Technical Validation

Hybrid Multi-cloud Artificial Intelligence (AI): IBM Watson Studio and Watson Machine Learning

Accelerating AI and Machine Learning (ML) Initiatives at Scale, with Speed and Simplicity

By Jack Poller, Senior Analyst; and Tony Palmer, Senior Validation Analyst
February 2019

Contents

ESG Technical Validations

The goal of ESG Technical Validations is to educate IT professionals about information technology solutions for companies of all types and sizes. ESG Technical Validations are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objectives are to explore some of the more valuable features and functions of IT solutions, show how they can be used to solve real customer problems, and identify any areas needing improvement. The ESG Validation Team's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments.
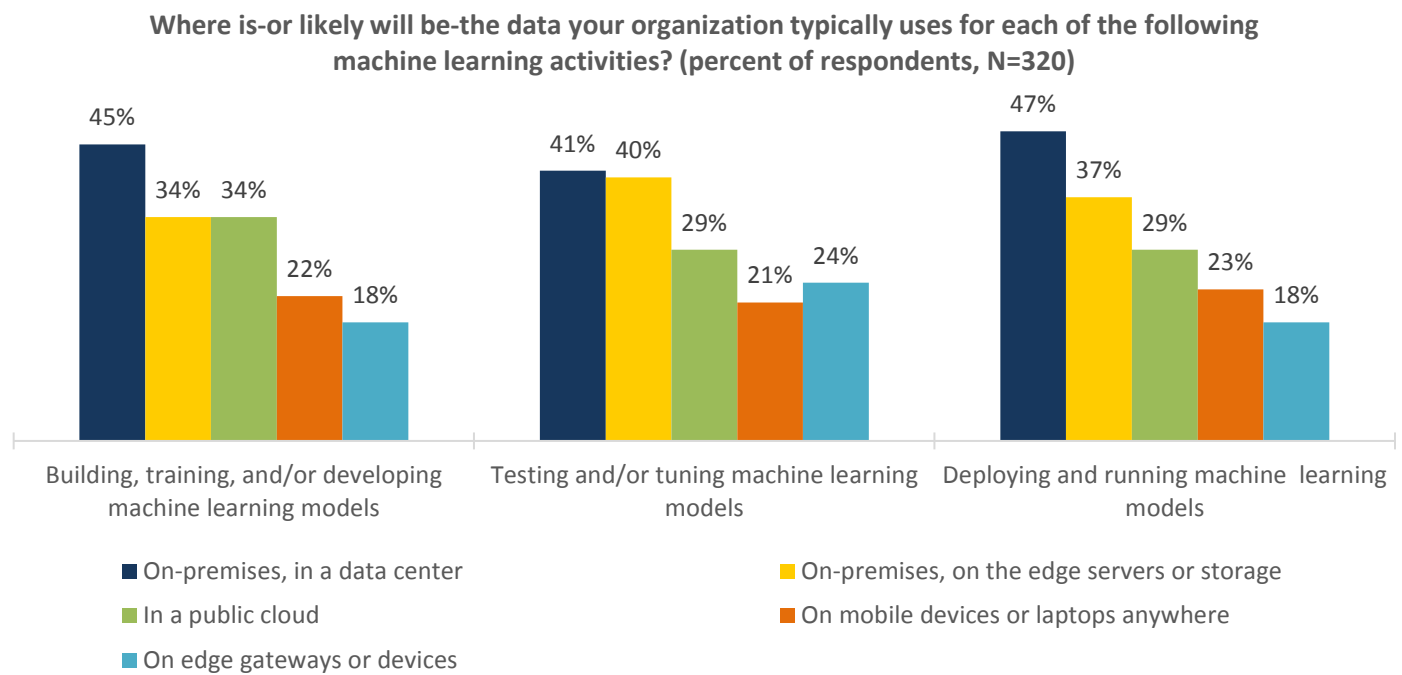
Introduction

ESG recently completed testing of IBM Watson Studio and Watson Machine Learning, which are designed to enable organizations to accelerate the value they can extract from AI more easily. Testing examined how IBM Watson Studio and Watson Machine Learning collect data, organize an analytics foundation, and analyze insights at scale—with a focus on the ease of operationalizing AI and data science to improve trust, simplify compliance, and speed monetization.

Background

Thanks to increased computing power, new algorithms, and massively parallel processing with graphics processors (GPUs), artificial intelligence (AI) and machine learning (ML) have moved from aspirational to key components of digital transformation and business modernization initiatives. According to ESG research, 59% of respondents expected their spending on AI/ML to increase in 2019, while 31% of organizations indicated that leveraging AI/ML in their IT products and services was one of the areas of data center modernization in which they expected to make the most significant investments in the next 12-18 months.[1]

Organizations looking to unlock the power of AI/ML face significant challenges, from the lack of experienced or trained staff to the lack of better IT infrastructure required to support ML. Over 64% of respondents reported using three or more tools for ML, which greatly complicates the iterative, cyclical learning process that drives ML. This leads to issues with time to business value, with more than half of respondents expecting ML to take more than a year to return significant business value, and nearly one in five expecting it to take more than two years. Most importantly, as shown in Figure 1, organizations have not consolidated their AI/ML IT infrastructure stacks. Organizations predicted that the three major phases of AI/ML initiatives—building/training, tuning/testing, and deploying/running in production—will occur in the data center, on local devices, on mobile devices, on edge devices, and in the public cloud.[2]

Figure 1. Location of Data for Machine Learning Activities

Where is-or likely will be-the data your organization typically uses for each of the following machine learning activities? (percent of respondents, N=320)

Legend:
- On-premises, in a data center
- On-premises, on the edge servers or storage
- In a public cloud
- On mobile devices or laptops anywhere
- On edge gateways or devices

Source: Enterprise Strategy Group

---

[1] Source: ESG Research, *2019 Technology Spending Intentions Survey*.
[2] Source: ESG Survey, *Machine Learning and Artificial Intelligence Trends*, June 2017.

