December 2020

Get more for less with on-platform analytics

© 2020 IBM Corporation

Increase your analytics insights

Using analytics to harvest key insights and business value can be a major differentiator for enterprises. However, various factors, such as the location of data, data access costs, and analytics solution capabilities, can inhibit lines of business from fully leveraging analytics for their data. Many enterprises store essential data in their IBM Z® environment and transfer data copies to a distributed server analytics solution in an attempt to taper costs.

This type of an off-platform analytics solution can minimize CPU bound analytics workloads on z/OS® but it also presents several drawbacks. Moving data from one platform to another can cause considerable extract, transform, load (ETL) charges as well as the likelihood of analyzing old data. If the data is not constantly refreshed from its system of record, the data will not be current, resulting in inaccurate analytic results. Additionally, other off-platform costs for storage, networking, floor space, power and labor are incurred to host the duplicated data on the off-platform servers.

IBM Open Data Analytics for z/OS (IzODA), an analytics solution on IBM Z, addresses the problem of moving data off-platform. IzODA is an assembly of open source and proprietary technologies that allow data scientists and application developers to analyze and visualize large volumes of data hosted on IBM Z.

IzoDA is made up of three major components, z/OS IzODA Spark, z/OS IzODA Optimized Data Layer and z/OS IzODA Anaconda. Two of the major components, Spark and Anaconda, are also two of the most popular code stacks used by the data science community. Apache Spark, an in-memory compute engine and analytics runtime where data scientists can use commonplace languages like Scala, Java™, Python, and R to run analytics. Anaconda, using Python and R, is a package and environment manager with many open source pre-built machine learning libraries and toolkits to analyze modern enterprise level volumes of data. The third component is the Optimized Data Layer (ODL)[1] which provides unique data abstraction capabilities that allow data scientists to reach key data sources on and off the mainframe through a common interface.

IzODA utilizes the Optimized Data Layer (ODL) to access existing Db2® data without adding any Db2 costs to the existing z/OS system. This is done by reading the storage directly through the use of IBM z Integrated Information Processor (zIIPs) and System Assist Processors (SAPs), without going through Db2. By accessing the data on-platform, data copies no longer need to be moved, eliminating the problem of dated/erroneous data, and avoiding off-platform infrastructure costs for re-platforming the data.

To examine how IzODA addresses data currency issues and lowers costs, we compared two analytics solutions, one using IzODA on IBM Z and another using open source Apache Spark on x86 using off-platform data that was stored in a proprietary database.

[1] Optimized Data Layer (ODL) also known as the Mainframe Data Service (MDS)

© 2020 IBM Corporation

Figure 1 shows the compared x86 solution that extracts data residing on IBM Z in an IBM Db2 z/OS database. Two 16-core Skylake x86 servers with 256 GB of memory run open source Apache Spark analytics and one 64-core Skylake x86 server with 256 GB. of memory runs a proprietary database to read data that is extracted, transformed, and loaded from IBM Z.

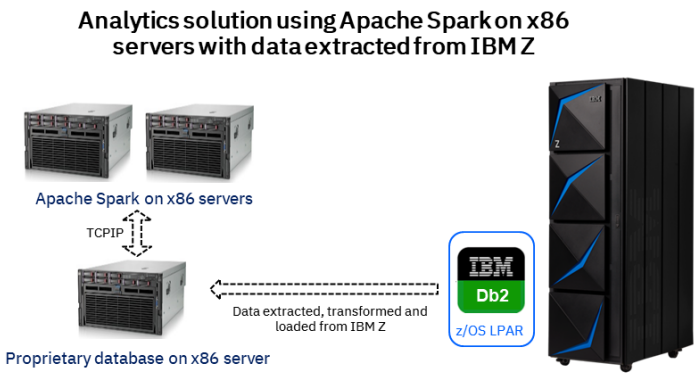

Figure 1: Overview of x86 analytics solution accessing data from IBM Z

Simplifying an analytics infrastructure

Figure 2 illustrates the IzODA solution in which the analytics engine is collocated with the data on IBM Z. IzODA runs on an IBM z15 T01 in the same z/OS LPAR as the Db2 database. Because most IzODA workload is zIIPs-eligible[2], 99% of the IzODA workload is dispatched to five zIIPs.

