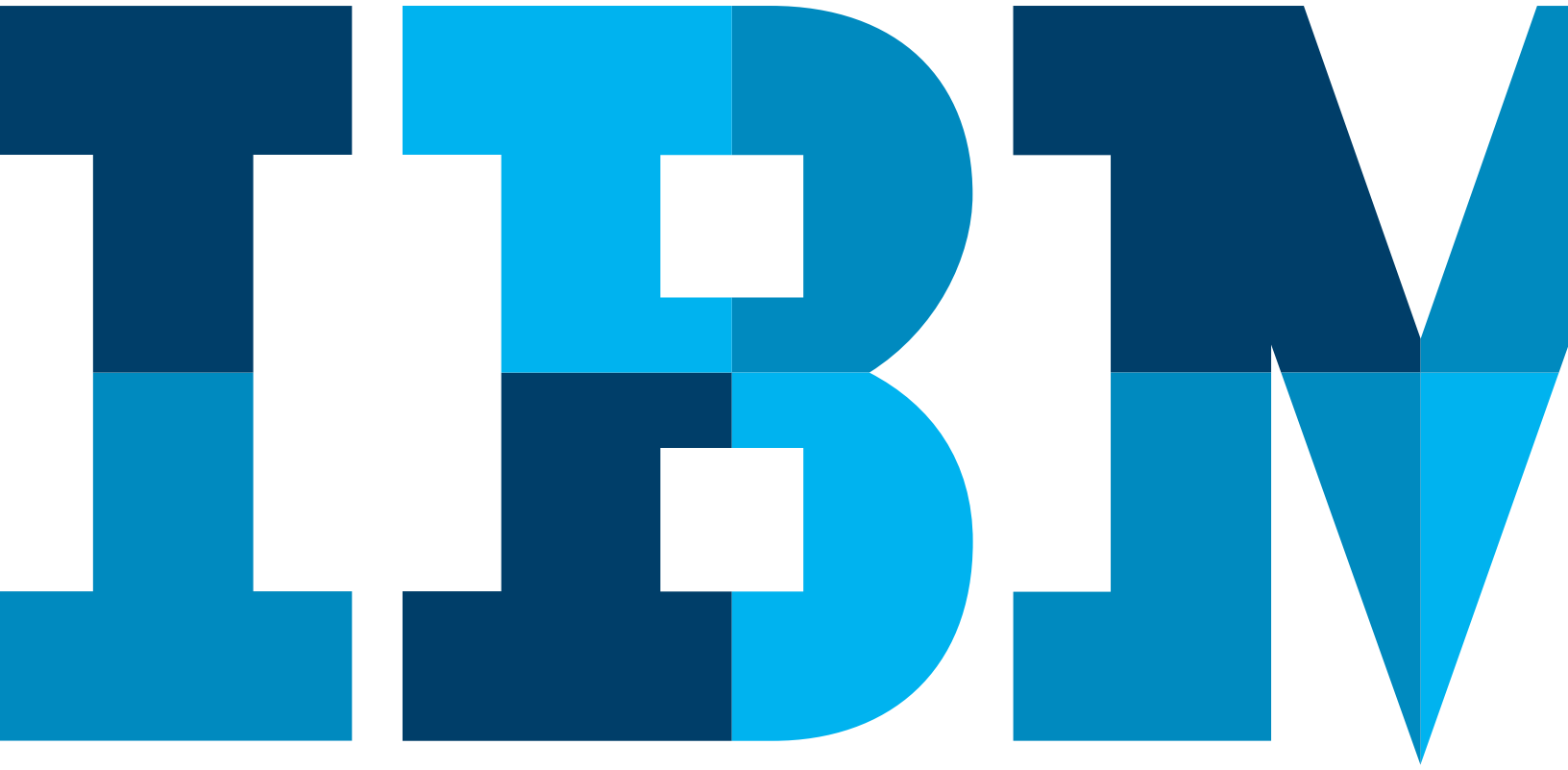
Cross-industry

The IBM analytics and machine learning advantage

Optimize the valuable data behind your firewall



IBM

Executive summary

IBM understands the value of your enterprise data and is committed to finding new ways for you to leverage your data to improve insights and increase competitive advantage. A Forrester Research study states, "Insight-driven businesses bring insight, not just data, into every decision, and they know exactly how to use them for greatest advantage across the entire customer life cycle. For these firms, digital insights, and what they do with them, are their secret weapons to disrupt your market and steal your customers."[1] Providing efficient, secure access to all relevant data sources—often in real time—is vital to delivering the right insight, at the right time, to the right person.