While machine learning can be used across an organization, the current use is still focused within IT itself, and IT is leading the way both in terms of using ML and defining the ML initiative strategy. What is needed — as companies expand ML beyond IT and into the lines of business and executive teams—is a solution that can span on-premises, hybrid, and multi-cloud environments.

IBM Watson Studio and Watson Machine Learning (WML)

IBM describes a prescriptive approach—called the AI ladder—to accelerating organizations' journeys to AI. The AI ladder describes four layers: simple, accessible data collection; data organization to support a trusted foundation for analytics; scalable analysis with AI everywhere; and infusion, or transparent operationalization of AI with trusted AI-driven business processes. Pre-built app services and built-in expertise help accelerate time to value.

As part of IBM's prescriptive approach to AI, IBM Watson Studio and Watson Machine Learning are designed to bring hidden intelligence to the surface to help organizations transform business operations, on any cloud.

Figure 2. The IBM AI Portfolio
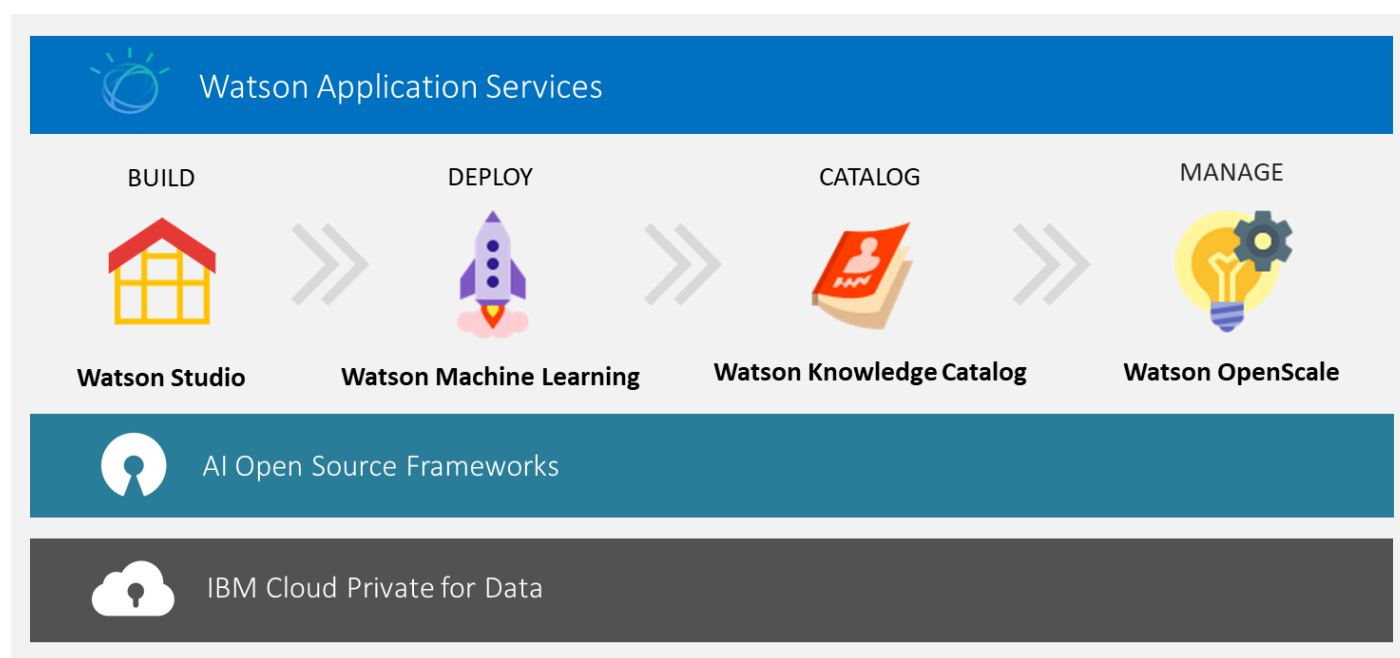
Source: Enterprise Strategy Group

IBM Watson Studio provides tools for data scientists, application developers, and subject matter experts to collaboratively and easily work with data to build and train models at scale. It is designed to give the flexibility to build models where the data resides and deploy applications anywhere in a hybrid environment, to operationalize data science faster.

IBM Watson Machine Learning (WML) is an enterprise machine learning offering focused on the deploy phase of the data science lifecycle. It is designed to enable a business to use trusted data to put machine learning and deep learning models into production. It allows business to leverage an automated, collaborative workflow to deploy AI-infused business applications easily at scale with more confidence. IBM WML Accelerator—formerly known as PowerAI Enterprise—is designed for organizations moving from exploration and single-node development to scale-out production environments for machine learning and deep learning technologies. WML Accelerator's goals are to shorten time to accuracy and improve the efficient use of resources across multiple data scientists, enhancing their productivity. WML Accelerator is available as an add-on to the WML base offering.

IBM Cloud Private for Data (ICPD) is an open, cloud-native information architecture for AI. Designed as an integrated, fully governed team platform, organizations can keep data secure at its source and add preferred data and analytics microservices as needed. ICPD can also be augmented with a Data Science premium add-on, a consumption-based value-added offering that includes IBM SPSS Modeler and Decision Optimization to further accelerate time to results and combine predictive and optimization models.

