IBM®

Making data smart with IBM Cloud Pak for Data and IBM Spectrum Discover

Highlights

- Leverage file and object data in IBM Watson Knowledge Catalog

- One-click integration leverages millions or billions of files and objects

- Leverage file and object data in IBM Cloud Cloud Pak for Data

- Use Watson Studio to analyze file and object data

- Leverage an integrated portfolio of storage solutions for data and AI

One click integration from Spectrum Discover to IBM Watson Knowledge Catalog
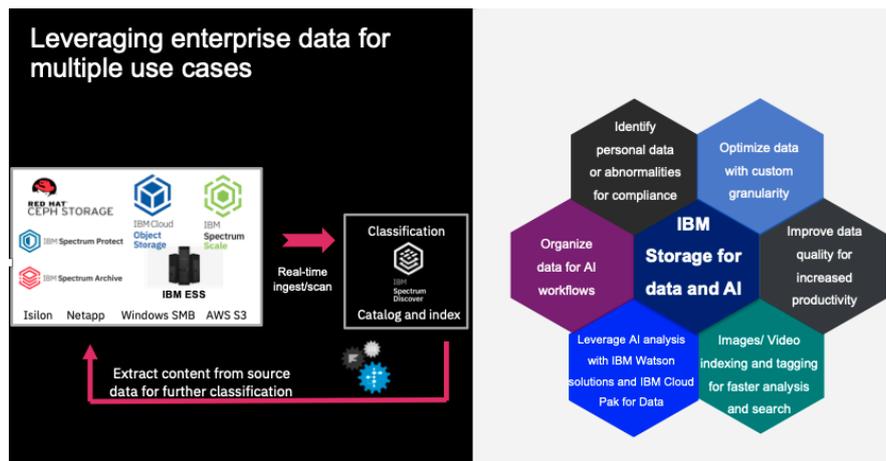
High quality and larger volumes of data are what make artificial intelligence (AI)-driven applications smarter. It has been reported that 81% of business leaders do not understand the data and infrastructure required for AI.

Now, IBM Storage is integrating file and object data to IBM AI solutions to make data smarter.

Enterprises looking to leverage the power of AI are learning that the accuracy and consistency – among other qualities – of the data fed into AI applications directly affect the quality of the results produced by AI. Better data in – better results out.

 But this isn't the only data-related challenge faced by organizations these days. Many companies have been deluged by so much business-generated data that they can't keep track of it all – or they don't know as much about their data as they should in order to make the most profitable use of these very valuable assets. Where does any particular data asset reside? What specifically is contained in each of those thousands or millions of files? How old is it? Who has access to it – and should they? Is it compliant with the latest governmental and international regulations?
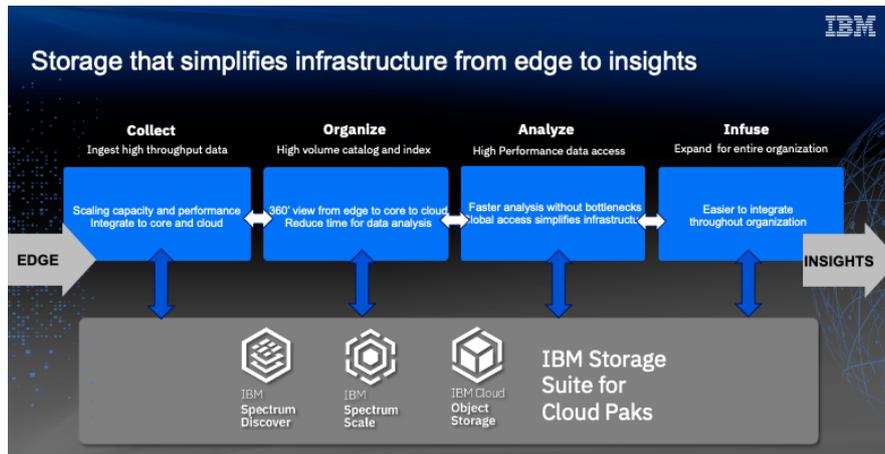
IBM Spectrum Discover

Leverage IBM Spectrum Discover for many use cases

IBM Spectrum Discover simplifies data and AI organization with a high volume data catalog and data management solution that provides a 360º view of data for higher productivity and helps optimize data for faster AI analysis.

Traditional metadata solutions for unstructured data aren't designed to provide the level of detail needed these days about aspects such as storage consumption and data quality or integrate well to other AI solutions. IBM Spectrum Discover serves the important role of classifying and labeling data with custom metadata that not only makes it easier to find and recall data for analysis but can actually increase the value of data by imbuing it with additional semantics and meaning.

IBM Spectrum Discover is an on-premises interactive data management catalog that offers a detailed, real-time view of data. It's non-disruptive to existing storage and applications and provides the ability to create custom indexes and reports for identifying and optimizing enterprise data.