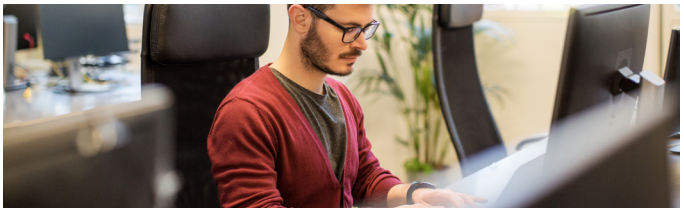
A significant amount of the freshest, most valuable data resides behind the firewall. Often, this more sensitive, personal data originates on IBM® Z® transactional systems—an industry leader in qualities of service, including availability and reliability in a security-rich environment. This data represents a unique combination of information about your clients and business that only your organization possesses. Leveraging this data for analytics and machine learning can give you the insights you need to outperform your competitors.

Organizations can more securely and efficiently take advantage of the high-value data originating on IBM Z by using hybrid transactional and analytical processing (HTAP) that brings the analytics to the data. The IBM Z HTAP approach, uses the IBM Db2® Analytics Accelerator to service transactional to complex analytics workloads in an integrated, security-rich environment with virtually no data latency. HTAP eliminates the need for data movement through multiple platforms that's typically associated with legacy analytic architectures.

This advantage enables organizations to process real-time analytics on their live transactional data without impacting transactional workloads and benefit from the platform's resiliency that's on par with their transactional applications. They can also combine data and insight from other sources, such as structured and unstructured data from other systems for use in analytics. This flexibility means more personalized insight, in real time, at the point of customer interaction.

IBM offers other products to complement the HTAP approach. The IBM QMF™ technology delivers scalable, enterprise analytics and data visualization. It is optimized for IBM Z data sources and can combine data from an extensive number of structured and unstructured data sources from IBM Z and non-Z platforms. The IBM Data Virtualization Manager for z/OS® technology can virtualize and combine IBM Z and other enterprise data sources in real time.

Another technology making a significant impact on the *insight-driven business* is machine learning. Organizations consider using machine learning when they have a complex task or problem involving a large amount of data and lots of variables, but no obvious set of rules or formulas to solve the problem. There are two fundamental types of machine learning: supervised learning and unsupervised learning. See the callout comparing the two types on the next page.



Supervised learning **builds a model that makes predictions based on probability. A supervised learning algorithm takes a known set of input data and known responses to the data, called labels, and trains a model to generate reasonable predictions of future behavior. Supervised learning typically uses classification and regression techniques. It helps you determine whether a customer might default on a loan or which product might interest a certain customer.**

Unsupervised learning **finds hidden patterns in data. It can identify clusters of characteristics or behaviors within data sets consisting of input data without labeled responses. An example of this type of algorithm is clustering. You might cluster your customers into groups with similar behavioral traits, providing better insight than just demographic market segments.**

Preparing for a machine learning initiative requires a few preliminary steps:

1. Define the business problem.
2. Identify the data to solve the business problem. Data can be derived from different sources and comes in different forms.
3. Transform the data into a consistent format.

Data transformation must happen prior to the machine learning process and can be quite time-consuming for the data scientist. A great deal of a data scientist's time is spent transforming, combining, consolidating, aggregating, formatting and changing the shape of data. As mentioned earlier, offerings, such as the IBM Db2 Analytics Accelerator, IBM QMF and IBM Data Virtualization Manager for z/OS, can reduce the time it takes to transform data and simplify this process.

After these steps are completed, the actual machine learning process can begin. This process includes the following procedures:

1. First, a set of algorithms must be chosen to test and then identify the one that predicts outcomes with the greatest accuracy. To make this determination, data scientists typically use a trial and error process. They apply each selected algorithm to the data set to determine which algorithm appears to render the best results. This identification method can be a very time-consuming, iterative process.
2. Once the best algorithm or model is identified, the data scientist tests the model with a new subset of the same data. The model is evaluated by comparing the accuracy of the tested results.
3. After comparing model accuracy, the data scientist must ensure that the selected model is not *over fit*. This step involves a validation process that retests the model with a new subset of known, labeled data.

The organization is now ready to use the trained and validated model to score new, previously unseen data within a transactional application. Often, this process involves passing the data through the modeling software's data engine. Alternatively, deployment could involve rewriting models in application code to be deployed within the operational systems. Either approach involves significant additional work to automate the results from the modeling process and incorporate them into the production systems. Once the model is deployed, it's continually monitored to ensure that the model quality doesn't degrade over time.