ESG Technical Validation

ESG performed evaluation and testing of IBM Watson Studio, Watson Machine Learning, and IBM Cloud Private for Data. Testing was designed to demonstrate how IBM's AI portfolio can help organizations accelerate their data science journeys, providing tools to simplify data collection, organization, and analysis, with a goal of operationalizing AI.
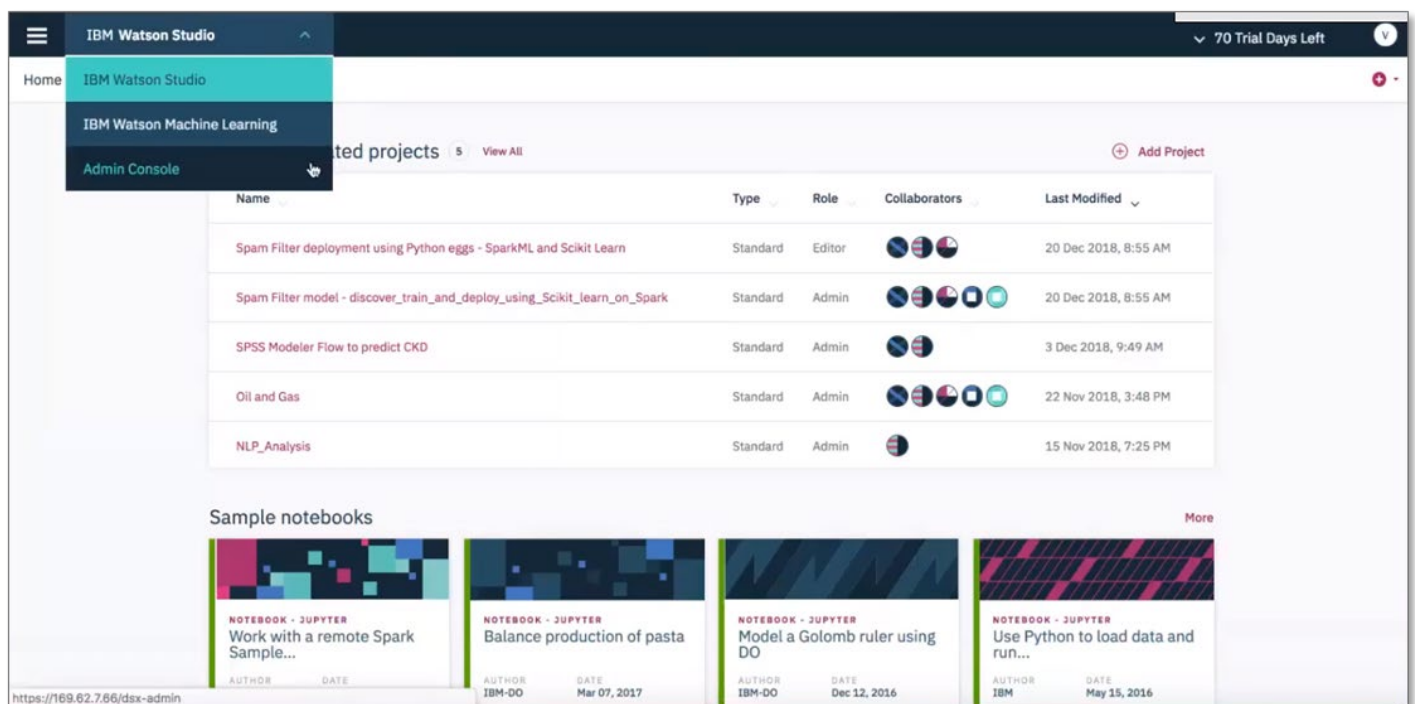
IBM Watson Studio

To support the needs of people who perform data exploration, data preparation, and modeling, IBM offers Watson Studio. Watson Studio provides visual and open source tools that help to accelerate the time to complete and iterate during the build phase.

ESG looked at how IBM Watson Studio can be used to accelerate the build phase using best-of-breed tools and frameworks from IBM and the open source community; scalable, flexible infrastructure; and the ability to train and deploy wherever the data resides, whether on-premises, in the cloud—AWS, Azure, or IBM Cloud—or as a fully managed service.

ESG Testing

The build phase in the data science lifecycle includes data exploration, preparation, and model development and testing. First, we logged in to the IBM Watson Studio dashboard.

Figure 3. The IBM Watson Studio Dashboard

Source: Enterprise Strategy Group

The dashboard shows a list of existing projects and is where a new project would be created. Selecting an existing project took us to the IBM Watson Studio project dashboard. The project dashboard allows users to connect directly via a terminal, or download the entire project as an artifact, to run the project wherever the data lives.

Figure 4. IBM Watson Studio Project Dashboard
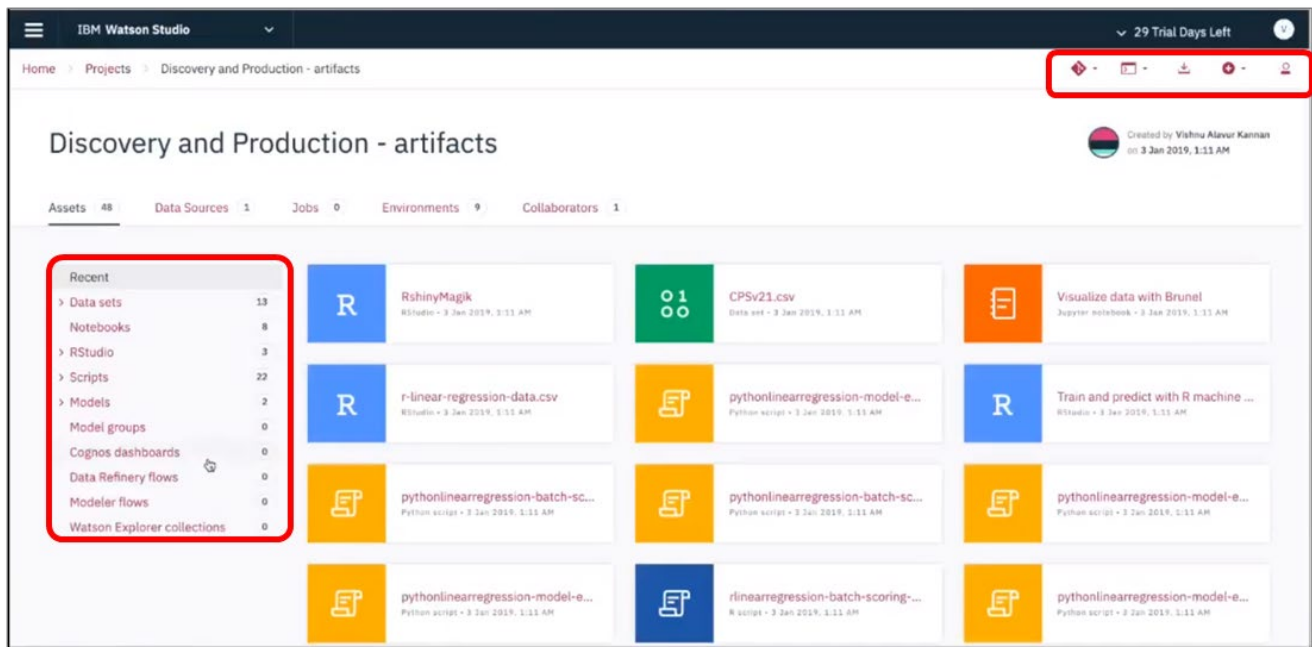
Source: Enterprise Strategy Group

