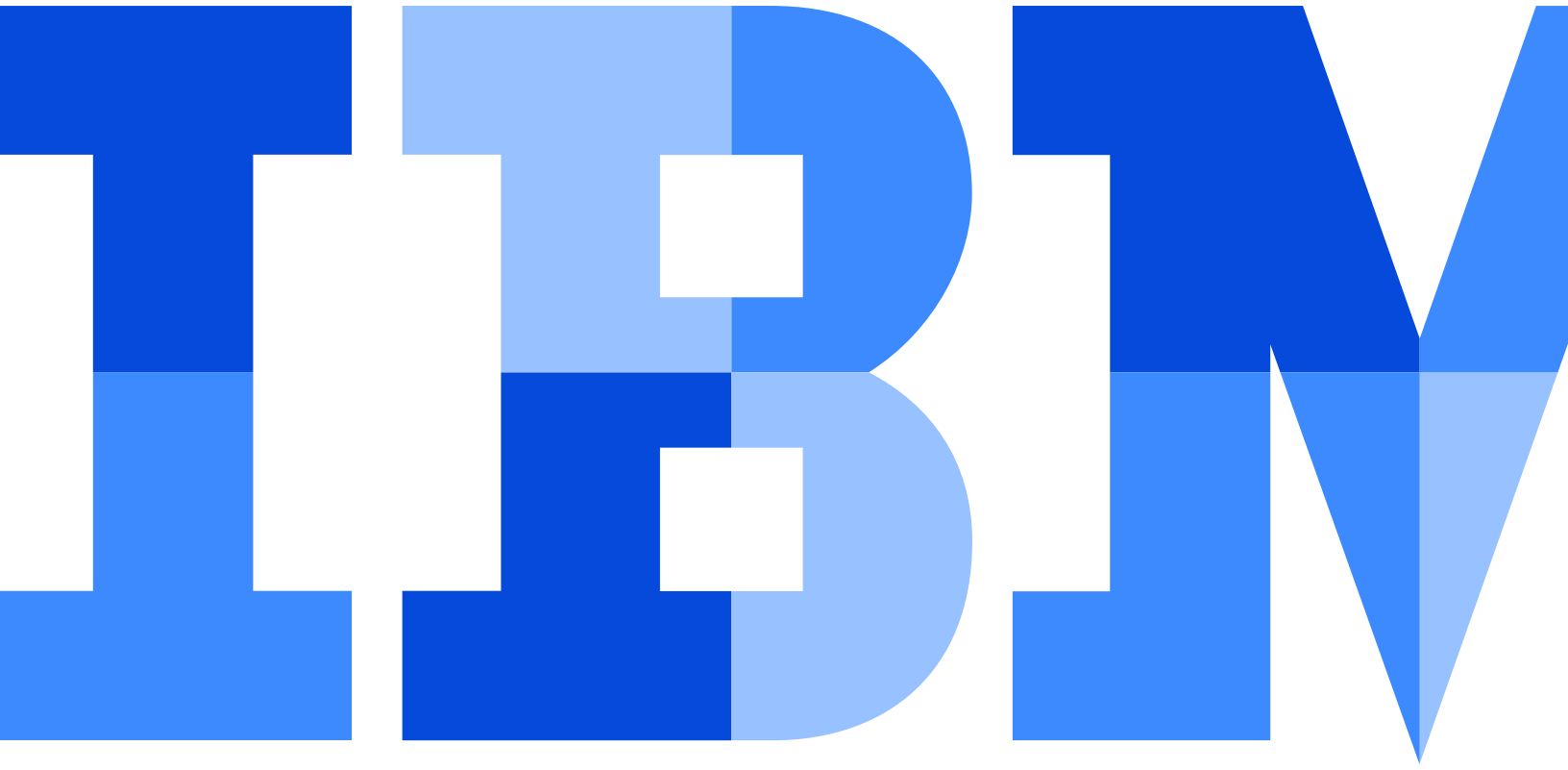
Infuse continuous intelligence into your organisation with IBM Machine Learning for z/OS

Explore all the ways in which the technology of tomorrow can help you make future-proof decisions for your business, based on existing data



IBM

Contents

What’s machine learning?

