NetApp

NETAPP
DATA CLASSIFICATION

Integrated data intelligence for your NetApp data estate. NetApp Data Classification gives you actionable insights into your data to maintain compliance, optimize storage, accelerate data migrations, and prepare data for GenAI and retrieval augmented generation (RAG).

Unleash the Power of Intelligent Data Governance

In today's digital age, data is the lifeblood of any organization. But as data volumes explode and environments become more complex, how can you ensure that your data is not just managed, but harnessed for its full potential? Enter NetApp Data Classification—your partner in transforming data chaos into data clarity.

Data Classification, a core capability within NetApp Data Services, is a robust data governance service that provides comprehensive visibility for managing data across your NetApp footprint more effectively.

Data Classification automatically maps your data, determining how much data exists, where it's located, and the types and categories of the data. This enables you to make intelligent decisions on your data in real time and take action to optimize storage, accelerate data migrations and prepare your data for GenAI and RAG—reducing risk and costs.

Using advanced AI, NetApp Data Classification simplifies data governance, giving you actionable insights to address data privacy, security and compliance requirements.

Quickly uncover compliance and security risks

Data discovery and classification

Identifying sensitive data is complex and especially important in enterprise environments. Often the sensitivity is specific to the organization (and potentially a specific domain or language). To define sensitive data accurately, AI is a must.

Data Classification goes much further than traditional pattern matching. Data Classification uses AI, machine learning (ML), and natural language processing (NLP) technologies to categorize and classify the data by sensitivity and compliance type, while highlighting potential security and/or compliance risks.

Personally Identifiable Information (PII)

Data Classification automatically identifies specific words, strings, and patterns in the data. It can recognize PII, credit card numbers, social security numbers, bank account numbers, and more.

To ensure accuracy, Data Classification uses proximity validation to validate its findings. Validation works by looking for one or more predefined keywords near the personal data that was found. For example, Data Classification identifies an Australian Tax File Number (TFN) as a TFN only if it finds a proximity phrase next to it, for example, "TFN" or "Tax File."

Sensitive personal data

Data Classification also automatically identifies special types of sensitive personal information as defined by privacy regulations such as articles 9 and 10 of the General Data Protection Regulation (GDPR). For example, information regarding a person's health, ethnic origin, or sexual orientation. With its NLP abilities, Data Classification can distinguish between "George is Mexican" (indicating sensitive data), versus "George is eating Mexican food".

KEY BENEFITS

Govern all of your NetApp data.

- Map, classify, and categorize your data for visibility and control.
- Perform data hygiene tasks holistically across your hybrid NetApp data estate.

Optimize storage and reduce costs.

- Archive stale data.
- Identify and remove duplicate data.

Accelerate data migration projects.

- Map data for migration.
- Identify sensitive data before moving to the cloud.

Maintain regulatory compliance.

- Map personally identifiable information (PII).
- Comply with privacy regulations, including GDPR, CCPA, PCI, HIPAA.
- Respond quickly to Data Subject Access Requests (DSARs).

Prepare data for GenAI and RAG.

- Find and remove irrelevant or stale data that can distort results.
- Identify and delete duplicate data to enhance training efficiency and prevent the model from assigning undue importance to it.
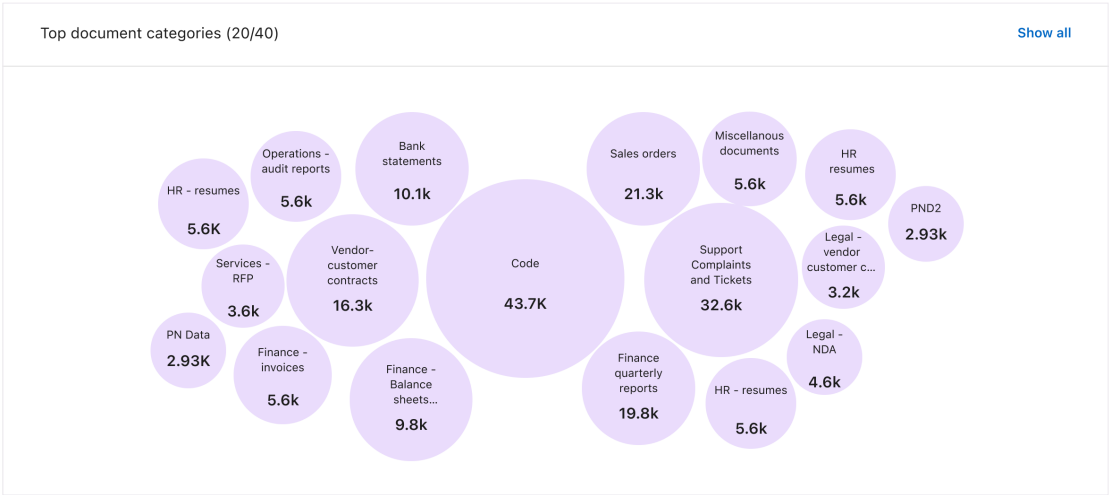- Identify PII and sensitive PII to avoid inadvertent use in training sets and results.

