# IBM Spectrum® Discover

Enable users to be smarter with the data you have

Highlights

- Gain new insights with data that is difficult to find or may be located in hard to reach locations
- Locate PB's of data on IBM® and non-IBM storage in seconds to speed AI workflows on-prem and in the public cloud
- One-click integration to IBM Cloud Pak® for Data and the IBM data fabric to bring more insights to data users
- Automatically classify data for faster AI and business analysis
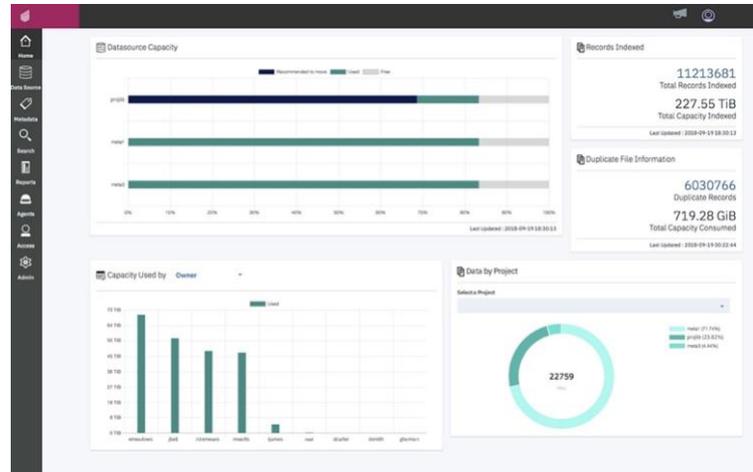
With digital transformation comes rapid growth of unstructured data. It is no surprise that data users are struggling to keep pace with rapid growth of unstructured data. These users often find themselves taking too much time to find the data they need or end up missing data that would have been critical to their analysis. Other users such as administrators are blind to important information about their data because of the massive amounts they are managing. They are unable to optimize their storage and many times just end up throwing more storage at the problem. They may miss vulnerabilities in their data such as improper data governance or data that is located on the wrong type of storage or wrong location.

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 simplifies data and AI organization with a high volume inteigent index engine 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 inteligent data service 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.

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

With the IBM global data patform and AI customers can:

- Simplify the journey to AI through a comprehensive global data platform portfolio
- Ingest data with the ability to scale with growth and provide modern data access
- Inteligently index data from IBM and non-IBM data sources and with real-time data views
- Create Higher performance analysis with fast access to GPUs/CPUs and evolving data science tools
- 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

How to leverage AI with IBM Spectrum Discover
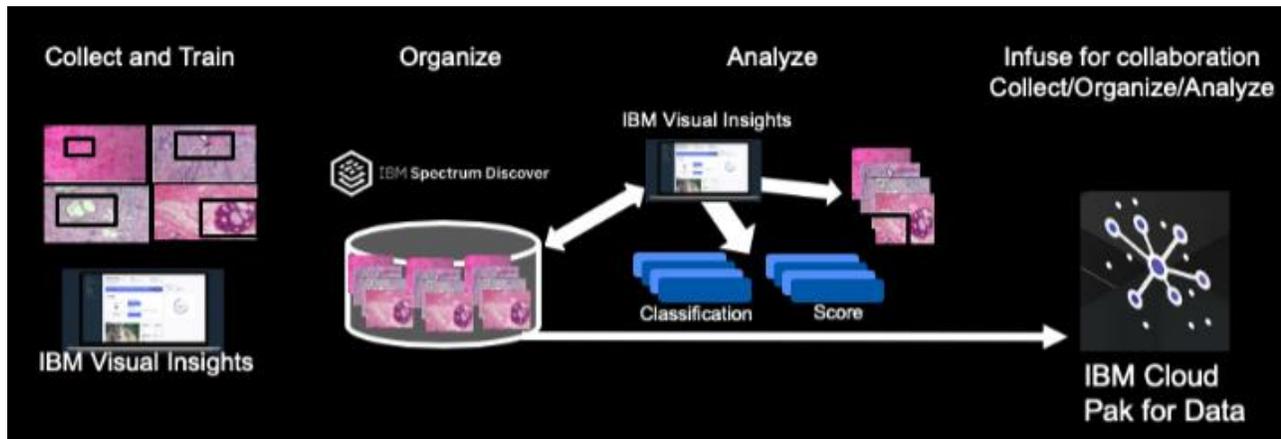