Sometimes, as humans, we make the right decisions based on the data we see and analyze. Sometimes we don’t. This is quite normal because our decisions are influenced by logical reasoning and emotional perception which can reinterpret every advisable conclusion. So, what if we could have computers augment our knowledge so that we can take the best decision every time? That’s what machine learning can help us do.

Machine learning helps data scientists identify patterns in historical data, build behavioral models from those patterns, and makes recommendations based on that insight.

Machine learning is taking center stage in application investment. In fact, Industry consultants and experts are expecting that machine learning is and will continue to disrupt all industries.

Traditional machine learning process

The figure below shows the traditional machine learning process, which requires that data engineers and data scientists follow a complex and iterative set of steps to build, train and deploy behavioral models. As part of this process, the human intervention piece can be very time consuming and costly, requiring significant development, deployment and management by data scientists and engineers.

It can take months or even years for a model to go from creation to deployment. The continuous fine tuning of a model can require more human intervention and processing.

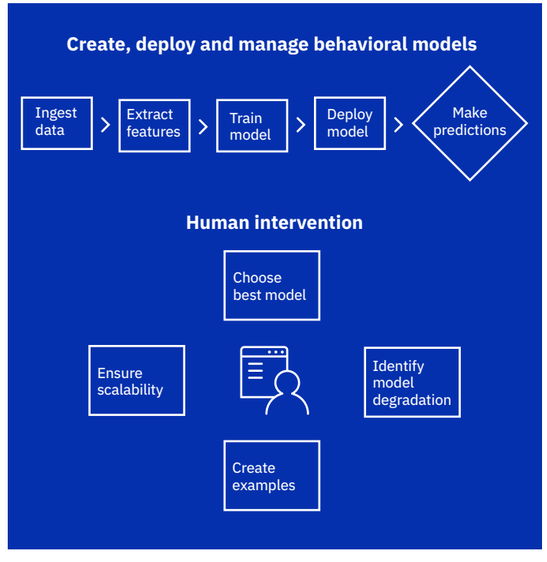


Figure 1: Traditional machine learning process.

IBM and machine learning

The IBM® vision is continuous intelligence throughout the enterprise. This consists of enabling customers to continuously make their business processes more intelligent, by combining the power of learning machines built in IBM Research with simplicity and ease of use. Even novice users can benefit from it.

A system capable of continuously infusing intelligence into your enterprise should offer the following:

- **Productivity**—Data Scientists are a valuable asset in your organization. Predictions tell that there will be a tremendous shortfall of data scientists by 2020. This means that you need to make the most out of your data science teams and equip non-data scientists with basic data science skills.
- **Trust**—As a business and technology leader, you need to know that the insights being produced are usable and based on current trends: your analytics models are only as good as the data fed in to it. If you can trust that your insights are accurate, you can deploy them in a variety of different situations.
- **Freedom**—as an enterprise and as a data scientist, you don't want to be locked in to a framework or platform. You need to have the freedom to choose a language, framework and platform of your choice.

IBM Machine Learning leverages open-source algorithms. The open-source community is providing state-of-the-art machine learning technology which is a top requirement for data scientists. IBM has started with Spark ML, a set of algorithms that can scale to big data volumes. IBM has recently added Python package support and is in the process of adding support for further popular packages for R and Java users as well as for major deep learning packages.

Furthermore, IBM intends to support users through the end-to-end machine learning workflow by providing tools to assist them and to optimize every workflow step. Data Science Experience is a tool which enables data preparation, model creation and model selection. Once the models are created, IBM Machine Learning provides tools to support model management, deployment and governance as well as monitoring the accuracy of deployed models.

Finally, IBM Machine Learning strategy is to deliver it on every platform that matters. This started with IBM Machine Learning for z/OS which includes extracted features and tools from the IBM WatsonTM Machine Learning, an IBM public cloud offering.

IBM has simplified the machine learning process

IBM Machine Learning simplifies the building and deployment of models through easy-to-use wizards. This allows data developers and data scientists to focus on the quality of the models, rather than on the complexities of the process.

One of the biggest challenges of building predictive models consists of selecting the right algorithm and setting optimal parameters. IBM has simplified this process with "Cognitive Assistant for Data Scientist" (CADS) capabilities that can suggest the best algorithm based on your data and your business scenario. The "Hyper Parameter Optimization" (HPO) feature can then help determine the best parameters to use for the selected model.

