July 2021

Anaconda on IBM Z and IBM LinuxONE

© 2021 IBM Corporation

Abstract
A clear trend is emerging in the era of hybrid cloud: winning enterprises will likely pull ahead by scaling the value of their data with AI. For many IBM Z® and IBM LinuxONE clients, the enterprise platform often serves as the system of record for their mission-critical data and applications. Data scientists often look for open-source solutions, and we are committed to embracing and bringing open-source AI capabilities to Z and LinuxONE. This includes activities at front-end of the data science process, such as data evaluation and visualization, through deploying real-time AI decision-making at scale.

IBM and Anaconda, provider of the leading Python data science platform, have brought to general availability Anaconda for Linux on IBM Z and LinuxONE. This paper will detail how to leverage Anaconda on IBM Z & LinuxONE and provide a detailed data science walk-through using industry standard frameworks.

Introduction
AI is driving better decisions and outcomes across many industries. While initially focused on the consumer space, AI is increasingly being leveraged by medium-to-large enterprises to drive better business performance and increase efficiencies spanning use cases ranging from internal process automation, client outreach and retention, through regulatory compliance and fraud detection. AI has created a new type of profession, data science, focused on harnessing the power of large datasets to drive these outcomes.

Data scientists leverage a plethora of tools, both open-source and commercial, and expect interoperability between platforms as they train and deploy their models to production. IBM Z and LinuxONE deliver next-level data privacy, security, resiliency. Enabling the most sensitive and critical workloads through pervasive encryption, full isolation, and extremely high system availability to provide safe and continuous delivery. IBM Z and LinuxONE provide the foundation for your mission-critical AI models, enabling data scientists to train anywhere and deploy on IBM Z and LinuxONE.

Summary of AI & Analytics Capabilities on IBM Z and LinuxONE
IBM Z and LinuxONE support the full data science tool-suite, with popular open source capabilities available through Anaconda on Linux® on Z, as well as IBM offerings like IBM Cloud Pak® for Data and IBM Watson® Machine Learning on z/OS®. Data scientists can leverage industry standard frameworks to discover, visualize, build, and train their models in tools and frameworks they're used to, such as scikit-learn, PyTorch, Matplotlib and more.

For production deployments, IBM Z and LinuxONE enable clients to take advantage of co-location of key business processes and data. Without replicating data to other environments for analysis, IBM's design goal is to enable low latency inference on platform for time-sensitive workloads such as in-transaction scoring and other real-time or near-real-time workloads. One example is fraud detection; for banks and financial markets, accurate detection of fraud during the transaction can result in significant savings. IBM is developing optimizations in software and

© 2021 IBM Corporation

hardware to meet low latency requirements and to enable clients to integrate AI tightly with IBM Z data and core business applications on Z. These technologies help clients embed AI in their applications with minimal application changes. Target use cases include time sensitive applications with high transaction volumes. In these transactional use cases, a main objective is to reduce latency for faster response time, delivering inference results back to the caller at a high volume and speed. We will provide an example later in this paper illustrating how such a problem could be addressed on the IBM Z and LinuxONE platform using Anaconda packages.

To further enable this strategy, IBM is architecting solutions to enable model portability to Z and LinuxONE without requiring additional development efforts for deployment. For instance, ONNX (Open Neural Network Exchange) technology, a standard format for representing AI models, allows a data scientist to build and train a model in their framework of choice without worrying about the downstream inference implications. To enable deployment of ONNX models, IBM has provided an ONNX model compiler that is specifically optimized for IBM Z and LinuxONE. In addition to this, we are optimizing key open-source frameworks such as TensorFlow (and TensorFlow Serving) for use on IBM Z and LinuxONE. This way our clients can continue to use their preferable tools. Additional information on the AI on IBM Z and LinuxONE strategy can be found here: link.

To summarize, our mission is to enable clients to easily deploy AI workloads on IBM Z and LinuxONE to deliver faster insights while deriving more value to the business. We are enhancing IBM Z as a world-class inference platform. We aim to help clients accelerate deployment of AI on Z by investing in seamless model portability, in integration of AI into Z workloads, and in operationalizing AI with industry leading solutions such as Anaconda and IBM Cloud Pak for Data for more flexibility and choice in hybrid cloud deployments.

Anaconda Overview

IBM and Anaconda, the leading open source Python data science platform, have recently brought Anaconda availability to IBM Z and LinuxONE. Anaconda, the world's most popular open-source package distribution and management experience, is leveraged by more than 20 million people, with a focus on enabling real data science and machine learning applications. Anaconda enables enterprises to leverage industry standard data science tools and frameworks across a variety of platforms, enabling training and deployment on the environment of choice. Anaconda provides multiple levels of products suited for practitioners ranging from beginner students through enterprise teams. Additional information on their products can be found here: https://www.anaconda.com/.

Key to Anaconda's mission is creating a trusted channel for data science packages. This includes following security and provenance practices, maintaining a strict chain-of-control for source code and built artifacts, and a quality assurance team that performs exhaustive testing for each release of Anaconda and Miniconda, including all installers and packages. Commercial edition provides Anaconda-curated packages and metadata, as well as Conda signature verification and guaranteed up-time support.

© 2021 IBM Corporation

This focus on security and trust aligns well with IBM Z and LinuxONE's mission to enable AI for critical and sensitive workloads.