A high-level view of projects is available in the project dashboard, with a single view of all artifacts, including data sets, notebooks, RStudio projects, scripts, models, etc. The project dashboard provides a view of how the entire ecosystem comes together, from discovery to production.

Figure 5. Adding a Data Source
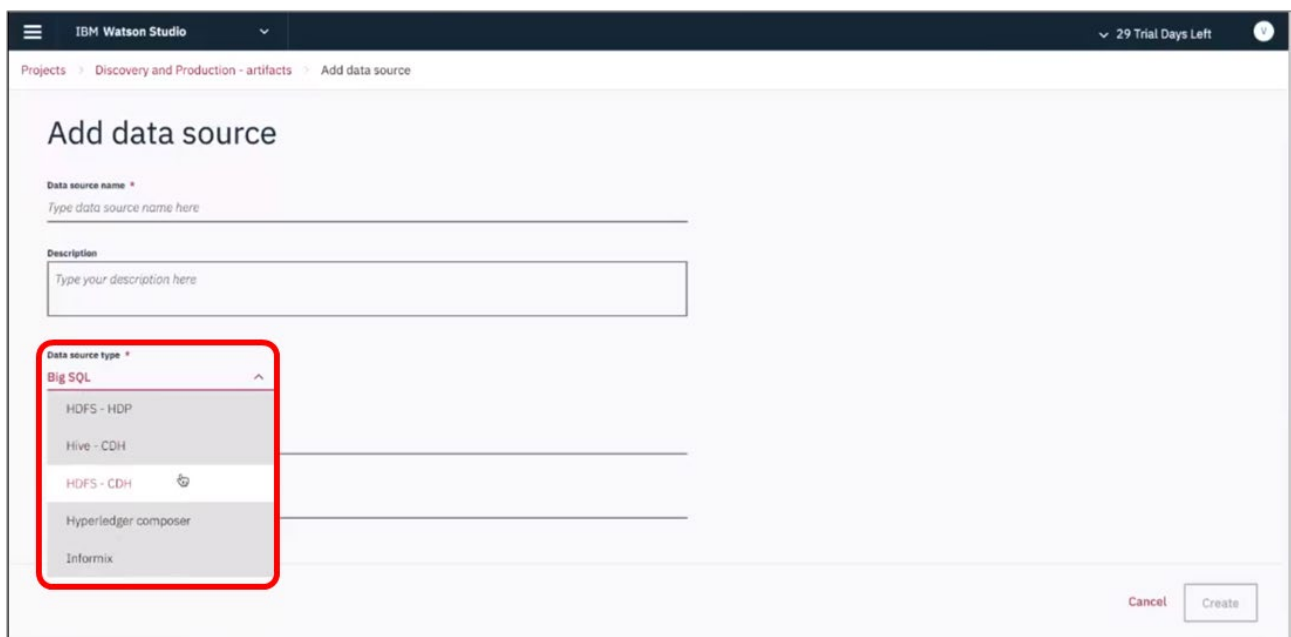
Source: Enterprise Strategy Group

IBM Watson Studio integrates with over 120 data sources via built-in connectors, to enable organizations to apply AI to their data wherever it lives.

> ### ⓘ Why This Matters
>
> AI and ML present organizations with significant challenges, exacerbated by a shortage of experienced or trained staff and the lack of the requisite IT infrastructure to support ML. Most respondents are using three or more tools for ML, which greatly complicates the iterative, cyclical learning process that drives ML. This leads to issues with time to business value. What is needed is a solution that gives organizations the flexibility to build and train models where the data resides, using familiar tools, simple enough for non-data scientists.
>
> ESG testing revealed that IBM Watson Studio provides simple tools data scientists, application developers, and subject matter experts can use to work together to build and train models at scale. Integrated data sources give the flexibility to build and train models where the data resides, and pretrained activities and notebooks accelerate the process. Extensive help and training resources keep it simple for non-experts.

IBM Watson Machine Learning

Once a model is built and trained, it can be deployed, auto-retrained, and managed using IBM Watson Machine Learning. WML focuses on the deploy phase of the data science lifecycle. It enables businesses to quickly start model management, speed time to deployment, and simplify the process of ensuring production-worthy accuracy as part of operationalizing the data science lifecycle. Organizations can leverage an automated, collaborative workflow to use trusted data to put machine learning and deep learning models into production. IBM Watson Machine Learning Accelerator—formerly known as PowerAI Enterprise—is designed to deliver improved and updated integration to IBM Watson Machine Learning especially targeted for deep learning workloads that require high computing power for training and inference. Available as an add-on to the WML base offering, WML Accelerator targets organizations moving from exploration and single-node deployment to scale-out production environments and provides:

- Distributed, scale-out processing of machine learning and deep learning workloads in a hybrid multi-cloud environment.

- Parallel hyperparameter optimization.

- Training visualization.

- Transparent scaling from a single server to many servers.