Another challenge faced by data scientists is that models tend to degrade in performance over time. What worked yesterday may not work tomorrow. With IBM Machine Learning, you can have a feedback loop which readily determines if your model performance is deteriorating and proactively retrains the model. IBM offers data scientists and engineers the ability to schedule continuous evaluations to monitor model accuracy over time and to be alerted when performance deteriorates. Feedback data is stored for retraining to continuously improve the model.

Furthermore, the data scientists face another challenge consisting of the varying tools that are required during the development of predictive behavior models. Data and application developers work in SQL, java and sometimes COBOL. Data scientists work in Python, Scala, R, SPSS and SAS. This represents a development risk where tools and applications need to be agreed upon before, during or after development; often requiring significant rework that hinders and delays the model development process.

With IBM Machine Learning, IBM has simplified and automated the management and deployment of models. A deployment interface allows to create a distribution of the model in one click or one API call. Thus, organizations can manage the often thousands of models that a modern enterprise environment will demand. A dashboard provides a health check across all models in the enterprise, offering insight into overall model performance and a quick view of those that need to be retrained.

The figure below highlights the different capabilities provided by IBM Machine Learning for z/OS mapped to each step of the machine learning workflow.

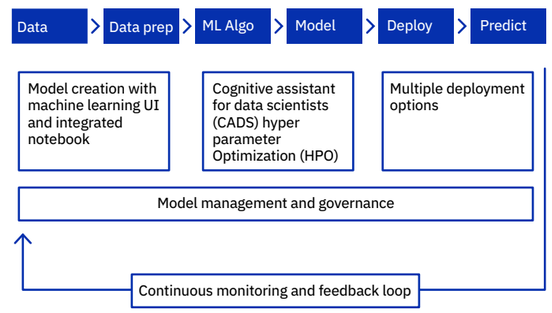


Figure 2: Capabilities provided by IBM Machine Learning for z/OS.

Why IBM Machine Learning for z/OS

In many industries, IBM Z is the platform for mission critical core applications and data. However, this role is perceived yet outside the machine learning paradigm. Due to this perception a lot of data gets offloaded to other platforms for analytical purposes. These offload procedures may result in duplicated data, additional processes, complexity, security intrusions and a long duration until the output can be processed. But what if it would make more sense to analyze data where it initially resides? What if real-time requirements make every offload to another platform inefficient or impossible?

IBM has extracted machine learning from IBM Watson, which is IBM's cognitive platform in the cloud, delivering that key capability to IBM Z where your data resides. Now, IBM Machine Learning enables you to monetize your data without having to worry about data movement, duplication, security or governance issues.

Deploying IBM Machine Learning on the z/OS platform enables you to fulfill real time requirements and access Z Systems data in place. You can combine that data with other sources of information, such as structured and unstructured data from other systems. By accessing live data on IBM Z,iu you can make optimal business decisions in the fastest possible way and sooner than your competition. In addition, you can minimize the ETL overhead in terms of cost and time and maintain the established level of security. This helps you focus on the analytical procedures and not get diverted by side tasks like ETL. In terms of scoring, predictive analytics demands prompt scoring results that have real time requirements and which need to be embedded in sub-second online transactions. Co-locating machine learning analytics with data on the same physical server enables immediate and accurate insights.

Financial services

- Fraud detection
- Customer risk scoring
- Credit monitoring
- Product recommendations
- Planning assistance
- Customized withdrawal limits
- Portfolio tax optimization
- Spending patterns
- Credit increase worthiness
- Customer retention
- ID theft detection
- Identity management
- Sentiment and news alalysis
- Spending impact influencers
- Documentation review

Retail

- New store locations
- Shelf, store and package optimization
- Seasonal planning and forecasting
- Consumer trends
- Optimal product blend
- Personalized promotions
- New label products and product categories
- Price optimization
- High-end purchase anomalies
- Product lifecycle

Manufacturing

- Responsive machines
- Rare failure reduction
- Increased production capacity
- Internal defect reduction
- Accelerated price determination
- Better integrated process flow
- Improving preventative MRO
- OEE improvements
- Quality production forecast
- Demand forecast accuracy
- Optimized product customization

Figure 3: IBM Machine Learning for z/OS use cases in 3 exemplarily picked industries.

