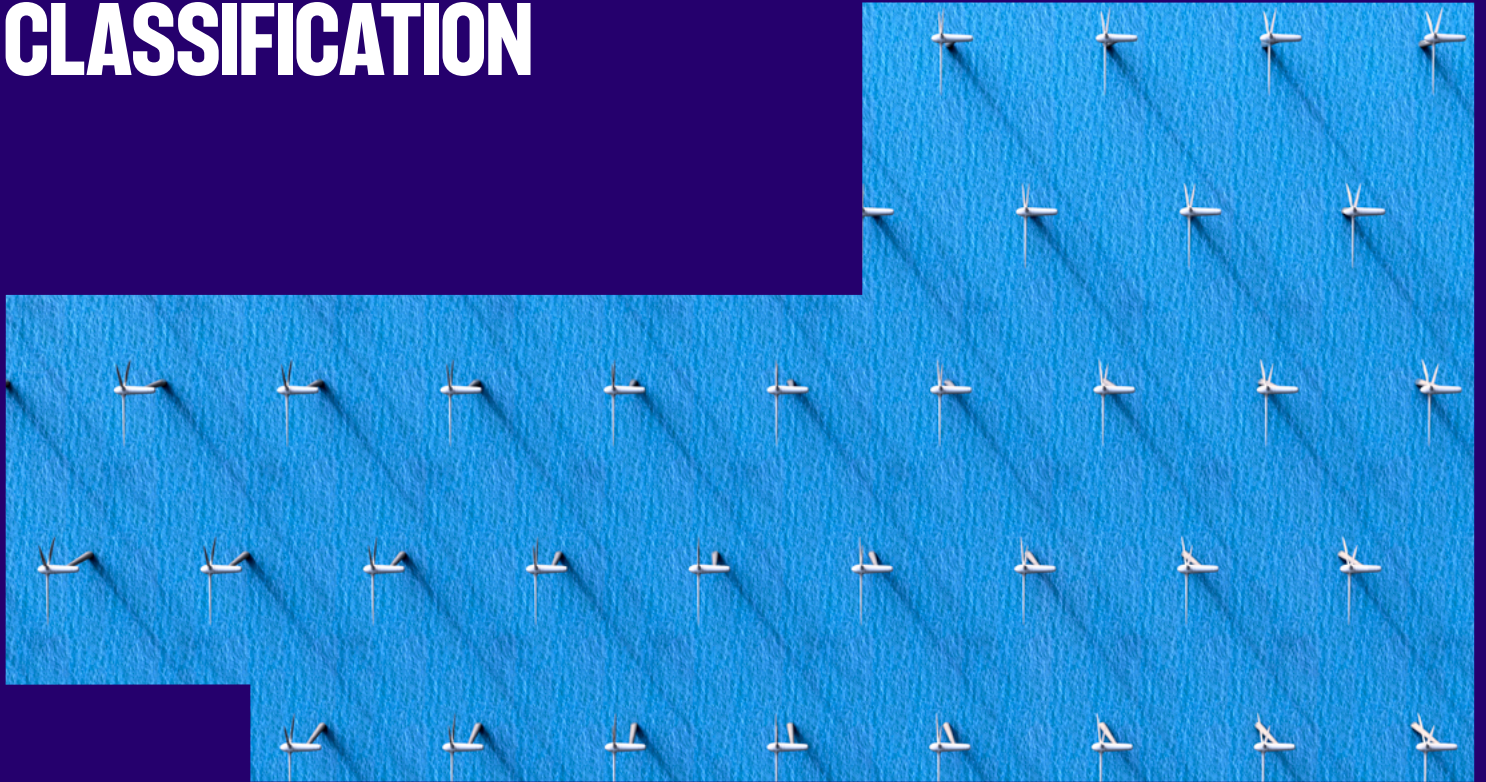**NetApp**

BLUEXP
CLASSIFICATION

Integrated data intelligence for your NetApp data estate. BlueXP classification gives you actionable insights into your data to maintain compliance, optimize storage, accelerate data migration, 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 BlueXP classification – your partner in transforming data chaos into data clarity.

BlueXP classification, a core capability within BlueXP, is a robust data governance service designed to provide comprehensive visibility, insights, and controls for managing data across your NetApp footprint more effectively. It uses advanced AI to simplify data governance, giving you actionable insights to address compliance requirements, optimize storage, accelerate data migration, and prepare data for GenAI and RAG.

With the simplicity of NetApp BlueXP classification, you empower your IT, storage, data privacy, and compliance teams to do their jobs in the most efficient way.