Making storage simple for AI

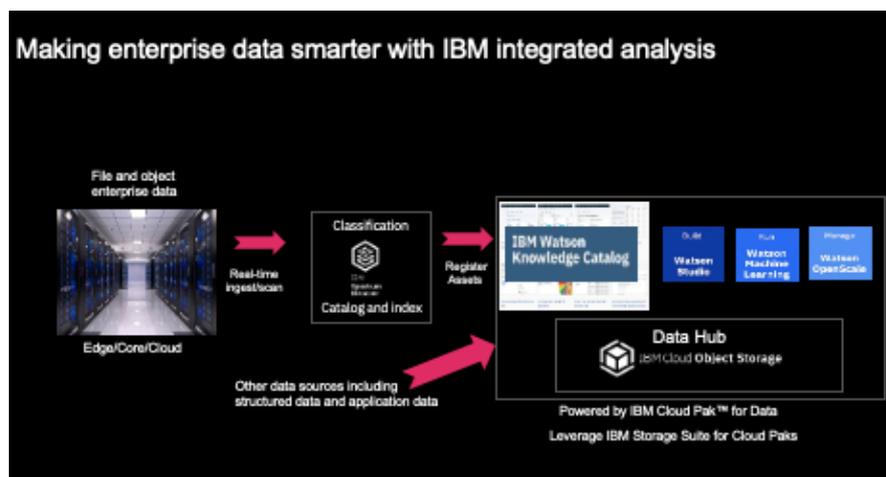*An integrated porfolio of storage solutions for data and AI*

IBM Spectrum Discover is part of a portfolio of products from IBM that simplifies the AI journey by providing an information infrastructure with optimized efficiency to drive faster results that are massively scalable and globally available from edge to insights.

With IBM Storage for data and AI customers can:

- Simplify the journey to AI through a comprehensive data and AI portfolio

  ○ Ingest data with the ability to scale with growth and provide modern data access

  ○ Catalog and index data from multiple sources and with real-time data views

  ○ Create Higher performance analysis with fast access to GPUs/CPUs and evolving data science tools

  ○ Expand the value and efficiencies using a simple and comprehensive infrastructure throughout the organization

- Simplify AI infrastructure for modern solutions from edge to core to cloud

  ○ Flexibility for file-based or POSIX-based application workloads or hundreds of validated S3-based object workloads

- ○ Reference architectures and blueprints for Splunk SmartStore, SAS Grid, NVIDA for AI, IBM Power Systems for AI

- ○ Simplify multicloud and container connectivity with integration to IBM Cloud Pak for Data

  - ▪ One click integration from Spectrum Discover to IBM Watson Knowledge Catalog for enterprise data integration

  - ▪ One product solution using IBM Storage Suite for IBM Cloud Paks with flexibility to choose storage options for file, block or object requirements that include Spectrum Discover

Make data smarter

IBM CLoud Pack for Data with IBM Spectrum Discover

Data scientist and data users can now integrate any information from Spectrum Discover with a single click directly into IBM Watson Knowledge Catalog. This means that data information in the form of metadata along with custom and automated tags from Spectrum Discover can be leveraged by IBM Watson Knowledge Catalog and all the tools that connect to the catalog.

IBM Watson Knowledge Catalog, powered by IBM Cloud Pak for Data, is a data **catalog** that is tightly integrated with an enterprise data governance platform. IBM Watson knowledge Catalog can help your data citizens easily find, prepare, understand and use the data they need. By providing an end-to-end experience rooted in metadata and active policy management, the solution can be leveraged to find success across top use cases like regulatory compliance for the California Consumer Privacy Act (CCPA) and the General Data Protection Regulation (GDPR),

governing data lakes, and self-service consumption of high-quality data.

IBM Watson Knowledge Catalog helps accelerate business value and AI projects as part of a DataOps practice by providing unique capabilities like:

- Real-time data virtualization support

- Automated data discovery and metadata generation

- ML-extracted business glossary from most common regulatory terms

- Dynamic data masking to protect sensitive data

- Automated scanning and risk assessments of unstructured data via Watson Knowledge Catalog InstaScan

IBM is committed to help clients deliver business-ready data to feed AI and analytics projects with IBM Watson Knowledge Catalog for IBM Cloud Pak for Data. The integration of enterprise file and object data information from Spectrum Discover helps to bring more data and more value to AI and analytics projects.

Work with assets in projects

Once information from IBM Spectrum Discover is moved to IBM Watson Knowledge Catalog and the data is moved to the data hub, data users can discover insights by working with the data or analytic assets. The data hub is an S3 enabled storage system and IBM Cloud Object Storage can become the storage system no matter where the deployment is located. You can move the data from Spectrum Discover along with other metadata to a project. You can also use a project as a staging area to curate data assets or create analytic assets before publishing them back to IBM Watson Knowledge Catalog. Projects usually contain a select subset of catalog collaborators.

You can add assets from a catalog to a project to work with them and add other tools to analyze data or create artificial intelligence models by using the Watson Studio application. Watson Studio and Watson Knowledge Catalog are fully integrated and are part of IBM Cloud Pak for Data.