IBM Machine Learning for z/OS Use Cases

IBM Machine Learning for z/OS is the integral tool for data empowerment. It enables enterprises and organizations of any kind to not only gain valuable insights into their data but to make sense of correlations and ultimately to make predictions. In that, IBM Machine Learning for z/OS acts as the enterprise's focal point and interface between internal operations and customers. The solution's capabilities allow for diverse use cases across all industries that eventually lead to revenue increase, cost reduction, risk minimization and a targeted market approach.

Let's consider a use case in the financial services industry. A large bank that is using an in-house credit card operations solution on Z is looking for a way to detect fraudulent transactions before they occur. With IBM Machine Learning for z/OS the bank has a solution that can be implemented on the core transactional data and allows for immediate intervention on a transaction basis.

For real time scoring IBM Machine Learning for z/OS leverages the customers' financial and purchase patterns, both historical and current, and combines it with information from external data sources like Google location sharing, social media and personality insight information from Watson. The implementation of IBM Machine Learning for z/OS could result in over 30 percent fewer fraudulent transactions and thus in a decrease in support costs and credit card replacements. Additional financial benefits result from the lower total cost of ownership (TCO) through the deployment on Z versus offloading the data. All in all, IBM Machine Learning for z/OS can generate tremendous savings and increase customer satisfaction for this bank.

Looking at other industries, the use cases are countless. For instance, the retail industry today is dominated by consumers' increasing online presence. Customers reveal their shopping habits and preferences online and retailers can access social media data and combine it with their own customer data on purchase history to better understand their customers, maximize spending and offer personalized promotions. Forward-thinking retailers see limitless possibilities in IBM Machine Learning for z/OS, spanning from sourcing, buying and supply chain management to marketing, sales and improved customer experience.

Taking a closer look on the manufacturing industry, it is experiencing a never seen increase in available data today. This data compromises a variety of different formats, semantics and quality, for example sensor data from the production line, environmental data or machine tool parameters. Through IBM Machine Learning for z/OS the manufacturing industry can benefit from the increased data volume for quality and safety improvements and for process optimizations and cost estimations. That way, the high complexity and dynamics involved in manufacturing can be handled more efficiently.

The data revolution affects and sometime even dominates all industries, enterprises and organizations. Future oriented players will seek for and deploy machine learning solutions to exploit the possibilities that big data offers. IBM Machine Learning for z/OS will be key to success.

IBM Machine Learning for z/OS Architecture

IBM Machine Learning for z/OS integrates powerful IBM Machine Learning capabilities, including IBM Open Data Analytics for z/OS which extends the IBM Z open analytics beyond Apache Spark on z/OS to include Python and Anaconda Python packages for data science (z/OS IzODA Anaconda).

The optimized data abstraction and virtualization layer included in IBM Open Data Analytics for z/OS enables efficient access to a broad set of structured and unstructured data sources through Spark APIs. With this capability, traditional z/OS data sources such as IMS, VSAM, Db2, PDSE, PS, or SMF data can be accessed in combination with data from other platforms in a performance-optimized manner with Apache Spark or Python/Anaconda applications.

As shown in the figure below, IBM Open Data Analytics for z/OS runs on z/OS and continuously ingests data into the machine learning for z/OS pipeline in real time. The application cluster runs on Linux and provides a web user interface for model training and management. Data scientists can use the user interface to select data sources for model building, training, and deployment. Application developers can see the list of saved models for deployment. Deployed models are used to make predictions. The models can be optimized as the data is pushed back into the machine learning workflow in a feedback loop. The application cluster components can run either on Linux on Z or on Linux x86 providing flexibility depending on the available IT landscape.

