

Mitigate Privacy Risks with Data Analytics

Insight AnalytiX - Privacy Risk Classifier

Trends and Challenges

Enterprises today are faced with a diverse infrastructure that ranges from on premise storage to the cloud. As organizations aim to leverage analytics and drive deep insights from their unstructured data, the growth of data and its complexity has left them exposed to risk and high costs.

Unstructured data is challenging to manage, as it does not easily lend itself to the more traditional models of data storage and analysis. However, if you can find a better way to manage this kind of data and understand its content, that unstructured data can be an extremely rich source of relevant information that can add tremendous value to your business.

Digital Subject Access Rights (DSAR) requests require enterprises to respond within **72 hrs.**

Enterprises are required to respond within **1 month** to GDPR requests.

Why Unstructured Data?

In the traditional data space, structured data is what most have dealt with and are comfortable with. Not only is it easy to identify and understand what data elements you have based on database schemes, but you also typically have a fairly good idea of where your data resides, because it is managed. Additionally, there are tools today that help you to manage and catalog your data in the structured world.

Unlike this structured world, where information is stored in a pre-defined data model, unstructured data:

- neither has a pre-defined data model nor is it organized in a predefined manner
- comes in many unrelated forms, such as documents, text files, spreadsheets, presentation files, video files, audio files, and pictures, which are the more common formats
- is typically stored in individual files without an organizational application organizing and managing the relationships between those files

There is currently no easy way for us to understand what is really stored in the unstructured space without physically opening the files and looking. There are, of course, machine learning and artificial intelligence tools available to help you parse some of the unstructured world, but those type of projects require time, money, and the right data science team to train and productionize the model.

Mitigate Data Risk with Insight AnalytiX - Privacy Risk Classifier

Insight AnalytiX is a platform that runs Privacy Risk Classifier which recognizes files containing Personal Identifiable Information (PII). It processes more than 200 file types, including:

- known file formats like XLSX, CSV, PPT, DOCX, and DOC
- images such as BMP, JFIF, or JPEG
- scanned documents as in PDF files

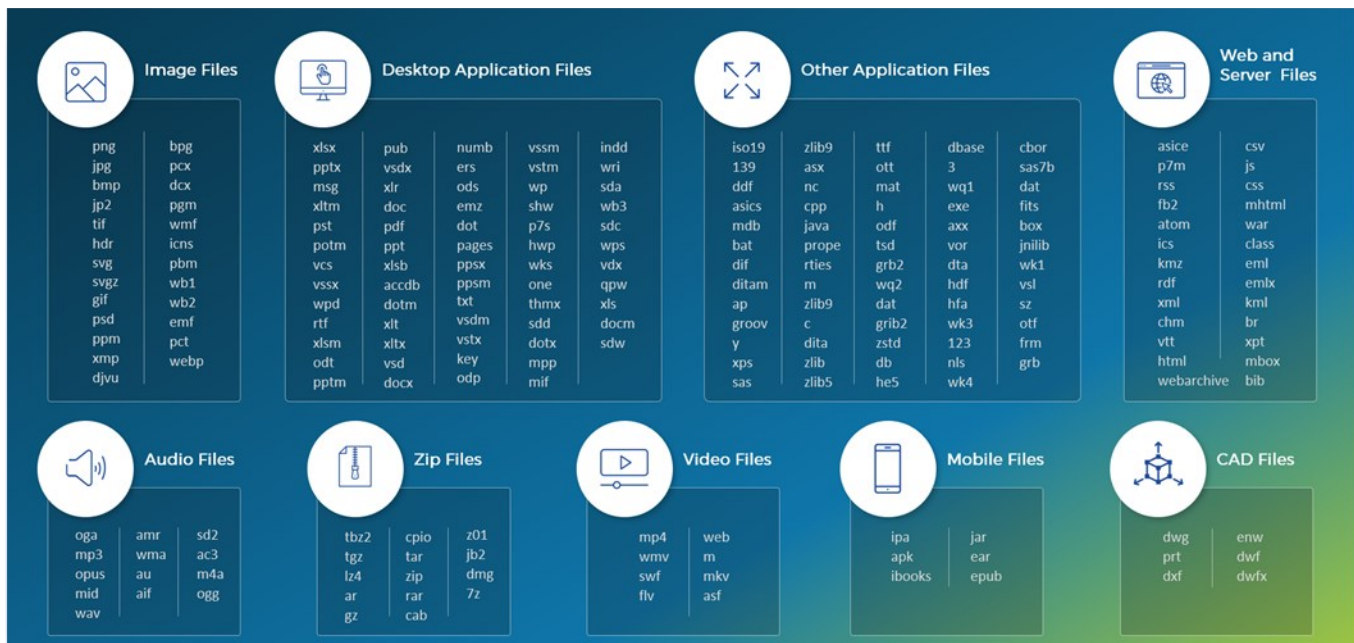


Figure 1: Insight AnalytiX Supports 200+ Unstructured Data Formats

Privacy Risk Classifier recognizes 49 different types of personally identifiable information, which include names, email addresses, mobile numbers, social security numbers (SSNs), and addresses.