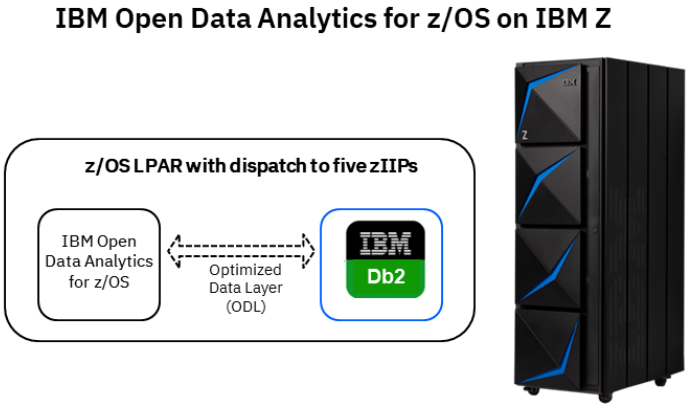

Figure 2: Overview of IBM Open Data Analytics for z/OS on IBM Z residing in same LPAR as z/OS data

With just 1% of the workload contributing to the MSU count, SW usage charges are dramatically reduced, and latency issues are avoided by accessing data on-platform. zIIPs usage in the IzODA scenario is also simple to implement since no changes are needed in IzODA to run zIIPs eligible workload. As with many other z/OS applications, zIIPs dispatching is transparent to the analytics application.

Reduce IT costs with IzODA

Just how much can IzODA save? In the modeled scenarios, IBM Open Data Analytics for z/OS delivers a 34% lower total cost of ownership than the compared x86 ETL proprietary database solution over three years.[3]

Figure 3 highlights the major cost drivers between the two modeled scenarios. In the x86 scenario software cost, comprised of proprietary database licenses, database tooling and Linux, is far the greatest cost in the x86 TCO. Even without inclusion of ETL charges or support costs for the open source Apache Spark application or other system management software, database

[2] Running the spark-sql-perf benchmark using IBM Open Data Analytics for z/OS (IzODA), 99% of the CPU consumed was zIIP eligible. IBM internal tests were run on a z15 T01 machine with 2 GCPs, 4 zIIPs and 100GB of memory. Spark-sql-perf v5.1 was used with 1 executor with 8 cores in cluster mode. IzODA 1.1.0 release level PTFs UI68206, UI68267, UI68264 was run on z/OS 2.4 with Db2 v11 and Java 8 or 6. zIIP eligibility is based on the CPU consumption of the work running on the Spark and MDS address spaces. Your results may vary.

[3] An IBM IT Economics model was used to examine one time charge and maintenance costs over three years for IBM Open Data Analytics for z/OS versus an x86 open source Apache Spark analytics solution with data extracted, transformed, and loaded from IBM Z to a proprietary database on x86. The IzODA workload scenario was comprised of five zIIPs with 256GB of memory running on an existing IBM z15 T01 using 9,300 MIPS, and IzODA that was colocated with Db2 data in a z/OS LPAR. IzODA scenario costs included purchase of zIIPs, IzODA and MLC charges for non-eligible zIIP workload (1% of total MSU count). The x86 analytics solution was comprised of two 16-core Skylake x86 servers with 256 GB of memory for Apache Spark analytics and one 64-core Skylake x86 server with 256 GB of memory for the database. x86 scenario costs included purchase of x86 servers, database software and Red Hat Linux. Both scenarios included hardware maintenance, floor space, energy, and networking costs. zIIPs pricing is based on U.S. prices as of 11/01/2020 from our website and x86 hardware pricing is based on IBM analysis of U.S. prices as of 11/01/2020 from IDC. Software, floor space, energy, and networking costs are based on data from IBM IT Economics assessments for clients. For additional information on the cost model, contact the IBM IT Economics Team at IT.Economics@us.ibm.com

© 2020 IBM Corporation

licensing and operating system costs become a significant expense due to per core and per socket pricing respectively.

Conversely, software costs for the IzODA scenario, comprised of IzODA and MLC charges for non-eligible zIIP, are significantly lower since the solution is able to leverage the existing z/OS environment without incremental DB cost. Furthermore, MLC charges are minimal, since just 1% of the workload contributes to MSU costs. The bulk, or almost all (99%), of the IzODA workload is offloaded to the zIIPs at no cost.

Another savings consideration for the IzODA scenario is that the IzODA solution is collocated in the same z/OS LPAR of the IBM Z system, avoiding incremental facilities costs. In contrast, the x86 scenario incurs incremental energy, floor space and networking costs in addition to software and hardware costs to deploy and support a new workload in the datacenter.