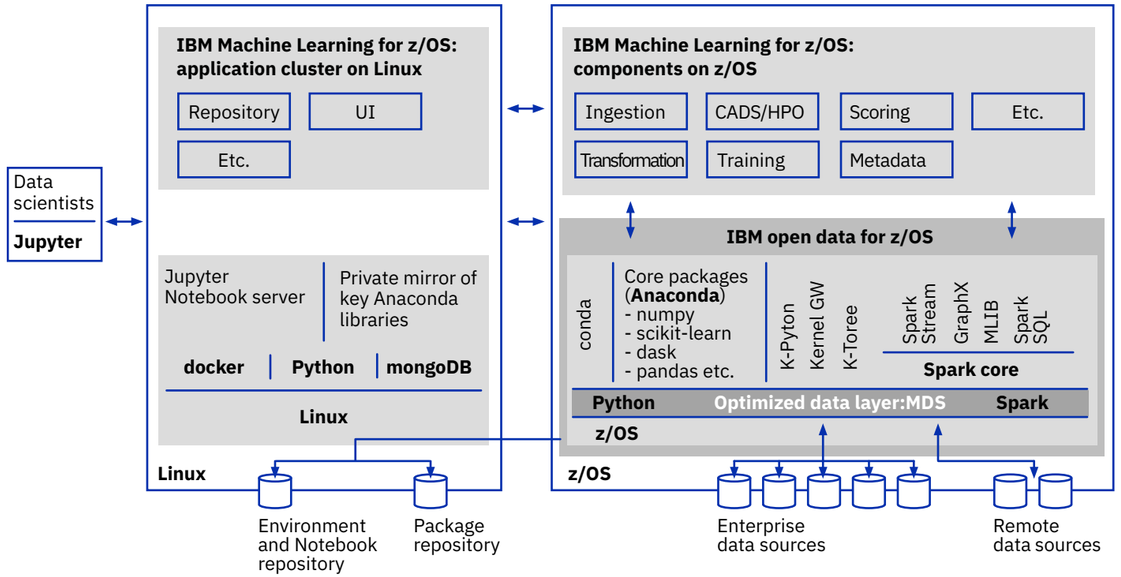


Figure 4: IBM Machine Learning for z/OS Architecture

In summary, Machine Learning for z/OS provides the following functionality to help you maximize the value from your enterprise data:

- Ingestion of data from different sources, including e.g. Db2, IMS, VSAM, and System Management Facilities (SMF)
- Transformation and cleansing of data to be used as the algorithm input
- Training of a model for the selected algorithm with the prepared data
- Evaluation, deployment, and management of the trained model
- A RESTful API for an application to embed the predictions that are made by the model
- Monitoring of the status, accuracy, and resource consumption of the model
- Automated feedback loop to feed new data for re-training of the model if the accuracy deteriorates
- An integrated notebook interface for data scientists to use machine learning APIs for interactive processing
- Security control through integration of authentication and authorization on Z

IBM Open Data Analytics for z/OS, the foundation of IBM Machine Learning for z/OS

The figure below shows the different components of IBM Open Data Analytics for z/OS which include the Spark components including the Spark Core where task dispatching, scheduling and I/O take place and where the Data Frames (DFs) are created. Spark includes also different libraries for various analytics workloads consisting of Spark SQL, Spark Streaming, GraphX and MLlib.

The optimized data abstraction and virtualization layer enables efficient access to a broad set of structured and unstructured data sources through Spark APIs. With this capability, traditional z/OS data sources such as IMS, VSAM, Db2 for z/OS, PDSE, PS, or SMF data can be accessed in combination with data from other platforms in a performance-optimized manner with Apache Spark or Python/Anaconda applications.

IBM Open Analytics for z/OS includes now Python and Anaconda Python packages for data science which enable IBM Machine Learning for z/OS to support the very popular Python programming language in addition to Scala and java. This allows to create, train, deploy, and monitor models leveraging popular libraries such as scikit-learn through the machine learning for z/OS framework.

By combining industry standard open source analytics frameworks with the Optimized Data Layer, data scientists are enabled to access mainframe data-in-place for analytics without requiring mainframe expertise. This creates a unified experience and expands the ability to extract insights from a wide range of data to a new generation of users who previously may not have had the skills needed to analyze data in z/OS.