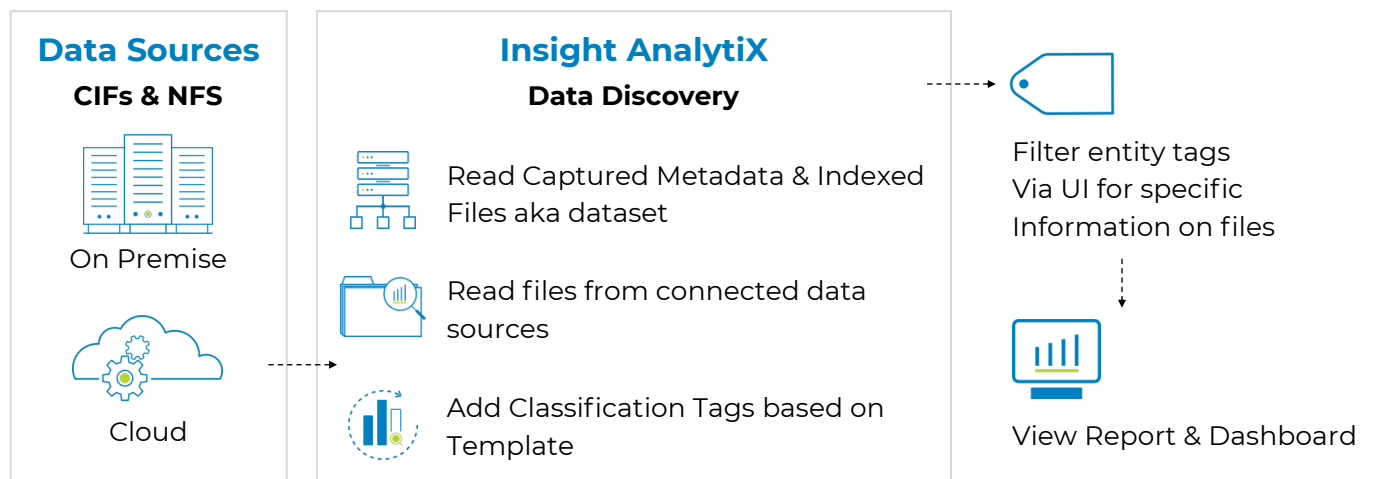
It is built on top of pattern recognition technology, keyword recognition technology, and artificial intelligence (AI). It comes with AI models, a pattern dictionary of 500+ advance patterns, and a 200,000+ keyword dictionary. Role-based access (RBAC) is also part of the functionality, as Privacy Risk Classifier has a built-in authentication mechanism that can integrate with an existing active directory server.

How Does it Work?

As a prerequisite, Insight AnalytiX requires StorageX to be installed on the same VM. A user creates an analysis set in StorageX before they can start using Insight AnalytiX's Privacy Risk Classifier. Privacy Risk Classifier works in coordination with StorageX and fetches lists of analysis sets from the StorageX server.

Once a user initiates a data discovery, Privacy Risk Classifier connects to network-attached storage (NAS) device and streams the files of the selected analysis set. It uses NFS (v3) and SMB (v2 & v3) protocols to stream files from NAS devices.

Privacy Risk Classifier reads the file content of supported file types and the file metadata for unsupported file types, with the help of various file readers. It also uses the built-in Data Science Engine to find out if any scanned file contains any personal identifiable information. The complete analysis is recorded in the application, and the user can see view the results in various graphical formats.



1. Analyze & Classify your Unstructured Data

After an analysis set is processed, the user can select a template within Privacy Risk Classifier and view the data analysis. Each template includes a set of specific PII that are grouped for different types of use. Privacy Risk Classifier analyzes file data using the below metrics:

PII Content	Personal Files vs Non-Personal Files vs Not analyzed Files
Risk Analysis	High Risk Files vs Low Risk Files
Top 10 PII values found	Top 10 PII values found for an analysis set
Top 10 file types found	Top 10 File Types values found for an analysis set
File size discovered	File Size ranges found for an analysis set
Last Modified Date	Files group in the three groups based, modified in last six month, modified in six to 12 months and modified before 12

The user can then download reports in PDF, XML, XSL, and CSV file formats, as desired.