Make intelligent data decisions in real time
BlueXP classification provides automated data governance by mapping data across your NetApp footprint, determining how much data exists, where it is, what type, and the profile of the data. This allows intelligent decisions to be made on data in real time, enabling rapid manual or automated actions to immediately optimize cost and reduce risk.

Data mapping and optimization

With BlueXP classification you can:

- Map all data in your NetApp storage and gain insights through visual dashboards plus detailed data investigation abilities.

- Map duplicate data to identify data redundancy for optimization insights.

- Map stale data to clearly understand what data might not be needed anymore.

- Map and remove data not related to the business for cost savings.

- Map personal, nonpersonal, and sensitive data for a clear understanding of files that might need stricter permissions.

File types

BlueXP classification takes the data that it scans and breaks it down by file type. Reviewing your file types can help you control your sensitive data, because you might find that certain file types are not stored correctly.

Quickly uncover compliance and security risks

Data discovery and classification

BlueXP classification maps and manages sensitive information.

Identifying sensitivities is complex. BlueXP classification uses AI, machine learning (ML), and natural language processing (NLP) technologies to categorize and classify the data by sensitivity and compliance type, constantly highlighting potential cyber and/or compliance exposures.

Sensitive data is especially important in enterprise environments. Often the sensitivity is specific to the organization (and potentially in a specific domain or language). To define sensitivities accurately, modern AI is a must. BlueXP classification does a lot more than traditional pattern matching and identifies many types of sensitivities.

Personal data

BlueXP classification automatically identifies specific words, strings, and patterns in the data.

BlueXP classification can recognize PII, credit card numbers, social security numbers, bank account numbers, and more. You can view the full list of private identifiers in the BlueXP classification documentation.

BlueXP classification uses proximity validation to validate its findings. The validation works by looking for one or more predefined keywords near the personal data that was found.

KEY BENEFITS

Govern all of your NetApp data.

- Map, classify, and categorize your data for visibility and control.

- Perform data hygiene tasks holistically across hybrid NetApp data estate.

Optimize storage and reduce costs.

- Archive stale data.

- Identify and remove nonbusiness and 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, stale, or nonbusiness 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.

For example, BlueXP 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."

Categories

BlueXP classification takes the data that it scanned and divides it into categories. Categories are topics based on AI analysis of the content and metadata of each file.

Categories can help you understand what's happening with your data by showing you the types of information that you have. For example, a category like résumés or employee contracts can include sensitive data. When you investigate the results, you might find that employee contracts are stored in an insecure location. You can then correct that issue.
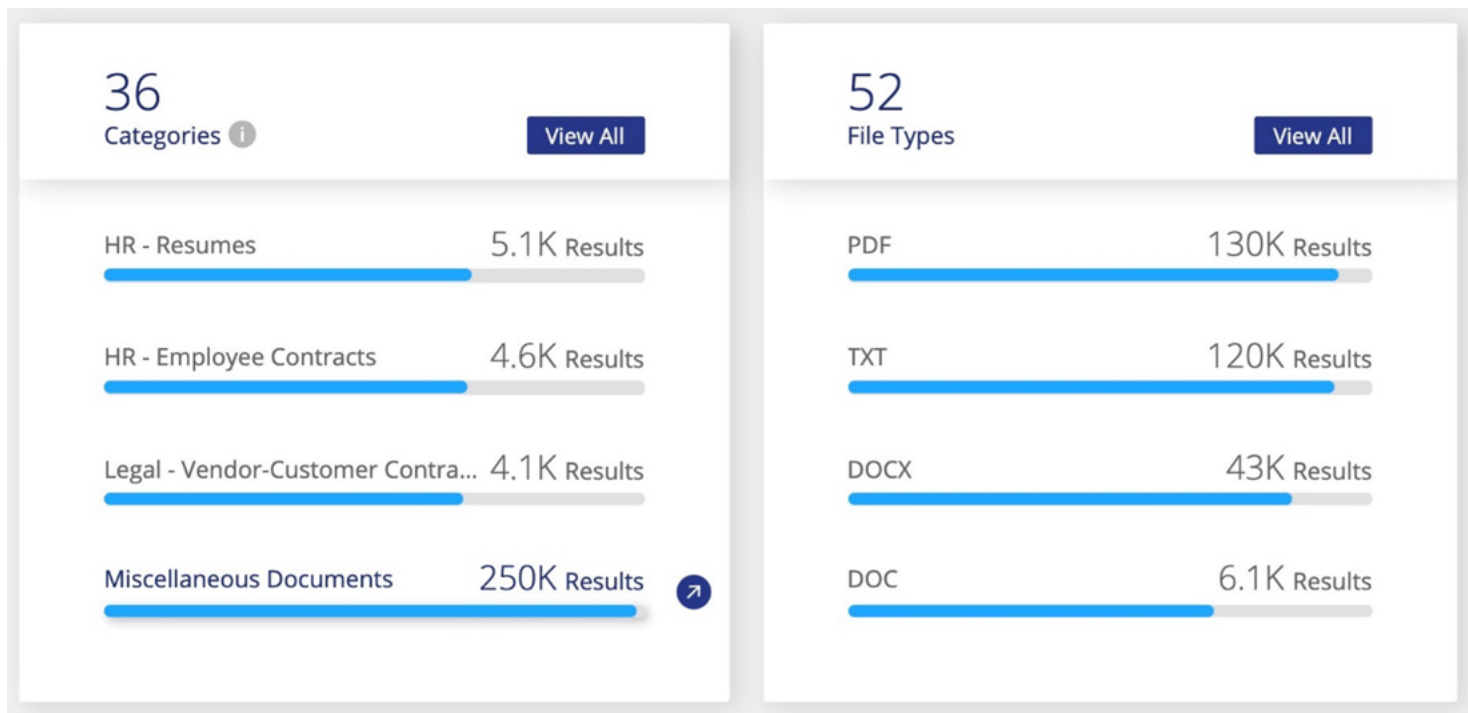
Figure 1: BlueXP classification data mapping