The predictive modeling process requires that data engineers and data scientists follow a complex and iterative set of steps to prepare data and then develop, deploy and monitor the models. This process of a model going from idea to deployment can take months.

For many organizations, machine learning is a time-consuming, costly process for data scientists and engineers. For this reason, the full benefit of machine learning cannot always be realized on an enterprise scale. Forrester Research states that, "Data scientists are the superheroes and unicorns of today's business. But data scientists are only human, and they are reaching the limits of productivity with current processes".[2]

Data scientists and engineers face several challenges:

- **With more than 50 different algorithms for data scientists to choose from, finding the best algorithm to solve a specific business problem can be extremely time-consuming.**
- **With a rapidly growing number of parameter options per algorithm to consider and test, model development can be labor-intensive and complex. For example, the computational cost for training a single support vector model (SVM) can exceed 24 hours.**
- **Model performance can have a significant impact on business applications. Data scientists and engineers are tasked with finding the optimal algorithm to meet performance demands. There can be a wide variation in model performance depending on what algorithm implementation a data scientist selects. For example, an IBM SPSS® implementation might perform differently than a Python or SPARK algorithm.**

Due to these and other complexities, such as workload demands, these algorithm and parameter selections are often based solely on the data scientist's or engineer's skill or comfort level with specific algorithms and parameters.

To meet the challenges faced by data scientists, IBM introduced IBM Machine Learning for z/OS to help organizations gain greater insight and value from their enterprise data on IBM Z—while keeping that data in a security-rich environment.

See figure 1 for a summary of the benefits and value of IBM Machine Learning for z/OS.

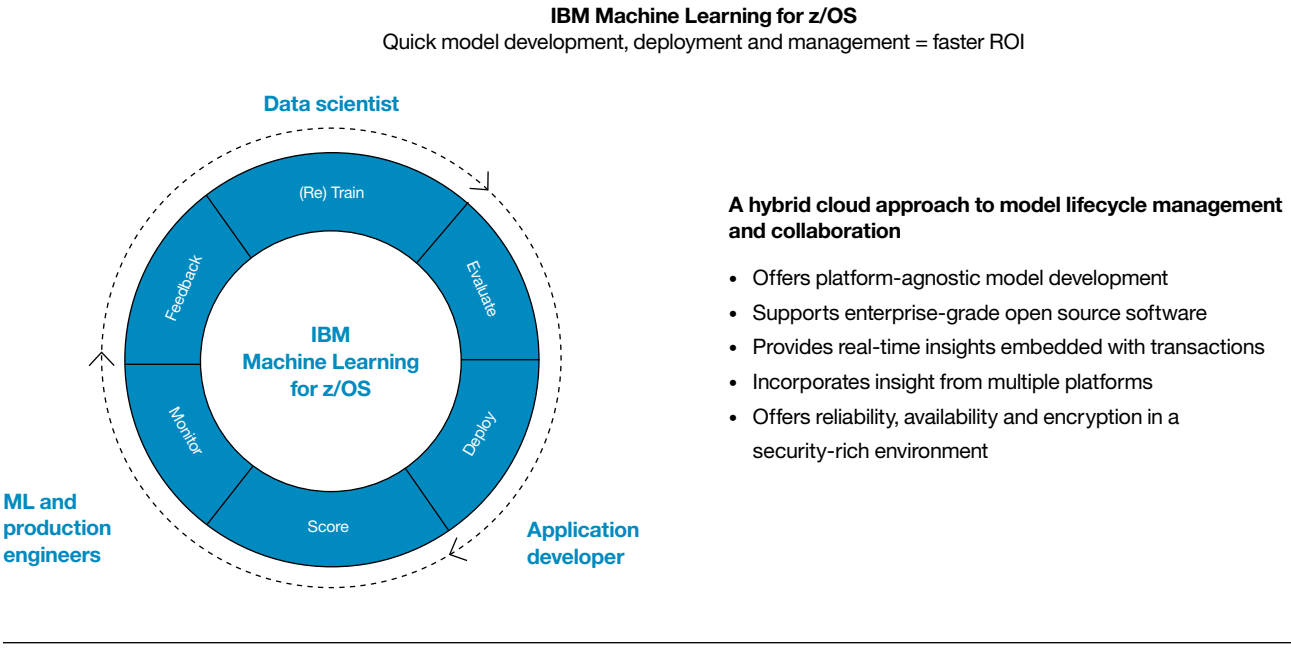


Figure 1: A summary of the benefits and value of IBM Machine Learning for z/OS.