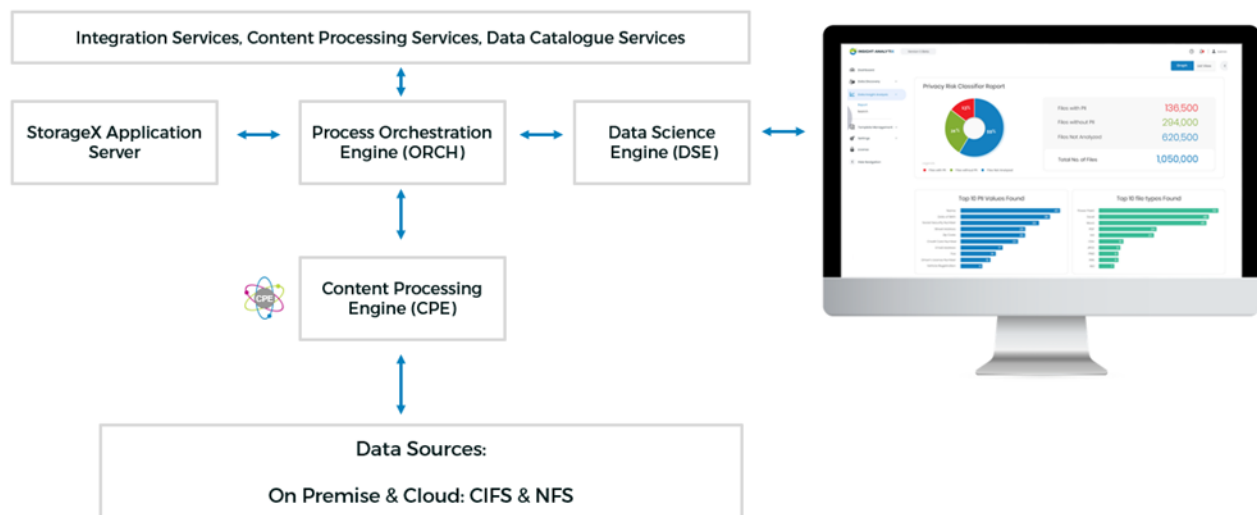


Figure 2: Main Components of Insight AnalytiX

2. Search

Insight AnalytiX has full text-search functionality. It allows a user to search a processed analysis set or all processed analysis sets in their environment. A user can also search with the help of “AND” and “OR” operators, giving the user the flexibility to search more intelligently.

For example, using the AND operator, a user can make all words in the search text mandatory in the search results; in the case of the OR operator, a user can make at least one of the search text words mandatory in the search results.

Search returns the search results using file type-based filters. It allows user to see results for a file type. The user can export results in CSV file format.

3. Role-Based Access

Insight AnalytiX includes three pre-defined roles and allows users to create roles based on their specific needs.:

System Administrator	Access all Insight AnalytiX – Privacy Risk Classifier functionality
User Administrator	Configure users, roles, and application configuration
User	Create templates, run scans, download results, and search analysis sets

Insight AnalytiX comes with its own user authentication mechanism or can be configured to authenticate users against an existing Active Directory server.

Conclusion

The key to managing your unstructured data – and therefore optimizing your storage footprint – is knowledge. Knowing what data you have in your environment is essential to figuring out what you can and should do with that data.

Since StorageX is tightly integrated with Insight AnalytiX, users have an unprecedented ability to dig deeply into their unstructured data and find data that could be potentially sensitive or need to be organized or stored differently.

Using Privacy Risk Classifier, you can quickly and easily find and categorize your data, then use the analysis results to determine what data may need to be migrated, what sensitive data may need to be secured, and what data should be removed entirely. This information powers data optimization on existing storage and make better plans for future storage usage, as well as avoid potential liabilities.

It is likely that these initiatives are now front and center within your 2020 ‘must haves’.

Find answers to you questions by contacting solutions@datdyn.com

See Insight AnalytiX in Action and [Book a Demo](#)

Data Dynamics is the pioneer of unified unstructured data management platform. With 26 of the Fortune 100 global customers worldwide, Data Dynamics enables organizations to analyze, move, manage, and modernize their data anywhere. Data Dynamics accelerates the adoption of hybrid, public and multi-cloud strategies, builds higher quality SLAs, and improves business process modernizations. For more information, visit www.datadynamicsinc.com.



Copyright © 2023 Data Dynamics, Inc. All Rights Reserved. The trademark Data Dynamics is the property of Data Dynamics, Inc. StorageX is a registered trademark of Data Dynamics Inc. All other brands, products, or service names are or may be trademarks or service marks and are used to identify products or services of their respective owners.