Sensitive personal data

BlueXP classification 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 is sensitive personal data.

BlueXP classification understands the content that it scans in order to extract entities and categorize the content accordingly.

For example, one sensitive GDPR data category is ethnic origin. Because of its NLP abilities, BlueXP classification can distinguish the difference between a sentence that reads "George is Mexican" (indicating sensitive data as specified in article 9 of the GDPR), versus "George is eating Mexican food."

Service Data Subject Access Requests in minutes

With BlueXP classification, you can respond to a DSAR by searching for a subject's full name or known identifier (such as an email address) and then downloading a report.

The downloaded report can help your organization comply with GDPR or similar data privacy regulations. When you perform a data subject search, BlueXP classification finds the files that have that person's name or identifier in it. After the search is complete, you can then download the list of files for a DSAR report. The report aggregates insights from the data and puts them into legal terms that you can send to the auditor or person/subject.
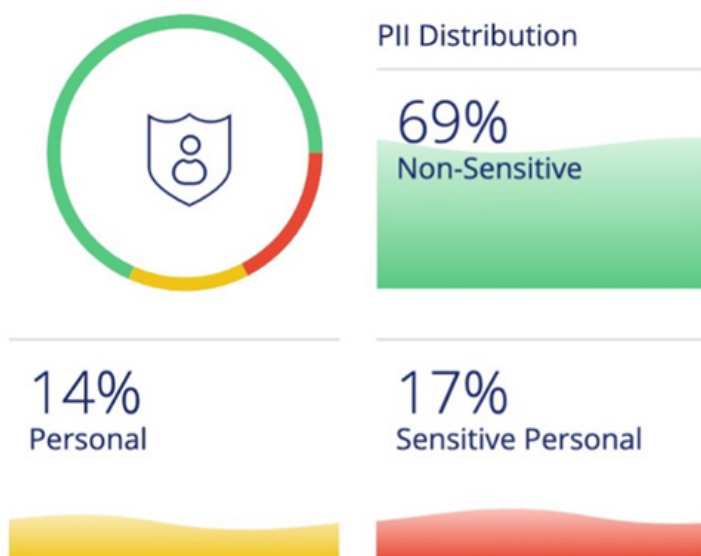