IBM Spectrum Discover example

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 ESS and IBM Cloud Object Storage 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 global data platform provides 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

The Global Data Platform
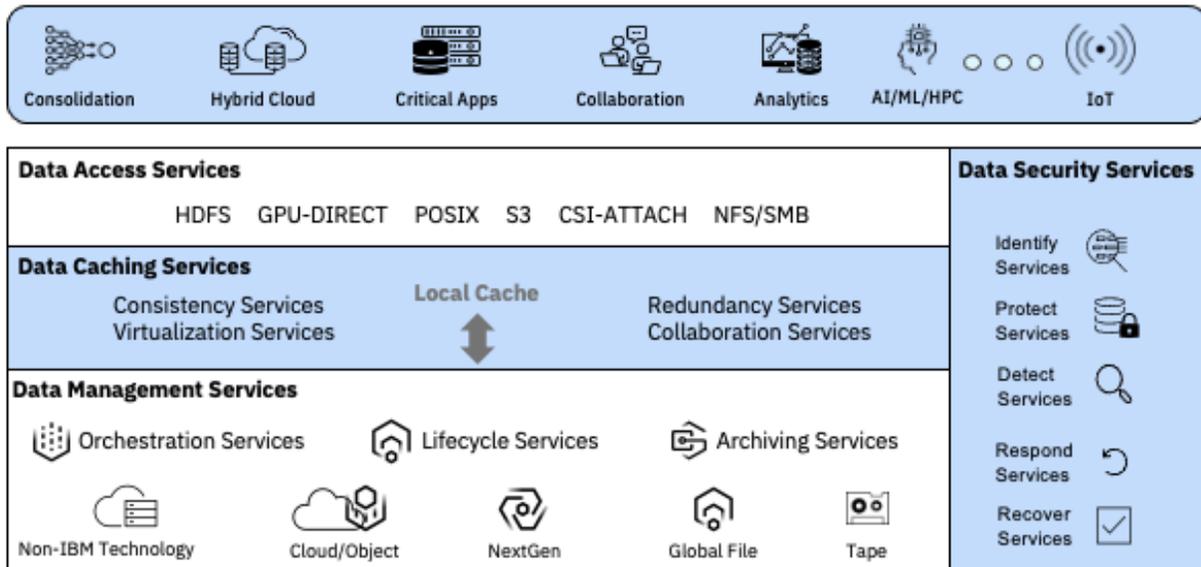
The details of the global data platform are actually not complicted but they contain over 30 years of investment and research at IBM. With over 4000 customers in almost every industry across the globe, the global data platform is a proven platform that has been used in the fastest supercomputers to banking systems that have never lost access to data in years. It has been used by many of the top corporations in the world with PBs of data to service millions of users and it also can be deployed in small businesses that only needs TBs of high perfromance data.

Contact your IBM Sales representative or value added business partner for details about pricing and your specific configuration and solution requirements. IBM Spectrum Discover is targeted at new easy to order, easy to install, easy to upgrade, easy to use, software that is available to deploy from edge to core to cloud.

Visit the IBM Spectrum Discover page for more information

Why IBM
IBM Storage for file and object data provides a way to solve diffult problems with more data and faster access to that data.

Each storage system from IBM offers a full-featured set of data services which are powered by the Global Data Platform. The core data services include Data Access Services, Data Caching Services, Data Management Services and Data Security Services.

Each core data service offers their own set of Custom data services that create an easy to manage and powerful set of capabilities that differentiate IBM's cloud scale products.

Find out more about our entire product portfolio including IBM Spectum Scale, IBM ESS, IBM Cloud Object Storage and IBM Spectrum Discover products.

Next steps
[IBM Storage Cloud Scale Solutions](#)