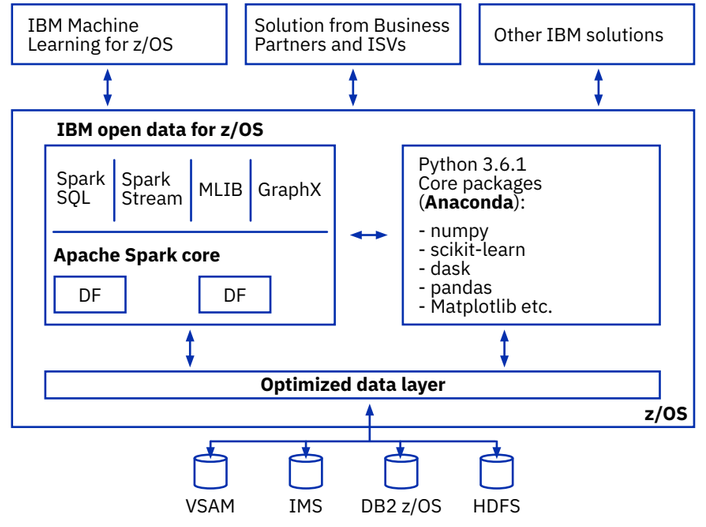


Figure 5: IBM Open Data Analytics for z/OS.

With z/OS system characteristics, there are various key benefits of IBM Open Data Analytics for z/OS e.g.:

- Real-time, fast, efficient access to current transactional data and to historical data.
- Avoid security and data governance intrusions, especially for sensitive data that are governed with z/OS security capabilities.
- Data in place Analytics allows analysis of both real-time operational data and warehouse data.
- Spark can be clustered across more than one Java virtual machine (JVM), and these Spark environments can be dispersed across an IBM Parallel Sysplex.
- Spark workload and a large majority of cycles used by Optimized Data Layer are zIIP-eligible for affordability
- Simultaneously Multithreading (SMT-2) for added thread performance and Single-instruction, multiple-data (SIMD) for better performance on selected operations.
- Possibility to integrate Spark with IBM Db2 Analytics Accelerator.

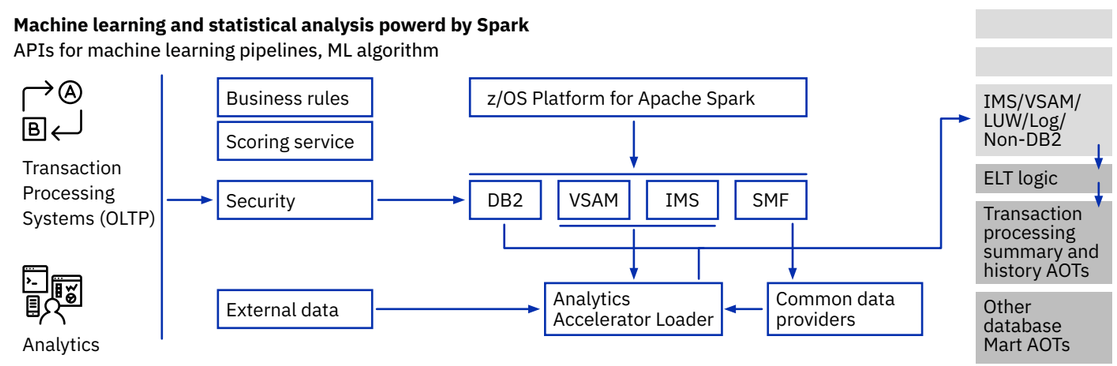


Figure 6: IBM Db2 Analytics Accelerator with IBM Machine Learning for z/OS

IBM Machine Learning for z/OS and IBM Db2 Analytics Accelerator

Many IBM enterprise customers are interested in leveraging the latest analytics technologies. They want to take advantage of their existing investment in people, processes and infrastructure on z Systems while also minimizing costly data movement and maintaining high degree of data governance and security.

IBM has made it possible for customers to satisfy these requirements while also benefiting from the latest analytics advancement like IBM Machine Learning for z/OS and the Db2 Analytics Accelerator.

The figure below features how combining IBM Machine Learning for z/OS along with the Db2 Analytics Accelerator enables you to access Z Systems data in place and combine that data with other sources of information, such as structured and unstructured data from other systems. You can then build models to predict behaviour and make the most optimal business decisions. And by accessing live data on z Systems you can be more agile, minimize ETL time and costs and maintain the highest levels of security.

The Accelerator excels in the data transformation processing necessary to support machine learning. Transforming data from the format necessary for transactions to the format necessary for analytical processing is a resource intensive endeavour. Therefore, the Accelerator is very complementary to machine learning for z/OS.

