IBM Z

The true cost of off-platform analytics

Organizations that rely on IBM Z often retain a wealth of enterprise data associated with their transactional systems on the platform. Using analytics to harvest key insights and business value from such data has proven to be a major differentiator for many.

Cost is a major factor when contemplating any IT investment and analytics is no exception. When faced with the decision as to whether analytical workloads, including machine learning, should be deployed on IBM Z or off-platform, there's often confusion about which cost elements should be considered.

IBM previously published a Redbook entitled *Reducing Data Movement Costs in IBM Z Environments*, which highlighted the costs associated with a daily Extract, Transform and Load (ETL) of 1 TB of data over a period of four years. The analysis highlighted the high cost of data movement to an off-platform analytics environment due to the significant CPU overhead of the ETL process.

As technology never stands still, we wanted to revisit the comparison and conduct our own calculations. As an alternative to ETL, we also wanted to evaluate CDC (Change Data Capture) as the main method for data replication, with ETL only being used for the initial transfer and load of data, and then being scheduled in such a way as not to impact on the IBM Z MLC. The numbers and projections used in this analysis are estimates and leverage tools used by the IBM IT Economics Team.

Leveraging the benefits of IBM Z

With strong security, low total cost of ownership (TCO), competitive performance and established governance mechanisms, IBM Z can be an effective keystone in an enterprise analytics solution. More often than not, organizations consider moving data from a system of record to an off-platform analytics environment with the belief that costs will be lower. Deploying analytics and machine learning on IBM Z will lead to the following benefits:

Data gravity

- Many organizations maintain vast amounts of high-value, sensitive data on IBM Z. IBM recognizes the tremendous benefits that data gravity can bring to enterprises, including reduced cost, shortened time to value and minimized security exposures, when analytical workloads are moved to where the data resides.

- The same thing cannot be said for off-platform analytics, as data quality problems are often introduced when data is replicated or in motion. This extends to the validity of data when currency is considered.

Industry-leading security

- In an increasingly intricate world of regulatory requirements and external threats, the security of client data and mission-critical workloads is paramount. Not surprisingly, security and compliance are two of the biggest concerns for many organizations today. IBM Z is already a highly secure system, and the latest IBM z14™ continues to enhance an already robust system with pervasive encryption, taking advantage of features such as the Central Processor Assist for Cryptographic Functions (CPACF) and the Crypto Express 6S cards for FIPS-4 certified encryption key management.

- Running analytics and machine learning with data on IBM Z is the safest choice to meet today's stringent compliance and security needs. The same thing cannot be said for distributed platforms, which by their very nature increase the risk of security exposure and information leakage by maintaining multiple copies of data across any number of servers.

IBM Z resiliency

- The implications of downtime can be considerable. Planned and unplanned system outages can negatively impact both customer loyalty and an organization's bottom line. IBM Z provides the highest levels of reliability, availability and security of any server platform on the market, as cited in the recent independent **ITIC 2017-2018 Global Server Hardware and Server OS Reliability Survey** which polled 800 organizations worldwide.

IBM **IT Economics Consulting & Research**

A cost comparison

Analytics

For the comparison we considered three configurations: small (S), medium (M), and large (L).

As the basis for the comparison, our starting point was IBM Open Data Analytics for z/OS®.

We assumed a medium sized analytics solution on an IBM z14[1] consisting of 5 (S), 10 (M), and 16 (L) zIIP specialty engines with 256 (S), 512 (M), and 768 (L) GB of memory.

Assuming pervasive encryption was enabled on the z14, we deducted an overhead of 2.6 percent before calculating the equivalent number of x86 cores[2] needed for our off-platform analytics environment.

We assumed an average 45 percent utilization for the x86 servers, including a 10 percent overhead for x86 platform encryption, which yielded a requirement of 117 (S), 223 (M), and 373 (L) workload cores on 3 (S), 5 (M), and 8 (L) 48-way x86 servers.

[1] IBM z14 5.2GHz (zIIPs and memory are priced as microcode upgrades only)
[2] Intel Xeon Platinum 8168 24-Core 2.7GHz (2ch/48co)

However, as IBM Z provides mission-critical reliability by design, we included an additional x86 server (n+1) to account for the failure of a single server in our x86 off-platform analytics cluster. This assumes that only one server and not the entire x86 cluster would fail at any one time, bringing the total x86 server count to 4 (S), 6 (M), and 9 (L). For comparison purposes, we assumed a commercial open-source vendor support.