Machine Learning for z/OS is an end-to-end machine learning solution that enables faster model training, deployment and monitoring for a quick return on investment. Organizations can use several differentiating Machine Learning for z/OS capabilities:

- **Cognitive Assistant for Data Scientists (CADS)** makes it easier for a data scientist to identify the correct algorithms and develop the best model. As mentioned, this process is usually done through lengthy trial-and-error testing.
- **Hyper parameter optimization (HPO)** helps data scientists identify and select the best parameters. This feature, in turn, helps them fully optimize the predictive capabilities of their models.
- Machine Learning for z/OS is built on **open source software**, enabling organizations to take advantage of the latest machine learning innovations without having to manage an open source software stack.
- A **continuous monitoring and feedback loop** evaluates the performance of models as they are exposed to new data. Data scientists can set a threshold for when they want to be notified, if, and when model performance deteriorates. Data scientists can also schedule regular model evaluations and feedback data can be stored for retraining to help continuously improve model performance.
- Using modern representational state transfer application programming applications **(RESTful APIs)**, Machine Learning for z/OS makes it easier to deploy a model for scoring new data—in just seconds.
- A **visual dashboard** automates and simplifies model maintenance and refresh. Organizations can easily manage the hundreds to thousands of models that a modern enterprise environment might need.

IBM Machine Learning for z/OS also offers a hybrid cloud approach to model lifecycle management and collaboration. This approach gives data scientists the flexibility to train and evaluate their models on IBM Z or an alternate platform of their choice. Organizations developing models on other platforms, using Spark or Python, can easily deploy these models on IBM Z—where the majority of their transactions occur and most valuable, sensitive data originates. Organizations can integrate scoring with transactional applications, without significant overhead, enabling them to optimize every interaction with their customers—in real time.

Machine learning production engineers can readily monitor resource usage and model accuracy to help ensure that the models remain accurate. All these features benefit from the security-rich IBM Z perimeter that provides enterprise-level resilience and encryption to help keep the data used in analytics and machine learning as secure as possible.

IBM Machine Learning for z/OS helps improve patient health at Argus

Argus Health addresses the complexity that healthcare payers, providers and pharmaceuticals face in maximizing clinical benefits for patients. As part of this initiative, Argus needed ways to encourage patients to keep their blood sugar, cholesterol and blood pressure under control. They wanted to score diabetes patients at the point of a prescription sale based on factors, such as average blood sugar level, cholesterol, blood pressure, whether they take their medications regularly and more. Depending on the patients' score, their copays might vary as they request refills of their medications.

The IBM and Argus team built a classification model to help predict the risk category of each patient by using IBM Machine Learning for z/OS. The model provides a score at the point of sale and the patient is encouraged to move up to the next category by improving their metrics so the patient can save money. Argus can positively impact the health of its patients, at a lower cost, while improving their customers' experience.

Conclusion

When deriving insights from highly sensitive data originating on IBM Z, you can take advantage of your existing investment in people, processes and infrastructure. At the same time, you can also minimize costly data movement and maintain a high level of data governance and encryption in a security-rich environment. There's an overall trend to move analytics to the data and IBM has made it possible for organizations to satisfy these requirements while benefiting from analytics advancements.

By using real-time enterprise data to make decisions, you can more quickly identify risk, improve the customer experience, reduce operational costs and deliver greater organizational agility.

Agility is all about making the right decision at the right time. IBM enterprise analytics and Machine Learning for z/OS help you leverage your IBM Z data—in place, in a security-rich environment, at the point of customer interaction—converting your most valuable, sensitive enterprise data into opportunity and opportunity into revenue.

For more information

To learn more about IBM enterprise analytics and IBM Machine Learning on z/OS, please contact your IBM representative or IBM Business Partner, or visit: **ibm.com**/analytics/us/en/enterprise

About the authors and contributors

The following people contributed to this paper:

Jonathan Sloan, IBM Analytics Solution Architect, z Analytics

Annette Zawacki, IBM Portfolio Marketing Manager, z Analytics

IBM Corporation
IBM Analytics
Route 100
Somers, NY 10589

Produced in the United States of America
November 2017

1 "Insights-Driven Businesses Will Steal $1.2 Trillion in 2020—It's Time to Join Them," Brian Hopkins, Ted Schadler and James McCormick, Forrester Research, 27 July 2016, https://img.en25.com/Web/Forrester

2 Forrester Research, 11 February 2016

Please Recycle

IMW14954-USEN-00