Additional branches of the IBM Watson offering include Watson Visual Recognition and Watson Natural Language Classifier from Watson Studio, which are integrated for transfer learning capability. Users can bring in new data to a base model specific to their domain and retrain. Neural network synthesis (NeuNetS) is a feature of the IBM Watson OpenScale offering that is also integrated with WML. Organizations can search deep learning architectures to find the best one for a specific data set and problem.

ESG Testing

ESG began by creating a model for automated machine learning (Auto ML). There are just three steps to model creation—data selection, training, and evaluation. In this example, we created a simple model to predict a binary yes/no condition from a small data set. As shown in Figure 6, we selected the WML instance, then selected Spark ML to create the model

builder, and chose the Automatic build method to choose the estimators to be used. Manual lets a user choose the estimators to add to the model themselves.

Figure 6. Creating a New Auto ML Model
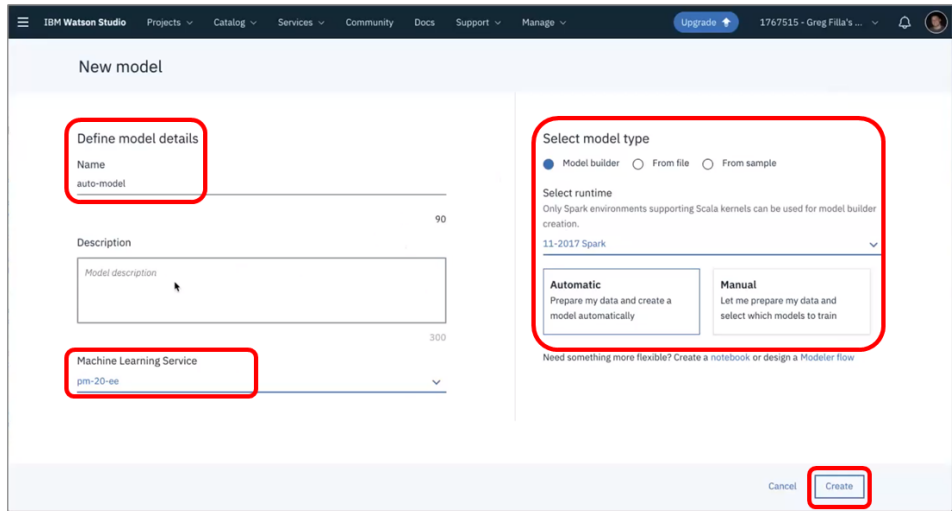
We clicked Create, which prompted us to select a data asset. Once data was selected, we selected training criteria.

For training, we selected the column value to predict, the columns to use to make the prediction, the type of classification—binary in this case—and the data split for fractional training. IBM is in the process of evolving the Auto ML capability toward Auto AI with additional sets of automation and productivity features.

**Figure 7. IBM Auto ML Training Report**
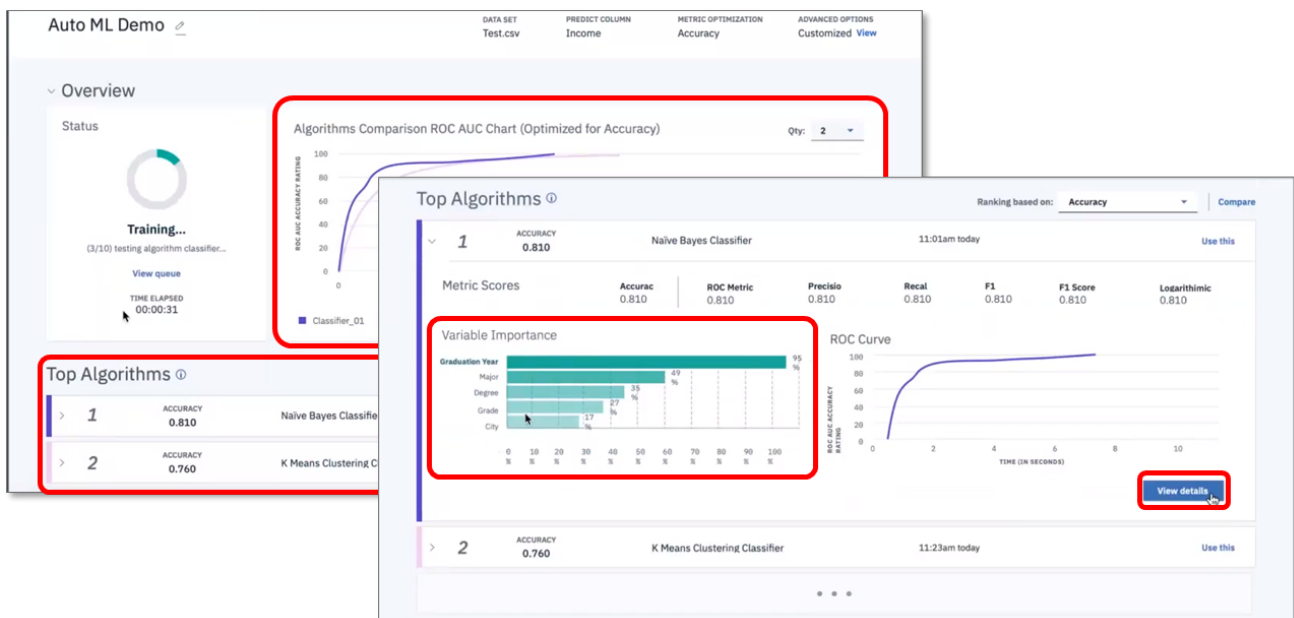
Once training was in progress, we were able to access a detailed report on the performance of the model, shown in Figure 7, including performance over time, leaderboard rankings of the algorithms, and the importance of the variables inside each algorithm. Clicking View Details brings up a detailed view using IBM's model viewer component. It's important to note that this is the same model viewer as found in IBM's SPSS modeler, so users get a consistent view and experience, wherever their data or models live. Once training is complete, the model can be saved with one click to the organization's Watson Machine Learning instance or to a local machine.