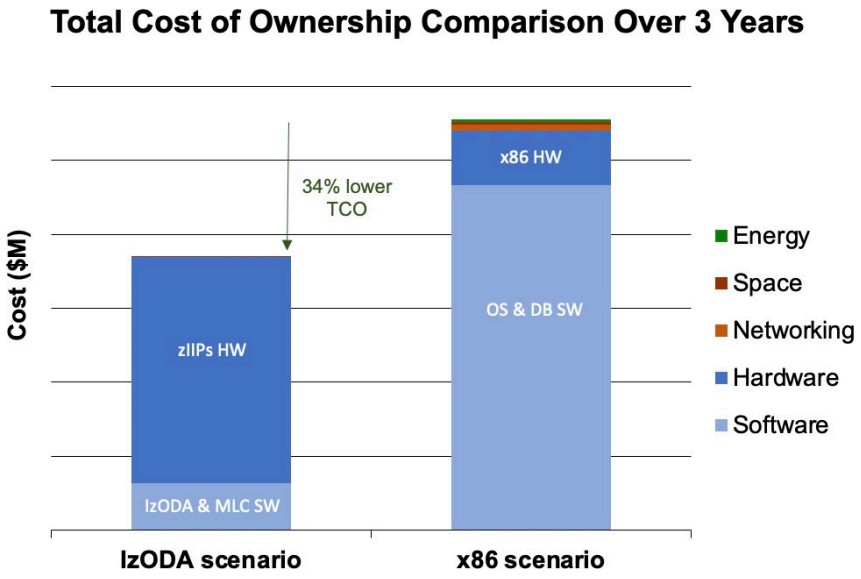

Figure 3: Three year total cost ownership for IBM Open Data Analytics for z/OS scenario versus x86 scenario with extracted data from IBM Z

Other considerations for on-platform analytics

In the compared TCO model, the IzODA solution was found to lower costs by one third (or 34%) in three years. However, other factors in IBM Z environments can potentially increase the savings delta between IBM Z and x86.

Non-Production environments: Other environments for QA, development and test were not factored into the TCO model. These non-production environments will require additional infrastructure (servers, floor space, energy, networking). Typically, IBM Z environments enable greater scaling efficiencies with resource sharing that enable workload consolidation and tiered MLC pricing advantages as workload activity grows. In most distributed server environments, however, these costs tend to be linear, causing infrastructure costs to increase at a higher rate than in a centralized server environment.

© 2020 IBM Corporation

Server administration: The TCO model excluded labor costs for server administration. Client data from IBM IT Economics assessments finds that distributed server administration effort is on average two times greater than centralized server administration.[4]

Disaster Recovery: The TCO model examined costs for production without inclusion of a corresponding disaster recovery environment. In order to replicate the production environment in an off-site location, equivalent hardware, software and facilities costs need to be added.
Qualities of service: The impact, or absence, of qualities of service such as reliability, scalability, security were not included in the TCO model. These attributes are quantifiable and can have a significant influence on the choice of a solution. If 24x7 availability with built-in redundancy and resiliency[5] or use of Hardware Security Module (HSM) Crypto Express cards at the highest level 4 of FIPS 140-210[6] are needed to meet business requirements, these capabilities should be factored into the model.

zIIP your analytics

On-platform analytics can unleash the full potential of analytics for your organization. IzoDA enables your lines of business to access data from your IBM Z system of record without latency penalties or exorbitant infrastructure costs. IzODA offers operational efficiencies and cost-savings benefits when compared to x86 solutions that extract data from IBM Z.

If your organization is examining analytic solutions, contact the IBM IT Economics team at IT.Economics@us.ibm.com for more information on IBM Open Data Analytics for z/OS and the advantages of on-platform analytics. Ask for a no-charge analytics assessment to determine the most effective infrastructure for your data and analytics.

About the authors

John Gustavson
CTO, Worldwide IBM IT Economics Consulting & Research
jgustav@us.ibm.com

Susan Proietti Conti
Program Director, Worldwide IBM IT Economics Consulting & Research
IBM Executive Project Manager, PMP®
sconti@us.ibm.com

[4] Labor observations for IT infrastructure, https://www.ibm.com/downloads/cas/7LOL5ZEG

[5] ITIC 2019 Global Server Hardware, Server OS Reliability Survey Mid-Year Update for LinuxONE and IBM Z found 0% annual unplanned server downtime of >Four Hours in 2019, https://itic-corp.com

[6] https://csrc.nist.gov/publications/detail/fips/140/2/final

© 2020 IBM Corporation

©Copyright IBM Corporation 2020
IBM Corporation
New Orchard Road
Armonk, NY 10504
U.S.A.
Produced in the United States of America,
12/2020

IBM, ibm.com, IBM logo, IBM Z, Db2, 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.

.

© 2020 IBM Corporation