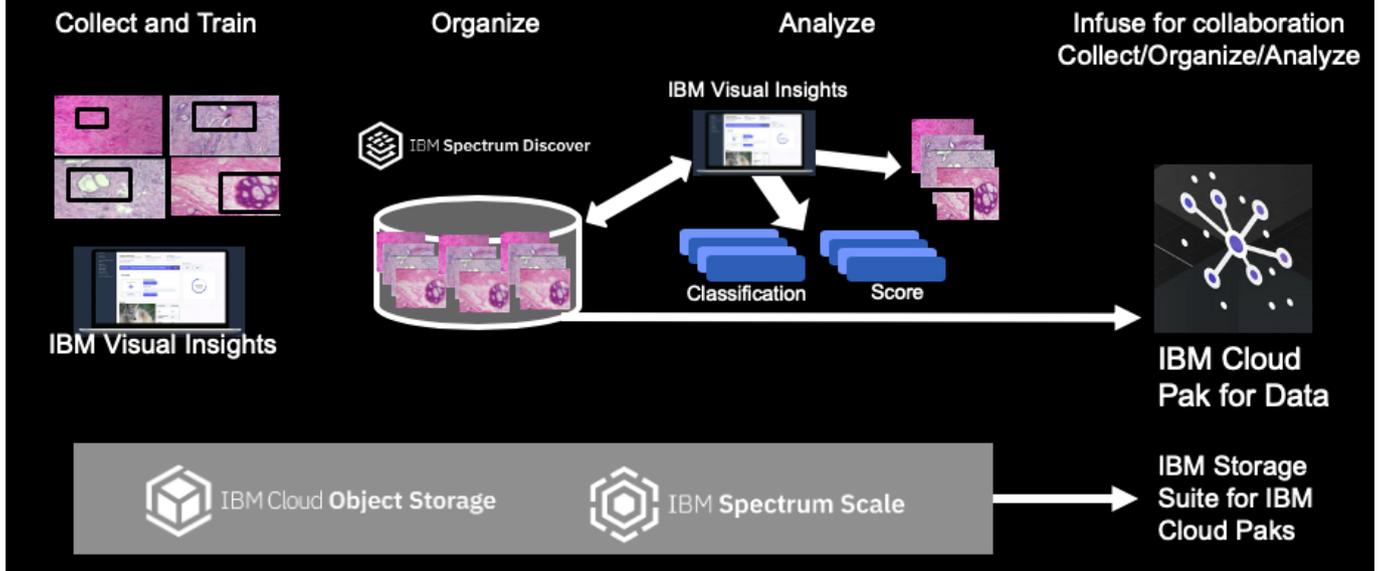## An integrated example leveraging IBM Storage for data and AI

In this example we first collect and train multiple data points from cancer images or even virus images like COVID-19.

This data is collected in IBM Spectrum Scale or IBM Cloud Object Storage. An inference model is developed using IBM Visual Insights that can identify defects or cancer or viruses based on an existing set of data images. Then data is sent to Spectrum Discover for the organize phase of the AI process. Spectrum Discover invokes a 'DeepInspect' policy (a capability built into Spectrum Discover) against the new data, activating the inference.   We now move to the first analysis phase where we will compare new images with our model to determine the probability of a match with our base data set. IBM Visual Insights is triggered to analyze new images and index them back in to Spectrum Discover. Spectrum Discover can then display results and can register the results to IBM Watson Knowledge Catalog for data governance and data collaboration and further analysis. Information and data from IBM Storage for data and AI can also be used to perform additional AI analysis based on other correlations such as geographical data, weather data, physical data or other data that can be used for additional analysis using tools in IBM Cloud Pak for Data such as IBM Watson Studio or IBM Watson for AI and ML. This makes the data even smarter and more powerful and shows the advantage of an integrated AI solution.

IBM solutions for data and AI provide an integrated approach across the enterprise with the ability to:

- provide event-based indexing and tagging of unstructured data across the enterprise

- automatically inspect and classify over 1000 unstructured data types, including genomics and imaging specific file formats

- automatically register assets with Watson Knowledge Catalog based on Spectrum Discover search and filter criteria and leverage assets in Cloud Pak for Data

- enforce data governance policies in Watson Knowledge Catalog in Cloud Pak for Data based on insights from Spectrum Discover and leverage assets in Cloud Pak for Data

In integegrated approach to AI

IBM®

Why IBM?

As an industry-leading provider of data storage products, IBM is investing in data management solutions that improve storage economics, data quality, data governance and data identification for large-scale analytics and AI. IBM Spectrum Discover is a key aspect of the overall IBM Storage for data and AI advantage, and provides powerful metadata management that brings visibility and classification to improve storage optimization and increase data science.

For more information

To learn more about any of the following products please contact your IBM representative or IBM Business Partner or visit

for Spectrum Discover:

https://ibm.com/marketplace/spectrum-discover

for IBM Watson Knowledge Catalog:

https://www.ibm.com/cloud/watson-knowledge-catalog

for IBM Cloud Pak for Data:

https://www.ibm.com/products/cloud-pak-for-data

for IBM Storage Suite for IBM Cloud Paks:

https://www.ibm.com/marketplace/storage-suite-for-ibm-cloud-paks

for IBM Visual Insights:

https://www.ibm.com/products/ibm-visual-insights

IBM.