Configuration Details	Small (S)	Medium (M)	Large (L)
zIIP specialty engines	5	10	16
GB of memory on IBM Z	256	512	768
x86 workload cores	117	223	373
48-way x86 servers	3	5	8
x86 servers for high availability (N+1)	4	6	9

Data Replication

For the comparison we considered three configurations: small (S), medium (M), and large (L).

To facilitate data replication between IBM Z and our off-platform analytics environment, we assumed that IBM InfoSphere® Data Replication for Db2® for z/OS would be installed on our z14 and that IBM InfoSphere Data Replication, IBM InfoSphere DataStage®, and IBM Db2 Enterprise Server Edition would be deployed in an n-tiered server architecture, each on separate x86 24-way servers with Db2 LUW being fully redundant.

Using the metrics published in the whitepaper entitled ***IBM InfoSphere Data Replication's Change Data Capture Version 10.2 (Db2 for z/OS) Performance Comparison to Version 6.5***, we estimated that a 1 (S), 2 (M), and 3 (L) TB a day transfer, sustained at 0.09 (S), 0.18 (M), and 0.28 (L) Gbps, would result in 66 (S), 133 (M), and 199 (L) MIPS usage on IBM Z and require 4 (S), 6 (M), and 10 (L) x86 cores[1] on the target IBM InfoSphere Data Replication server.

[1] Intel Xeon Platinum 8168 24-Core 2.7GHz (1ch/24co)

It is important to recognize that CDC interrogates the *Db2 for z/OS* log files to detect changes rather than querying the database directly. As a result, there is minimal processing impact on the actual database compared to the traditional approach of ETL. IIDR for z/OS was collocated with a typical IBM Z software stack consisting of CICS®, MQ® and Db2.

Performance Details	Small (S)	Medium (M)	Large (L)
TB a day transfer	1	2	3
GB per second transfer (Gbps)	0.09	0.18	0.28
MIPS usage	66	133	199
x86 cores on IBM InfoSphere Data Replication server	4	6	10

Security

For the comparison we considered three configurations: small (S), medium (M), and large (L).

One might think that our comparison would now be complete. However, as IBM Z delivers unparalleled security, being the most secure commercially available platform in the industry, it was necessary to consider security for our off-platform analytics environment, to achieve a like-for-like comparison.

To best assimilate IBM Z pervasive encryption, which leverages the Central Processor Assist for Cryptographic Functions (CPACF), standard on every core, and the new Crypto Express 6S hardware security module (HSM) found on our IBM z14, we assumed a commercial off-the-shelf transparent encryption agent offering for each of our x86 servers along with a fully redundant commercial Data Security Manager (DSM) with an embedded HSM.

In addition, to best assimilate the RACF® (Resource Access Control Facility) and security server components found in z/OS, we assumed a deployment of IBM Security Identity and Access Assurance Enterprise Edition. This required 4 (S), 4 (M), 4 (L) additional 24-way x86 servers to accommodate various components and was licensed for 200 (S), 400 (M), 600 (L) users accordingly.

Server and licensing details for security	Small (S)	Medium (M)	Large (L)
Additional 24-way x86 servers	4	4	4
Licensed users	200	400	600

A cost comparison

Manpower

For the comparison we considered three configurations: small (S), medium (M), and large (L).

A total of 13,694 (S), 13,694 (M), 13,694 (L) FTE hours were included for on-platform analytics, as opposed to 33,517 (S), 33,858 (M), and 41,356 (L) FTE hours for off-platform analytics. Off-platform analytics attracted significantly more labor overhead due to the effort required to architect, install, configure and maintain multiple software components across multiple servers.

We did not include any direct labor for data engineers or data scientists for either case, as we considered that these headcounts would be dictated by the business and in theory would be very similar across platforms

Manpower full time equivalent (FTE) hours	Small (S)	Medium (M)	Large (L)
FTE hours for on-platform analytics	13,694	13,694	13,694
FTE hours for off-platform analytics	33,517	33,858	41,356

Results

For the comparison we considered three configurations: small (S), medium (M), and large (L).

The estimated 5-year cost for IBM Z Analytics is $2.10M (S), $2.90M (M), $3.96M (L) compared to $5.37M (S), $6.50M (M), $8.75M (L) for off-platform analytics.

This amounts to a cost avoidance of over $3.27M (S), $3.59M (M), $4.79M (L), or 156% (S), 124% (M), 121% (L), for IBM Z Analytics over five years.

