Data Discovery
Methodology

Machine Learning plus Regex offers the best solution for identifying data, formats, tagging, and compliance.

cpl.thalesgroup.com

THALES
Building a future we can all trust

Data Discovery and Classification is the process of identifying and classifying your data and the personal information within it, allowing you to be more secure and more compliant. Data discovery and data classification are the two most essential components of data management.

Data discovery is the process of identifying and locating data sources, understanding data formats, and assessing data quality. Data classification is the process of labeling and tagging data based on its sensitivity, value, and regulatory requirements. Data is most commonly classified as public, internal, confidential, or restricted.

The Challenge

Traditional data discovery methods have served organizations well in detecting well-structured patterns. However, as data formats evolve, these rule-based methods reveal critical weaknesses:

- Lack of Flexibility: Regex patterns struggle with unstructured or dynamic data, such as addresses, organization names, or proprietary documentation lacking of standard patterns.

- Context Blindness: Pattern matching lacks contextual awareness—distinguishing "Apple" as a fruit versus a company is challenging.

- Manual Maintenance: Every change in data format requires updates to regex rules, making it resource-intensive and unsustainable.

- Data Evolution Complexity: As seen with telephone numbers, formatting changes over time, requiring frequent updates to detection mechanisms. As businesses navigate these challenges, the need for an adaptive, intelligent, and scalable solution has never been clearer.

The Solution

Thales Data Discovery Methodology, powered by AI/ML provides the following to you.

- Adaptive Learning: Models like BERT understand context and continuously improve based on labeled training data.

- Context Awareness: AI recognizes the difference between similar terms based on surrounding content.

- Scalability: Once trained, models can detect sensitive data without requiring constant rule modifications.

By integrating AI-driven data discovery with Regex data discovery, businesses gain greater accuracy, reduced manual intervention, and improved compliance.CipherTrust Data Discovery and Classification enables developers to have a single source and process for secrets discovery, that eliminates downtime, and promotes security and compliance. Become more efficient with CipherTrust Data Discovery and Classification.

Benefits

Greater Visibility of Seamless Compliance

Enhance data visibility to effortlessly navigate broad regulatory frameworks like GDPR and CCPA, ensuring compliance with greater confidence

Smart Access Control

Improves security strategy by implementing **precise access control across** different departments based on document categories.

Operational Efficiency

Reduce the response time for data subject requests **from days, weeks, or months to just minutes.**

CipherTrust Data Discovery and Classification (DDC) provides complete visibility into the location of sensitive data enterprise, empowering you to make better use of your data while mitigating risks and ensuring compliance

70%

Almost 70% of enterprises are able to classify only 50% or less of their sensitive data.
Source: 2024 Data Threat Report

Features

Data Discovery & Classification

Data discovery is the process of identifying and locating data sources, understanding data formats and assessing data quality.

Data Classification is the process of labeling and tagging data based on its sensitivity, value, and regulatory requirements. Data is most commonly classified as public, internal, confidential or restricted.

Enhanced Security & Risk Analysis

Risk analysis is the process of understanding what risks your data presents.

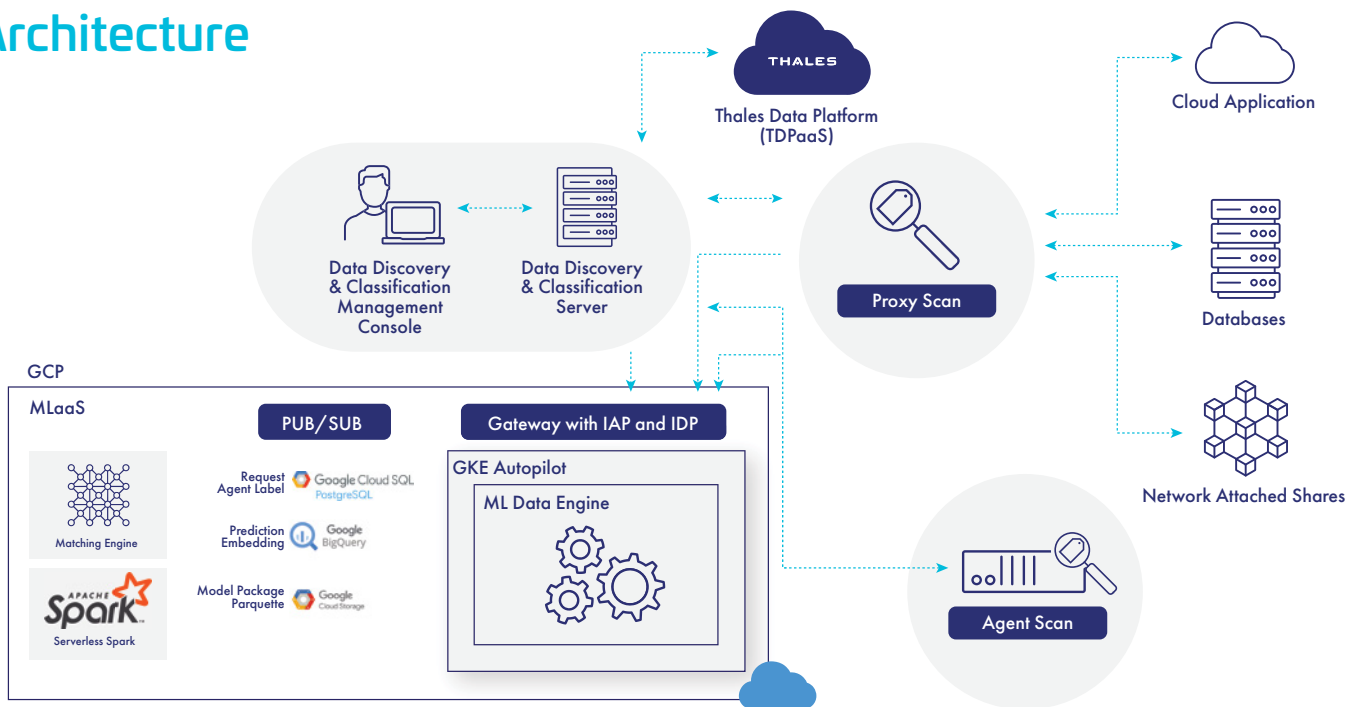**Enhanced security** is the process of taking advanced security measures based on its risk and classification.

Compliance & Governance

Compliance is the process of following a set of rules, laws, or standards.

Governance is a system of rules and practices that help guide how data is managed.

Architecture

Thales Data Platform (TDPaaS)

Cloud Application

Data Discovery & Classification Management Console

Data Discovery & Classification Server

Proxy Scan

Databases

GCP

MLaaS

Matching Engine

Serverless Spark

PUB/SUB

Request Agent Label — Google Cloud SQL PostgreSQL

Prediction Embedding — Google BigQuery

Model Package Parquette — Google Cloud Storage

Gateway with IAP and IDP

GKE Autopilot

ML Data Engine

Network Attached Shares

Agent Scan

Why choose Thales for Data Discovery and Classification?

Thales' CipherTrust Data Discovery and Classification (DDC) is chosen for its ability to provide comprehensive visibility into sensitive data across diverse environments, enabling organizations to enhance security, accelerate compliance, and manage data effectively.

cpl.thalesgroup.com

Contact us – For all office locations and contact information, please visit cpl.thalesgroup.com/contact-us