Figure 2: NetApp BlueXP classification sensitive data classification

Personal Results

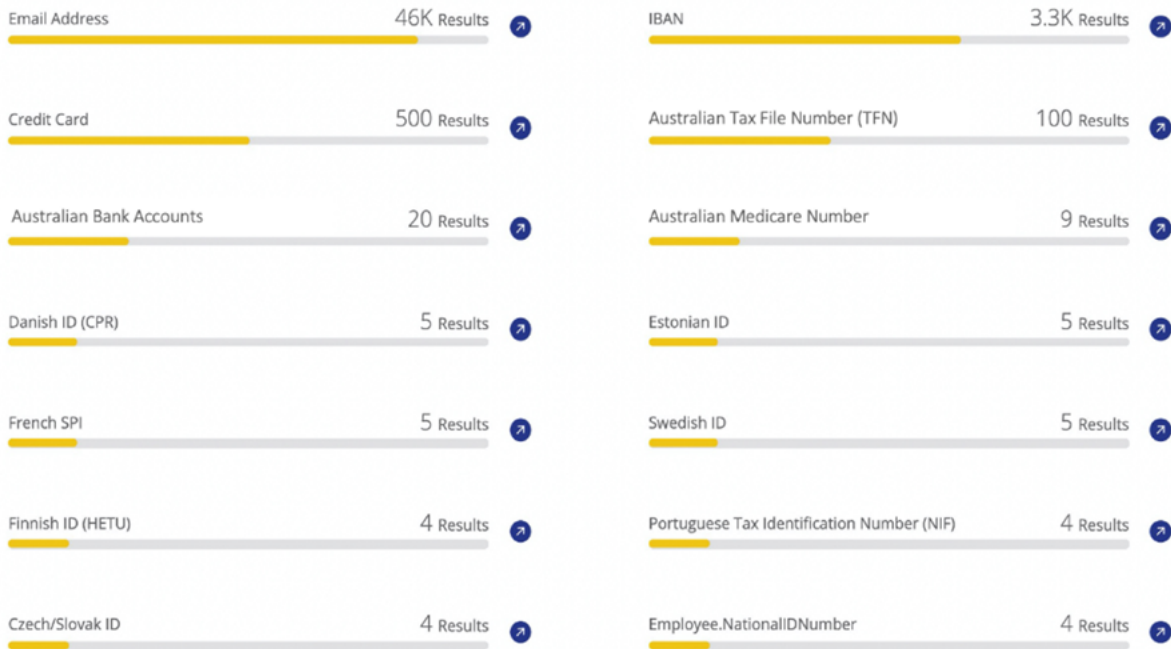27 Types | 47.6K Results found in All working environments

Email Address — 46K Results	IBAN — 3.3K Results
Credit Card — 500 Results	Australian Tax File Number (TFN) — 100 Results
Australian Bank Accounts — 20 Results	Australian Medicare Number — 9 Results
Danish ID (CPR) — 5 Results	Estonian ID — 5 Results
French SPI — 5 Results	Swedish ID — 5 Results
Finnish ID (HETU) — 4 Results	Portuguese Tax Identification Number (NIF) — 4 Results
Czech/Slovak ID — 4 Results	Employee.NationalIDNumber — 4 Results

Figure 3: BlueXP classification personal data mapping

Analyzing your privacy program

To help you slash the manual time, cost, and risk of error to achieve compliance parameters, BlueXP classification provides pre-formatted, legally compliant, and customized reports to address a number of compliance and reporting scenarios, including:

Privacy Risk Assessment Report

The Privacy Risk Assessment Report provides an overview of your organization's privacy risk status, as required by privacy regulations such as GDPR and the California Consumer Privacy Act (CCPA).

PCI DSS Report

The Payment Card Industry Data Security Standard (PCI DSS) Report can help you identify the distribution of credit card information across your files.

HIPAA Report

The Health Insurance Portability and Accountability Act (HIPAA) Report can help you identify files containing health information. It is designed to aid in your organization's requirement to comply with HIPAA data privacy laws.

Top reasons to use NetApp BlueXP classification:

- **Enhance Visibility**
  Get an up-to-date view of what kind of data makes up your NetApp storage footprint –all from one dashboard.

- **Control costs**
  Optimize storage costs by identifying and eliminating duplicate, stale, and nonbusiness data.

- **Migrate efficiently**
  Organize and cleanse data before a data migration project.

- **Stay compliant**
  Automate time-consuming PII identification and reporting for PCI DSS, HIPAA, GDPR, CCPA, and more.

- **Fuel innovation**
  Clean data to create readily usable assets to power GenAI and retrieval augmented generation (RAG) operations.

Figure 4: BlueXP classification sensitive personal data mapping

"*NetApp is positioned in the Maturity/Platform Play quadrant, boasting a comprehensive set of functionalities that extend beyond basic data cataloging to address diverse data governance needs. Its commitment to AI and ML integration reflects a forward-thinking approach, empowering users with AI-powered data analysis and automation. It's a Leader and an Outperformer, with the company's continuous innovation evident in recent advancements, including enhanced AI capabilities.*"

GigaOm Radar for Unstructured Data Management Solutions, March 2024

Contact Us

NetApp