Top document categories (20/40)   Show all

Category	Value
HR - resumes	5.6K
Operations - audit reports	5.6k
Bank statements	10.1k
Sales orders	21.3k
Miscellanous documents	5.6k
HR resumes	5.6k
PND2	2.93k
Services - RFP	3.6k
Vendor-customer contracts	16.3k
Code	43.7K
Support Complaints and Tickets	32.6k
Legal - vendor customer c...	3.2k
PN Data	2.93K
Finance - invoices	5.6k
Finance - Balance sheets...	9.8k
Finance quarterly reports	19.8k
Legal - NDA	4.6k
HR - resumes	5.6k

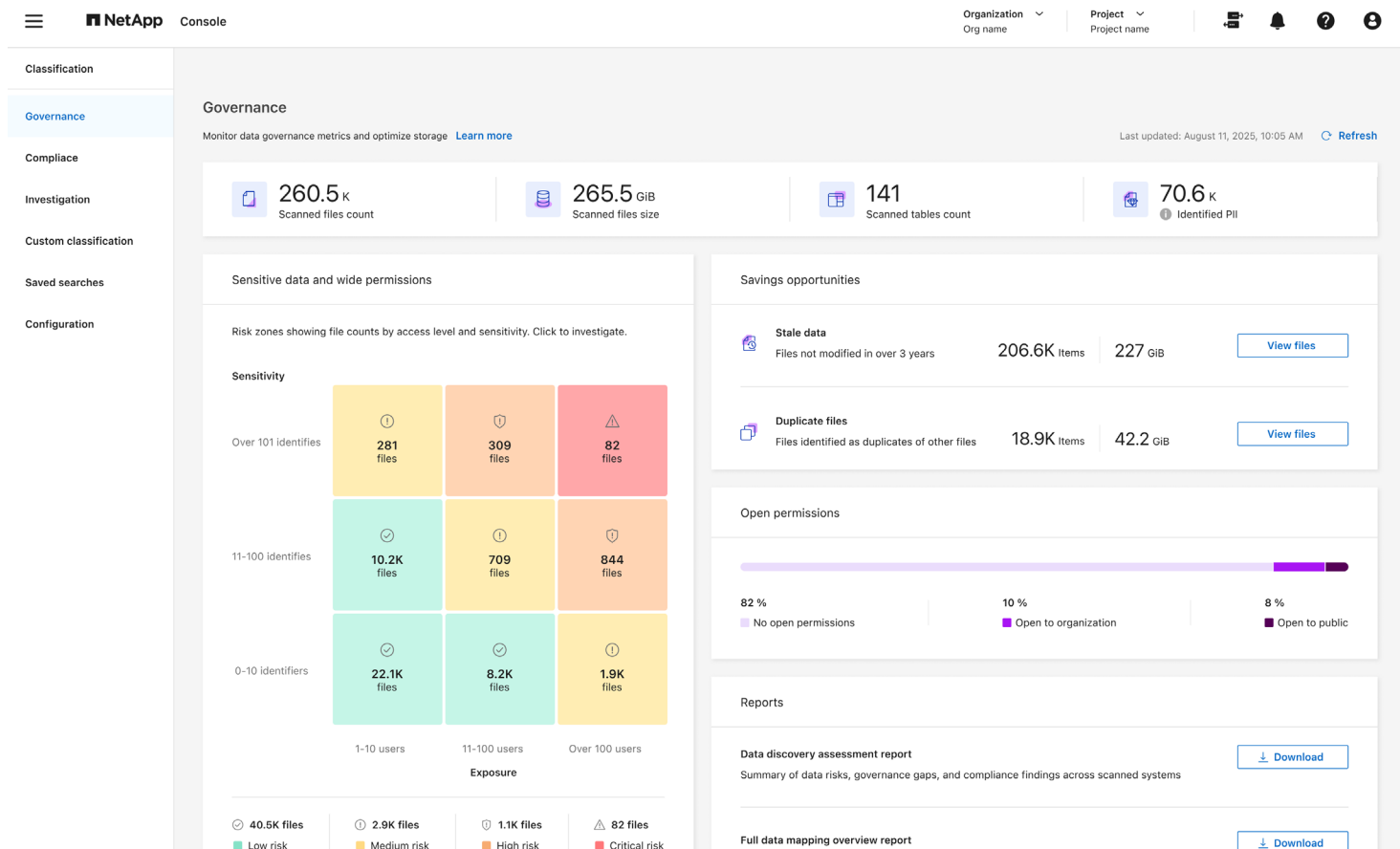Figure 1: AI-based classification for easy data management

Figure 2: NetApp Data Classification dashboard

Get actionable reports

Data Classification offers ready-to-use and custom reports for compliance that reduce manual work, cost, and errors. These include:

- **The Privacy Risk Assessment report:** Provides an overview of your organization's data privacy risk status to support privacy regulations such as GDPR and the California Consumer Privacy Act (CCPA).

- **The Service Data Subject Access Requests (DSAR) report:** Helps comply with GDPR and similar data privacy regulations by finding files that have that person's name or identifier in it.

- **The Payment Card Industry Data Security Standard (PCI DSS) report:** Helps identify credit card information within your data.

- **The Health Insurance Portability and Accountability Act (HIPAA) report:** Helps identify files containing health information.