Figure 8. IBM SPSS Modeler
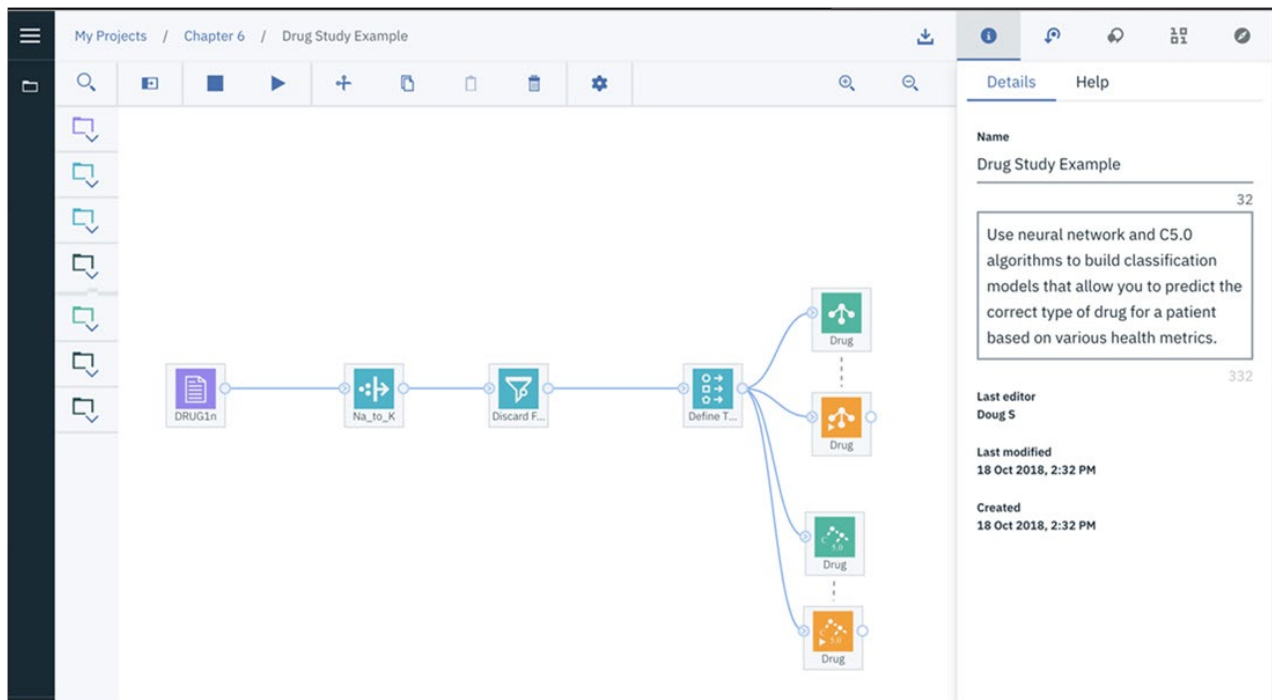
Source: Enterprise Strategy Group

Models can be visually examined and modified using SPSS Modeler, IBM's visual data science and machine learning solution. SPSS modeler helps speed up operational tasks for data scientists. Organizations can leverage data assets and modern applications with complete algorithms and models that are ready for immediate use in hybrid environments and designed to meet robust governance and security requirements.

Finally, in Figure 9, ESG walked through committing a project to production and confirmed that discovery and production artifacts are available in one location for visibility and traceability of models. These integrated features between Watson Studio and Watson Machine Learning help an organization to simplify the go-live process, accelerate AI/ML lifecycle management, and scale deployment tuned to business demand.

Figure 9. Commit to Production and Repository of Project Releases
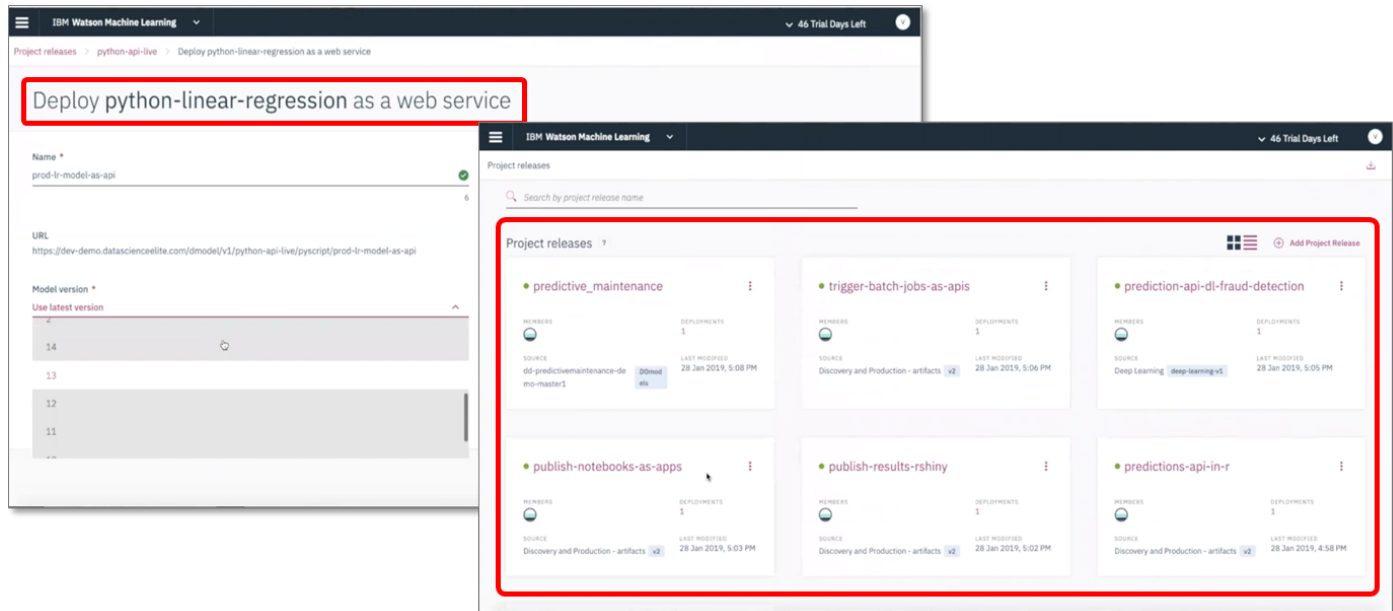
Aligning data science and application development with or without DevOps is often a challenge given the organizational dynamics, cultures, and skill set variances. The integrated development environment (IDE) for AI and data science from Watson Studio and Watson Machine Learning can help bridge the gap between these two areas and empower a business to choose the optimal models for applications to monetize their investments.

