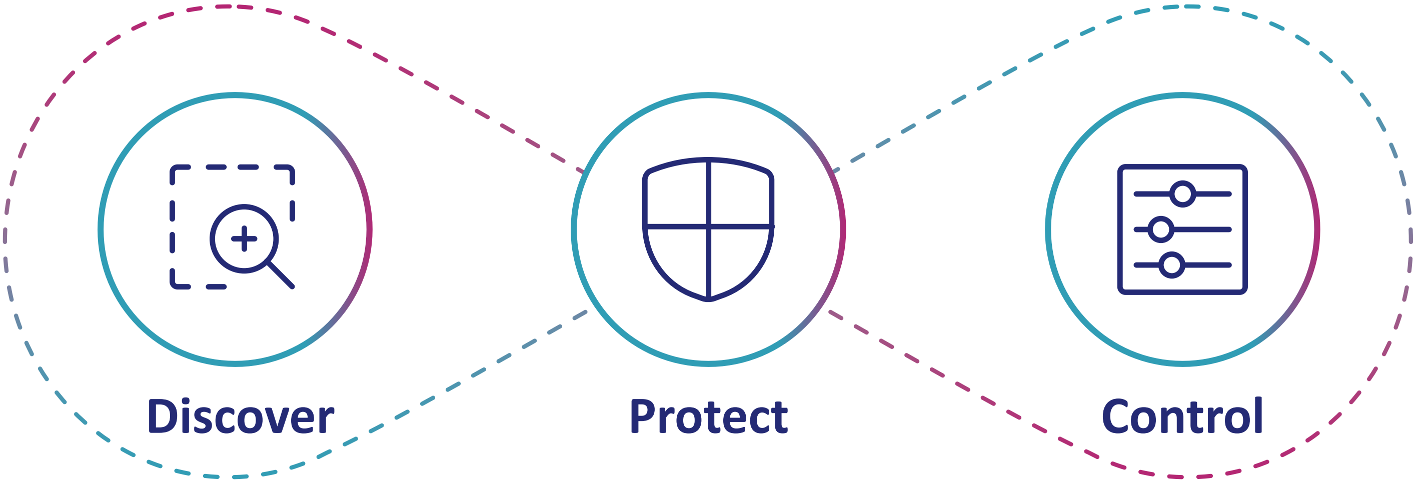
We often get asked…
**'What is the difference between CipherTrust Data Discovery and Classification and other Data Loss Prevention (DLP) solutions?'**

The DLP solutions focus on preventing sensitive data from leaving the organization's perimeter. DDC focuses on data privacy - identifying sensitive data, and getting a clear understanding of data and its risk. This enables appropriate steps to be taken to protect data and comply with data privacy and data security regulations.

THALES

# Key takeaways

## CipherTrust Data Security Platform



Your data management and protection strategy can ultimately be improved significantly using CipherTrust Data Discovery and Classification. In conjunction with CipherTrust Manager and CipherTrust Transparent Encryption it provides considerable visibility and insight into your sensitive data while offering quick and effective protection options. It enables you to:

- Block unnecessary data access to reduce risk of data breaches
- Keep up to date with the evolving regulatory landscape to remain in compliance
- Address digital transformation concerns relating to cloud security to support business and technology goals

**www.cpl.thalesgroup.com**

Try our free 90 day evaluation to see for yourself how it can help your organization now

THALES

# Abbreviations

1/3

| | |
|---|---|
| **ACL** | Access Control List |
| **CDSP** | CipherTrust Data Security Platform |
| **CIFS** | Common Internet File System |
| **CIP** | CipherTrust Intelligent Protection |
| **CM** | CipherTrust Manager |
| **CTE** | CipherTrust Transparent Encryption |
| **DB** | Database |
| **DDC** | CipherTrust Data Discovery and Classification |
| **GUI** | Graphical User Interface |
| **NAS** | Network-attached Storage |
| **REST API** | Representational State Transfer Application Programming Interface |
| **TDP** | Thales Data Platform |
| **NAS** | Network-attached Storage |

THALES

# Glossary

2/3

| | |
|---|---|
| **Classification profile** | A classification profile uses a list of infotypes to define what kind of sensitive information to search for during a scan |
| **Connector** | A connector is a generic term which relates to the different licensable products or components (which includes DDC and CTE) that are managed using the CipherTrust Manager console |
| **CTE agent** | A CTE agent is a software component that is used to encrypt data associated with GuardPoints defined using CTE policies |
| **Data match** | A data match occurs during a scan when a qualified instance of an infotype actively included in the scope of the search is found |
| **Data object** | A file or database table located inside a data store is known as a data object |
| **Data store** | A data store is the entity where the data is actually stored, with DDC supporting various types - local, network, database, Big Data and cloud |
| **DDC agent** | A DDC agent is a software component that is used to scan a data store for specific types of data defined using infotypes in the associated classification profile |
| **Discovery** | Data discovery is the process of mapping an organization's data assets including their locations, types and sensitivity levels |
| **False negatives** | False negatives are where the discovery process fails to identify one or more valid instances of sensitive data |
| **False positives** | False positives are where the discovery process wrongly identifies sensitive data matches where none exist |
| **GuardPoint** | A GuardPoint specifies the list of folders or paths to be protected – access to files and encryption of files under the GuardPoint is controlled by security policies |
| **Infotype** | An infotype is used to categorize specific data (such as passport numbers or email addresses) to look for during a discovery scan, forming an integral component in the definition of a classification profile |
| **Policy** | A policy is a collection of rules that govern data access and encryption |
| **Remediation** | Remediation refers to the process of protecting (or securing) data that has been identified as vulnerable as the result of a scan – typical remediation methods include encryption, tokenization, data masking and access control |
| **Resource set** | A resource set refers to the files or directories to which the policy will apply together with the associated governing key rules |

# Glossary

3/3

| | |
|---|---|
| **Risk** | A risk is the presence of a sensitive data object in a data store and is directly related to the data matches found in the data object or data store |
| **Scan** | A scan is part of the discovery process that is used to search for sensitive data within data stores using criteria defined in classification profiles |
| **Sensitive data object** | A data object that contains any data match is known as a sensitive data object |
| **Sensitivity level** | The sensitivity level is a mandatory parameter (when defining data stores and classification profiles) relating to the degree of potential vulnerability of any given data object if exposed |
| **Structured data** | Structured data is highly organized and easily understood by machine language – examples include names, dates, addresses, credit card numbers, stock information, geolocation and more |
| **Tag** | A tag helps group data together and can be specified when creating data stores and classification profiles |
| **Thales Data Platform** | The Thales Data Platform (TDP) is a Big Data platform based on Hadoop technology which is used by DDC for various tasks including storage of scan results |
| **Unstructured data** | Unstructured data is information that either does not have a pre-defined data model (or schema) or is not organized in a pre-defined manner, making it unsuitable for storage in a relational database |