The rest of this paper will provide an overview on how to install Anaconda on IBM Z and LinuxONE, and a data science example using Python and Conda.

How to use Anaconda on IBM Z and LinuxONE
Anaconda system requirements on IBM Z and LinuxONE:

1. What are my deployment options?
   Deployment options:
   - Linux VM on IBM Z and LinuxONE - https://www.ibm.com/it-infrastructure/z/zvm
   - z/OS Container Extensions (zCX) on IBM Z - https://www.ibm.com/support/z-content-solutions/container-extensions/
   - IBM Hyper Protect Virtual Servers on IBM Z and LinuxONE - https://www.ibm.com/products/hyper-protect-virtual-servers
   - IBM Cloud Hyper Protect Virtual Servers - https://www.ibm.com/cloud/hyper-protect-virtual-servers

2. What are the System Hardware requirements to leverage Anaconda on Z?
   IBM z14® and equivalent LinuxONE, and above are supported. Linux VM minimum configuration should be 2 vCPU, 4 GB RAM, 50 GB disk.

3. What are the Operating System requirements? Do I need RHEL or OpenShift®?
   No, Anaconda can run on any Linux flavor. Currently it is certified on Ubuntu and RHEL.

4. Where can I find installation instructions?
   Installation Instructions can be found here:
   https://docs.anaconda.com/anaconda/install/linux-s390x/

Anaconda in a Container
The deployment options outlined above include using Anaconda in a containerized configuration. Container images must be built for the s390x architecture using any of the deployment options detailed above. s390x images may be deployed to any of these platforms using an OCI-compliant command line interface, like the Docker CLI. Container images may be built on any of the supported IBM Z Linux platform flavors, and deployed to any of them as well. All environments provide the Docker Command Line Interface (CLI) to manage images, while Podman can be used on Linux on Z and LinuxONE.

© 2021 IBM Corporation

Application developers can build container images for Anaconda on IBM Z or LinuxONE and then push them to an image registry or save them to a tar file. Docker/Podman administrators can deploy these images to a target IBM Z or LinuxONE environment by pulling the image from the registry or loading from a tar file.

A simple Dockerfile recipe for an Anaconda container image is available at the Ambitus open source project (https://github.com/ambitus/linux-containers/tree/master/examples/Anaconda).

You can use this example as-is to experiment with Anaconda or modify it to make Anaconda part of a larger application workflow.

Anaconda on IBM Z and LinuxONE Guided Example

Financial organizations around the world face challenges to accurately predict the risk of credit default. This covers all aspects of financial lending, including personal, corporate, and public finance. To intercept fraud, great interest has been placed on the usage of data science and machine learning to predict the clients most likely to default.

These lending processes mainly run on an institution's core business systems, with IBM Z and LinuxONE providing a leading platform for security, resiliency, and scalability. With Anaconda on IBM Z and LinuxONE clients can run their machine learning model co-located with their key processes, without the need to move data off platform, improving data currency while minimizing security exposures associated with transferring data.

To that end, we have created an example credit default model leveraging Python and Conda with Linux on Z. We performed data processing and the required transformations to create a Logistic Regression model using scikit-learn.

The example uses a notebook to demonstrate key steps, from dataset visualization to data preparation, training, and inferencing, all leveraging the packages in Anaconda for Linux on Z and LinuxONE. You can use the example yourself through downloading and installing Jupyter on your system.

Please find the detailed walk-through here: link

Conclusion

As you have seen, Anaconda provides the industry standard open-source data science tools for AI and machine learning initiatives on IBM Z and LinuxONE. These capabilities enable your business to infuse AI into your mission critical workloads, all within the security-rich Z environment.

If you have any questions and/or are interested in leveraging Anaconda on Z, please reach out to your local IBM sales representative or to the AI on Z team at aionz@us.ibm.com.

© 2021 IBM Corporation

(C)Copyright IBM Corporation 2021
IBM Corporation
New Orchard Road
Armonk, NY 10504
U.S.A.
08/21

IBM, ibm.com, the IBM logo, IBM Cloud Pak, IBM Watson, IBM Z, z14 and z/OS are trademarks or registered trademarks of the International Business Machines Corporation.

A current list of IBM trademarks is available on the Web at https://www.ibm.com/legal/us/en/copytrade.shtml, and select third party trademarks that might be referenced in this document is available at https://www.ibm.com/legal/us/en/copytrade.shtml#section_4.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

InfiniBand and InfiniBand Trade Association are registered trademarks of the InfiniBand Trade Association.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the OpenStack website.

Red Hat®, JBoss®, OpenShift®, Fedora®, Hibernate®, Ansible®, CloudForms®, RHCA®, RHCE®, RHCSA®, Ceph®, and Gluster® are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

RStudio®, the RStudio logo and Shiny® are registered trademarks of RStudio, Inc.

TEALEAF is a registered trademark of Tealeaf, an IBM Company.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Worklight is a trademark or registered trademark of Worklight, an IBM Company.

Zowe™, the Zowe™ logo and the Open Mainframe Project™ are trademarks of The Linux Foundation.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this documentation is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this documentation, it is provided "as is" without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this documentation or any other documentation. Nothing contained in this documentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM (or its suppliers or licensors), or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

References in these materials to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in these materials may change at any time at IBM's sole discretion based on market opportunities or other factors and are not intended to be a commitment to future product or feature availability in any way.

© 2021 IBM Corporation