ℹ️ Why This Matters

According to ESG research, over 64% of respondents are using three or more tools for ML, which greatly complicates the iterative, cyclical learning process that drives it. This leads to issues with time to business value, with over 50% expecting ML to take more than a year to return significant business value, and nearly one in five expecting it to take more than two years.[3] What is needed is a solution that is focused on the deploy phase of the data science lifecycle, designed to put machine learning and deep learning models into production quickly.

IBM Watson Machine Learning brings analytics to any cloud or data center environment with anytime, anywhere access—without having to worry about cloud cost overages for private data center deployment or about significant hardware investment for public clouds. WML Accelerator enhances this capability, using distributed, enterprise-ready, scale-out processing of machine learning and deep learning workloads in a multi-tenant, hybrid cloud environment to provide accurate results faster, increase resource utilization, and simplify management.

ESG testing revealed an optimized user experience with the right tools for all roles involved with AI/ML—data scientist, app developers, and IT admins. Organizations can deploy applications anywhere in a hybrid environment to operationalize data science faster, at scale. The IBM SPSS modeler helps enterprises accelerate time to value and achieve desired outcomes with complete algorithms and models that are ready for immediate use. ESG found that Watson Studio and Watson Machine Learning provide an end-to-end environment that helps organizations apply learning from production and quickly iterate while ensuring visibility across data science, application development, and business teams.

---

[3] Source: ESG Survey, *Machine Learning and Artificial Intelligence Trends*, June 2017.

IBM Cloud Private for Data (ICPD)

The pace of innovation in AI is accelerating, and while there is a lot of attention on governance of data assets and compliance with policies, little attention has been given to governing the complexity associated with the lifecycle management of data science and machine learning. The range of open source frameworks in data science makes governance difficult for the typical enterprise. This is an area where ICP for Data can provide seamless governance to both asset curation with quality assessments ("governance for insight") and policy-based enforcements ("governance for compliance"), through its integrated platform, encompassing key services from IGC, Data Stage, and DSX Local. ICPD is designed to bring the power of IBM analytics to a simple validated package, repackaging proven technologies. As a foundational component of the AI ladder, ICPD helps to operationalize collection, organization, analytics, and infusion of AI seamlessly.
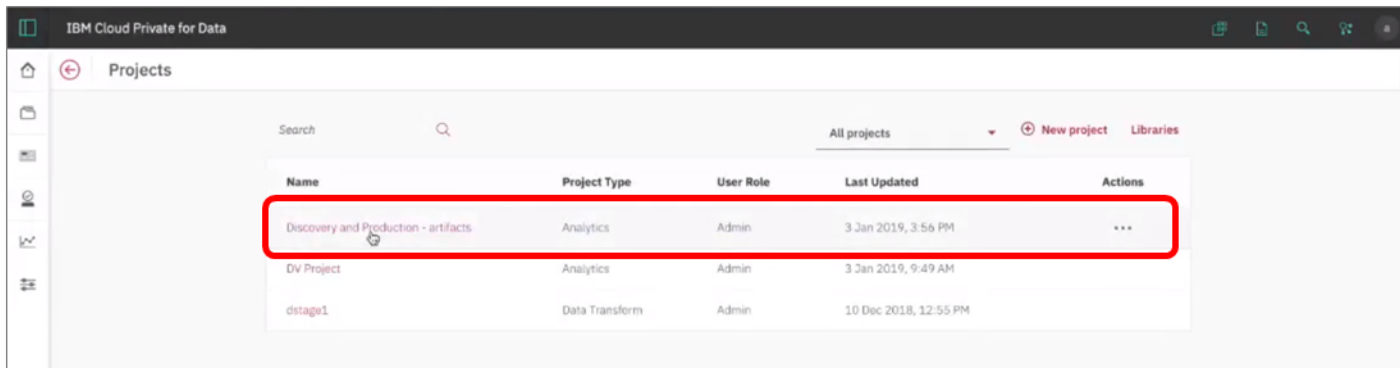
ESG Testing

With IBM Cloud Private for Data, organizations can modernize their information architecture (IA) and start their journey to becoming AI-driven. ICPD supports and governs the end-to-end AI workflow, as shown in Figure 10.

