Understand General Availability of IDM

Contents

Introduction

This document describes the general availability of the Index Document Match (IDM).

Overview

The IDM, is an advanced DLP data classification technique that significantly enhances the organization's ability to effectively safeguard documents containing sensitive data.

With IDM, organizations can index and fingerprint the contents of documents that hold their sensitive data. By creating a fingerprint repository of this data, our Data Loss Prevention (DLP) product can efficiently identify complete or partially matching documents during content evaluation.

The advantage of IDM over traditional pattern matching using regular expressions and keywords is significant. Instead of matching against anything that can resemble sensitive data, IDM allows you to match against your actual sensitive data. This targeted approach reduces the number of low-significance DLP incidents and enables organizations to focus their security operations and resources on high-value investigations.

How is IDM different than EDM?

IDM (Indexed Document Match) and EDM (Exact Document Match) differ in terms of the type of data they fingerprint.

EDM specifically focuses on fingerprinting tabular data, which is structured data organized in a table format. This means that EDM is designed to handle data with a specific structure, such as databases or spreadsheets. For example, an organization can use EDM to fingerprint a corporate credit card table, ensuring that only those corporate credit cards are monitored and protected.

On the other hand, IDM is used for indexing and fingerprinting freeform documents, which are unstructured data that do not use a specific format. IDM is capable of processing and fingerprinting documents that are not organized in a table-like structure, such as text files, PDFs, or Word documents.

In summary, IDM is used for unstructured data fingerprinting, while EDM is used for structured data fingerprinting.

What are common use cases for using IDM?

Some common scenarios include fingerprinting and safeguarding intellectual property, such as source code repositories, patent filings, or sensitive corporate information like HR employee forms, corporate documents, and legal documents.

Does IDM generate fingerprints based on the file or its textual content?

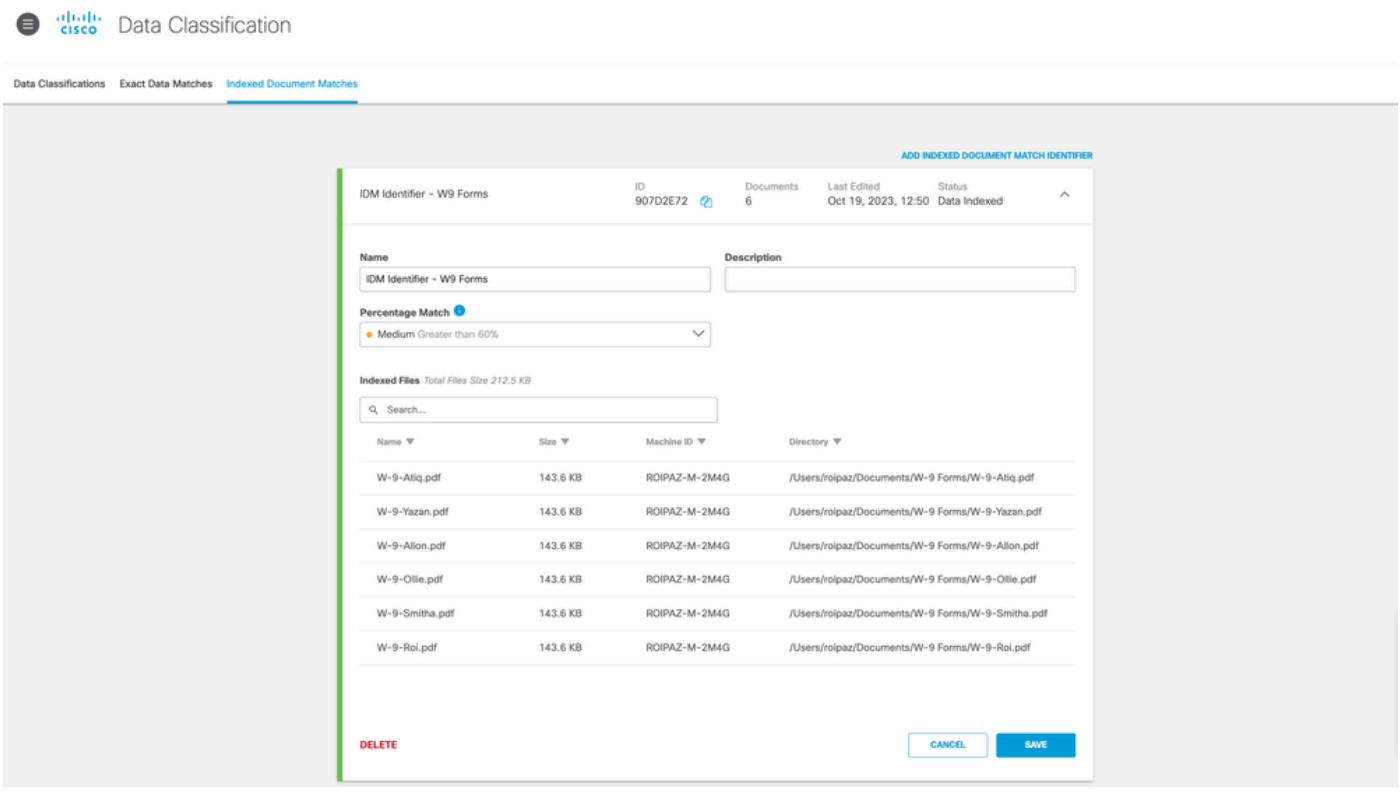
IDM indexes and fingerprints the textual content of the document rather than the file itself. This enables IDM to partially match against evaluated content, even if some of the sensitive data is copied and pasted into a new file. You have the flexibility to specify the extent of matching required to trigger a violation, selecting from a predefined list of options (20%, 60%, 80%).

How to use IDM?

Indexed Document Match (IDM) in Umbrella operates by generating hash fingerprints of the extracted text from sensitive documents. These fingerprints are then used by the various scans of Multi-Mode DLP to completely or partially identify the content of the documents. To generate these fingerprints, you need to download and use Cisco's DLP Indexer tool locally.

The indexer, a command-line interface, extracts text from the documents, performs fingerprinting and indexing operations and then hashes the indexed text. The tool subsequently uploads the hashed fingerprints to Umbrella or Secure Access.

The output of using the indexer tool is a new IDM data identifier type to be used in custom data classification. These classifications are applied with both Real-Time DLP rules and SaaS API DLP rules to effectively protect both data at rest and data in motion.



Can the DLP Indexer Tool be scheduled to fingerprint new data periodically?

The Indexer tool can be run in monitor mode as a background process. This mode enables the DLP indexer to automatically reindex at regular intervals, ensuring that source data is regularly updated in Umbrella without the need for manual operation.

Where to access IDM and download the DLP Indexer Tool?

1. Log in to the Umbrella dashboard.
2. Navigate to Policies > Policy Components > Data Classification > Data Classification.
3. Click on the Indexed Document Match tab.
4. In this section, you can create IDM Identifiers and download the DLP Indexer.

Which file types are compatible with IDM?

IDM supports all file types that are supported by DLP. You can find the comprehensive list of supported file types in the documentation. It is worth mentioning that IDM also supports Unicode characters.

What limitations must be taken into consideration when using IDM?

The total amount of indexed text for all the IDM data identifiers in an organization must not exceed 1 GB. The Indexed Document Matches tab on the Data Classification page displays warnings as the allotted quota is reached.

Where can I find more information?

Umbrella Documentation