Off-platform analytics attracts significant costs for data replication [$2.23M (S), $3.02M (M), $4.45M (L)] that is not required with IBM Z, and security [$1.51M (S), $1.59M (M), $1.85M (L)] that is an integral part of IBM Z.

5-Year TCO Comparison

Legend:
- z Analytics + Security
- "Off Platform" Analytics
- "Off Platform" Data Replication
- "Off Platform" Security

Y-axis: Cost (Millions) USD (0.0 to 10.0)

Z (Small)	x86 (Small)	Z (Medium)	x86 (Medium)	Z (Large)	x86 (Large)
US$ 2.10M	US$ 5.37M	US$ 2.90M	US$ 6.50M	US$ 3.96M	US$ 8.75M

156% Higher · 124% Higher · 121% Higher
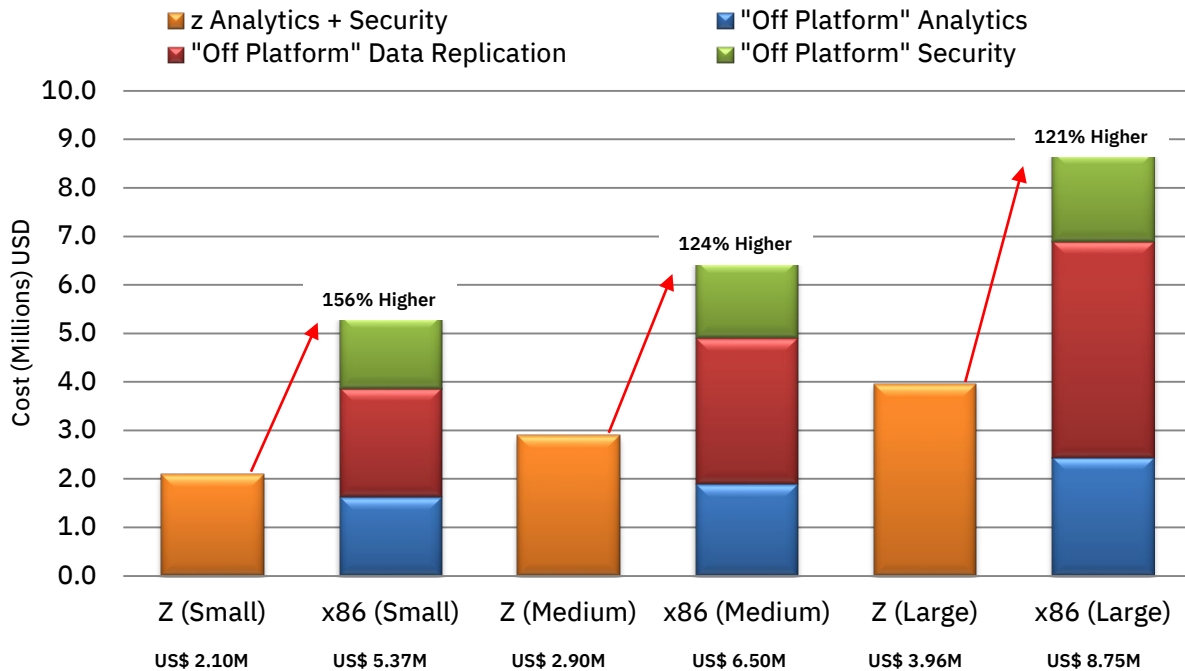
Conclusion

When evaluating IBM Z Analytics against off-platform analytics, it's important to include the cost of data replication and security in addition to any analytical software.

IBM Z is the only platform that offers pervasive encryption and delivers a robust security model that provides access control and auditing functionality built into the OS, out of the box.

Attempting to replicate these capabilities in an x86-based environment quickly becomes complex and expensive.

About the author

James Roca is an Executive IT Economics Consultant for the worldwide IBM IT Economics team.

James partners with IBM client CIOs / CTOs and their executive leadership teams to identify, evaluate, and define major enterprise-wide digital transformation programs that deliver tangible and long lasting business value.

During his career with IBM, James has held a number of senior technical and architectural leadership positions, which have included working at the IBM Hursley UK Development Lab, the IBM China Development Lab, in a regional capacity with the IBM ASEAN Growth Market Unit, and twice with the IBM Corporation in the United States.

For further information contact the IBM IT Economics team at IT.Economics@us.ibm.com