Figure 10. How ICPD Maps to Real-world AI
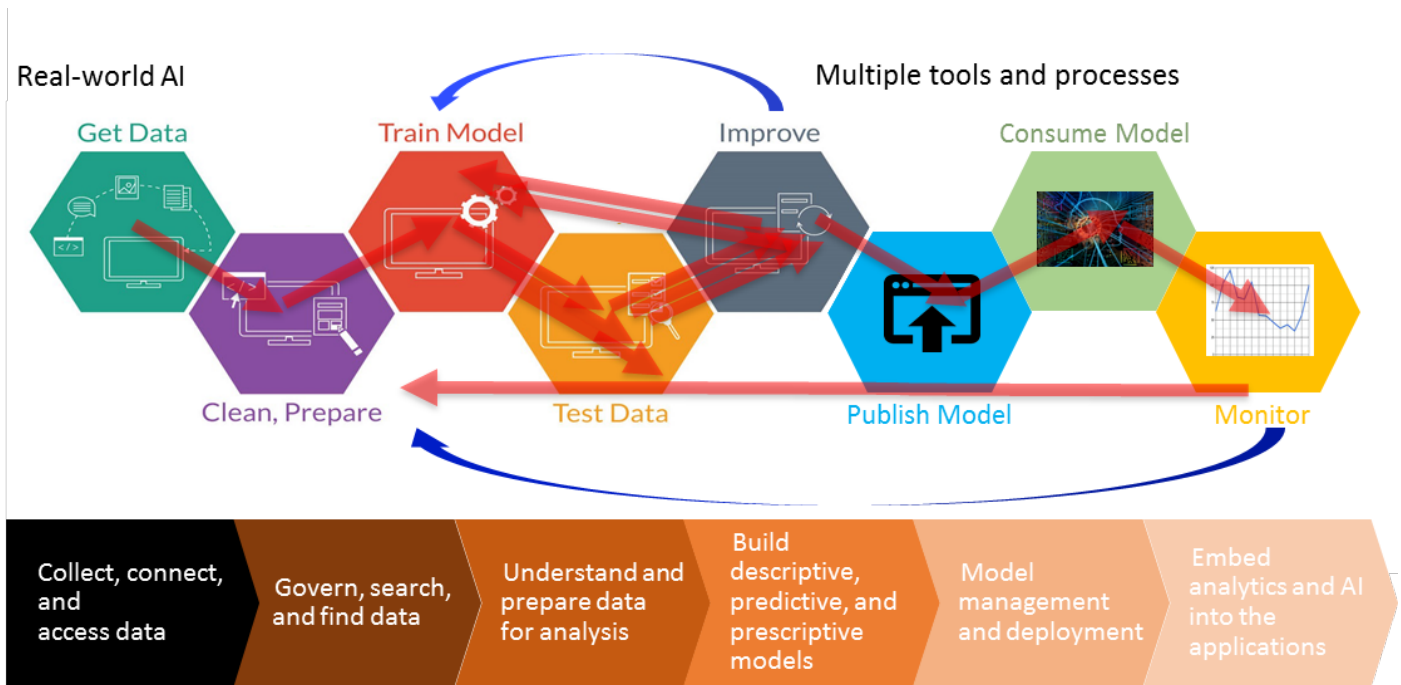
Source: Enterprise Strategy Group

The Enterprise Catalog enables the right users to find the right data and analytics assets—indexed for search, with lineage, usage metrics, and quality profiles. Users can pull these assets into their analytics projects, where they can cleanse, shape, understand, and model their data. ICP for Data supports open source and IBM frameworks—Spark, TensorFlow, IBM SPSS Modeler, CPLEX, etc.—with model management and deployment capabilities providing governance across dev, test, staging, and production. Models are versioned and scaled automatically through load-balancing to meet SLAs. Model performance is automatically monitored and can trigger model retraining and redeployment to be released as rolling upgrades.

Figure 11. WML Project Embedded in ICPD

Figure 11 shows the project we created in Watson Studio. Watson Studio and Watson Machine Learning—along with all their assets—are embedded in ICPD. Organizations can accelerate time to value and increase productivity of data science and business teams by purchasing a Data Science premium add-on to ICPD. This Data Science premium add-on includes SPSS Modeler, Decision Optimization, and Watson Explorer and is designed to enable the business to mix and match capabilities based on the consumption model.

ⓘ Why This Matters

In addition to the plethora of disparate tools, lack of experienced, trained personnel presents an obstacle to organizations working toward operationalizing AI and ML. IBM Cloud Private for Data (ICPD) is an open, cloud-native information architecture for AI.

ESG validated ICPD's cloud-native, fully governed platform, verifying that organizations can keep data secure at the source and add preferred data and analytics microservices as needed. ICPD's consumption-based model enables organizations to move among tools easily—IBM SPSS to Decision Optimization on Cloud, to Watson Explorer, for example.

The Bigger Truth

It wasn't too long ago that AI was science fiction. Now, any organization, large or small, can go beyond just proof of concept to reaping the benefits of operationalized AI/ML programs. However, AI/ML is not a silver bullet, and organizations face significant challenges in implementing AI. One-quarter of organizations lack trained or experienced staff; 19% need better IT infrastructure capabilities; 18% say there is a lack of proven technologies; and 15% say there is a need for better data science tools.[4]

IBM's prescriptive "AI ladder" approach to accelerating organizations' journey to AI describes four layers: simple, accessible data collection; data organization to support a trusted foundation for analytics; scalable analysis with AI everywhere; and infusion, or transparent operationalization of AI. IBM Watson Studio, Watson Machine Learning, and Cloud Private for Data are designed to help organizations scale the rungs of the AI ladder.

ESG testing validated that IBM Watson Studio and Watson Machine Learning made it easy to apply intelligence to data to help organizations transform business operations, on any cloud. WML Accelerator enhanced this capability, using distributed, enterprise-ready, scale-out processing of machine learning and deep learning workloads in a multi-tenant, hybrid cloud environment to provide accurate results faster, increase resource utilization, and simplify management. IBM Cloud Private for Data's open, cloud-native information architecture for AI enabled ESG to securely add preferred data and analytics microservices quickly and flexibly.

Organizations seeking a scalable AI/ML stack with a simple onramp that works on-premises, in hybrid- and multi-cloud environments, on laptops, edge, and mobile devices, and that want their AI practitioners, data scientists, and line-of-business leaders' staff to quickly and easily operationalize AI projects, should investigate how IBM Watson Studio and Watson Machine Learning can help transform business processes and modernize operations with AI initiatives.

---

[4] Source: ESG Survey, *Machine Learning and Artificial Intelligence Trends*, June 2017.

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

© 2019 by The Enterprise Strategy Group, Inc. All Rights Reserved.

www.esg-global.com   contact@esg-global.com   P. 508.482.0188