Put the power of AI and data to work for your business

IBM Watson Studio: Accelerating value for AI in the enterprise

Highlights

- Data Science teams of any skill set can use the top open source and advanced data science tools across the analytics lifecycle

- No need to move your data—build and deploy where you want in a multi-cloud environment, whether your data exists behind the firewall or in the cloud

- Data science can be put into production faster with ease of deployment, model management and auto-retraining to increase your team's productivity

Whether your organization is just getting started with AI or looking to expand early projects into production environments, IBM Watson® Studio can provide a platform for success. Watson Studio is an integrated environment designed to make it easy for builders who work with data to develop, train and manage models and deploy AI-powered applications.

With Watson Studio, data scientists, developers and analysts get preconfigured tools drawn from across open source and IBM technologies, along with powerful data preparation, cataloging and governance capabilities. This seamless working environment leads to enhanced collaboration and big productivity gains that save both time and money in building machine learning models and AI applications.

Deploy where it makes sense for you with our flexible multi-cloud architecture

Watson Studio is built for the enterprise, with an architecture that can span multiple environments. It can run on IBM's public cloud, behind your firewall in your data center or on your workstation. It can integrate with a wide range of data stores—on premises or on the cloud, both IBM and non-IBM—using over 30 built-in connectors. This means, if you have access to credentials, you can set up a new connection in minutes. This hybrid, multi-cloud environment allows you to take advantage of differing services, performance levels, security and redundancy. For example, you can:

Protect sensitive data—train on-prem, deploy on public cloud: When security or data gravity are an issue, you can train models using data that sits behind your firewall, then deploy on a public cloud where the model can be invoked by cloud-native applications. To prevent latency for scoring calls, deploy models close to the consuming app, near the edge of the network or outside of the firewall.

Harness public data—train on public cloud, deploy on-prem: Train models on the public cloud to harness high-volume, non-proprietary publicly available data. Pull down the latest version of the model and integrate it with on-prem proprietary applications like systems of record.

Work with a wide variety of open source data science tools

Watson Studio integrates tools like Apache Spark, Jupyter Notebooks and RStudio®. Jupyter Notebooks enable data scientists to create and collaborate on Python, R and Scala projects while RStudio enables fast development of R scripts. Using this shared environment makes it easier for data scientists to collaborate, share results and create reproducible research.

With Watson Studio, open source tools can run in highly configurable environments that allow you to assign compute resources based on team needs. For example, they can run in an elastic single node cloud environment that allows you to scale with a single click. Or you can scale with Apache Spark powered by services from IBM Cloud or remote clusters.

Experiment-centric deep learning and advanced modeling with neural nets

Watson Studio includes Neural Network Modeler, an intuitive drag-and-drop interface for designing neural architectures using the most popular deep learning frameworks: TensorFlow, Caffe, PyTorch and Keras. Users can quickly capture network designs and then export them for experimental optimization.

Experiment Assistant is a suite of tools that guide and manage the experimental process, simplifying the hundreds to thousands of training runs it takes to find the right combination of network layer configurations and hyperparameters. Each training run is automatically started, monitored and stopped upon completion. Training history and assets are tracked, then automatically transferred to object storage.

Distributed training capabilities allow training of a single neural network across dozens of GPUs to reduce compute time from days to hours or minutes, speeding time to market with new AI solutions. Support for distributed training is provided using native TensorFlow, IBM Distributed Deep Leaning framework and Uber's Horovod.

Enrich apps with integrated Watson AI services

Watson Studio is designed for tight integration with IBM Watson services, enabling users to create a service instance within Studio and associate projects with services to enable collaboration and use within notebooks and other tools. One such service is Watson Visual Recognition. Use this machine learning service to quickly and accurately tag, classify and train visual content. Visual Recognition analyzes images for scenes, objects, faces, colors, food, text, explicit content and other attributes that you select and train to provide insights into visual content.

Give business analysts a drag-and-drop solution for insights without coding

IBM® SPSS® Modeler, incorporated into Watson Studio, offers business analysts and data scientists an easy-to-use, interactive way to develop predictive models without the need for programming. It provides automated modeling with out-of-the-box, industry-leading algorithms as well as a range of advanced analytics, including text analytics, geospatial analysis and optimization.

Built-in data preparation and profiling with Data Refinery

Before the real work of data science can begin, data must typically be cleansed, refined and enriched. In fact, it's estimated that up to 80 percent of a data scientist's time is spent on this kind of data preparation work, leaving only 20 percent for real analysis. Data Refinery, included with Watson Studio, can help data scientists and data analysts push that ratio significantly toward greater productivity.