Situation #1	**Situation #2**	**Situation #3**	**Situation #4**
ML for z/OS	ML for z/OS	ML for z/OS	
	Db2 analytics accelerator	Db2 analytics accelerator	Db2 analytics accelerator
• Small amount of Z data • **zIPP / memory with sufficient capacity** • **Scala, Pyton for data prep required** • Jupyter Notebook of ML for z/OS used • **No SQL** skills or SQL not appropriate	• Large amount of Z data • **DataStage** (or similar tool) already used • SQL ok for data preparation • Z capacity insufficient for data preparation • Ml for z/OS **Jupyter Notebook** used for addtl. data prep (similar to SAS data cube build) • Supporting broader **data lake topologies** • Accomodating more data clie to **archiviрg** • Need for limited R support	• Large amount of Z data • **Db2 Analytics Accelerator already deployed** • Addtl. points similar to situation #2 • Need for limited R support	• Large amount of Z data • Batch scorirg in the Accelerator • **SPSS Modeler** and SPSS C&DS already used (integrates with the Accelerator) • Interest in **Accelerator in-DB Analytics** • Need for limited R support • Supporting broader **data lake topologies**

Figure 7: IBM Db2 Analytics Accelerator with IBM Machine Learning for z/OS—complementing values

Four types of applications can benefit from leveraging the Accelerator:

1. Data transformation
2. Batch predictive model building and scoring within the Accelerator
3. R processing (limited language support) and predictive model building within the Accelerator
4. Loading and storing SMF data in support of "IT Predictive Analytics"

The figure below provides some more details regarding the decision making as to whether Db2 Analytics Accelerator, or machine learning for z/OS, or both may be a good fit for a specific set of requirements.

IBM Machine Learning for z/OS newest capabilities

IBM is continuously improving the machine learning for z/OS offering by adding new functions and features. Following are the new functions included in machine learning for z/OS 1.1.0.4:

1. Integrated online scoring service in CICS
2. IT Operational Analytics (ITOA) Solution Template - System Health Tree sample application: a set of assets that can be used as a template to be further customized to provide advanced analytical insight regarding the Z operational environment. The IBM Machine Learning for z/OS ITOA Health Tree application is an add-on feature that runs on a Linux or Linux on z system
3. Multiple model versioning deployment support: different versions of a model can be deployed on different scoring services or clusters
4. Export & import models across different machine learning platforms: a model trained on machine learning for z/OS can now be exported, and then imported on other IBM data science/ML platforms like Watson Studio with Watson machine learning for IBM Public Cloud or Data Science Experience (DSX) Local for x86, Power, or Linux on IBM Z
5. Visual Model Builder (VMB) enhancement
6. Support Scikit-learn model schema and key-value pair input
7. Support Remote Distributed Spark clusters (leveraging the Apache Levy API) as training runtime
8. Add compute node to machine learning for z/OS
9. Full support for Linux on IBM Z

Following are the new functions included in machine learning for z/OS 1.1.0.5:

1. Performance enhancement for online scoring deployment in CICS integrated scoring service
2. Save and Publish models to other platforms from UI
3. Scoring service creation and configuration using z/OSMF workflow through administration dashboard
4. Upgrade enhancement
5. Up & running enhancement
6. Scalability enhancement for scoring process with uWSGI server for scikit-learn models, where uWSGI is an WSGI (Web Server Gateway Interface) standard based container to execute Python web applications
7. PySpark support
8. Administration dashboard refresh and updates
9. UI design refresh and updates

Authors

Khadija Souissi

Certified IT Specialist, Analytics on Z, IBM Client Center, IBM Germany Lab

souissik@de.ibm.com

Sarah Assmann

Solution Sales Professional, Analytics & Cognitive Solutions on System z

sarah.assmann@de.ibm.com

Benedict Holste

Client Technical Professional, Analytics Software on IBM Z

benedict.holste@de.ibm.com

Eberhard Hechler

Executive Architect, Member IBM Academy of Technology Leadership Team

ehechler@de.ibm.com

IBM Corporation
Software Group (or appropriate division, or no division)
Route 100
Somers, NY 10589

Produced in the United States of America
May 2018

19016419GBEN-00