It provides a powerful set of self-service capabilities that can be used to explore datasets, iteratively cleanse and refine them, and visualize the results to make sure it is what was intended. The solution also automatically tracks and documents every step in the refinement process to provide end-to-end data flow and makes it easy to save and share the output with other knowledge workers across the organization.
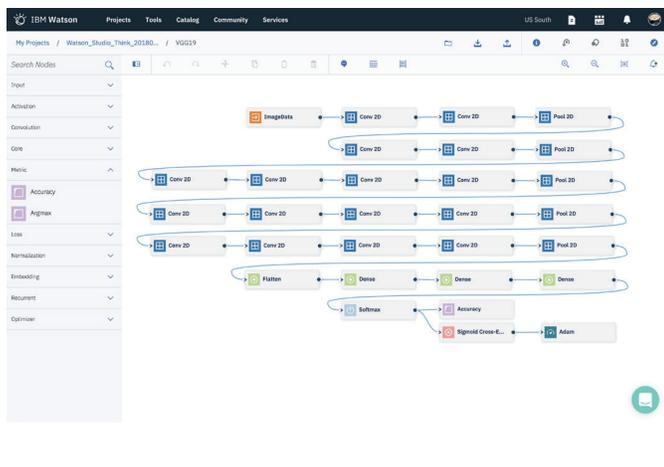
Figure 1. Drag-and-drop interface of Neural Network Modeler

Integrated data labeling services

Machine learning requires labels, or annotations, that describe an object (text, image, audio, video) to the algorithm. Most training data is not annotated and, hence, requires extensive human resources to label it. To simplify this task, IBM has integrated services from third-party providers like Defined Crowd and Figure Eight within Watson Studio. These provide "human-in-the-loop" data labeling that transforms raw, unstructured data into valuable training data.

Act on insights in real time

Streams Designer is another integrated tool within Watson Studio, which provides a simple approach to defining the intent of an app through visual composition. Apps can then be deployed to the IBM Streaming Analytics service to analyze data in motion. This enables real-time scoring of models and analysis using Python code created in other parts of Watson Studio, allowing users to act on new insights in real time.

Visualize insights with dashboards

Integrated analytic dashboards can enable analysts to discover insights from data and turn the analytical results into user-friendly visualizations for sharing with a broad audience of business users.
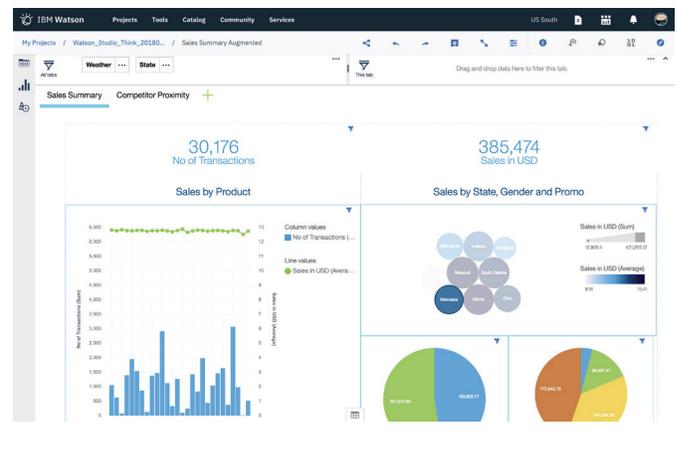

Figure 2. Visualize analytics with built-in dynamic dashboarding

Dashboards can be created from within a project using drag-and-drop tools accessing available data assets. The system provides automatic recommendations for effective visualizations based on the type of data selected, while built-in templates and styles make it fast and easy to format reports as needed.

Deliver self-service access to data and other knowledge assets

Integrated with Watson Studio is the Watson Knowledge Catalog, which includes tools to automate and simplify data discovery, facilitate curation and provide active governance. Intelligent AI-powered search capabilities help users find the structured and unstructured data, notebooks and other knowledge assets they need, while metadata such as tags, comments, and quality metrics help them decide whether a data set will be useful to them and how best to extract value from it. Lineage of assets, including models, is automatically captured to give consumers the ability to understand where an asset came from, where it was used and what the inputs were.

Integrated active governance capabilities give users confidence that they are permitted to use a given data set while automatically masking sensitive data so they are not able to see it. This helps ensure that the assets in the catalog are used responsibly by others in the organization.

Accelerate analytics development with IBM Analytics Engine

IBM Analytics Engine is another key component of IBM Watson Studio. It is the next generation of IBM's Apache Spark and Apache Hadoop cloud-based service that enables data scientists to rapidly provision, manage, run and retire Apache Hadoop and Apache Spark clusters. It increases flexibility by keeping compute and storage infrastructure separate, so each can scale independently to prevent loss of data if a compute cluster fails. The data is stored in IBM's Cloud Object Storage service, and the Hadoop and Spark clusters connect to the object storage repository when they need to access it. It simplifies the analytics infrastructure and streamlines workflow.

Choose the right deployment and pricing for your organization

Whether you are an individual user looking to learn about data science, an AI professional working alone or with a small team, or an enterprise with department-level AI programs, there is an IBM Watson Studio edition to meet your needs, in the cloud or behind your firewall.

For more information

To learn more about how IBM Watson Studio can help you develop new business solutions based on AI and deep learning, contact your IBM Business Partner or visit our website:
ibm.com/cloud/watson-studio

11021911USEN-00