Reducing Data Movement Costs in z Systems Environments
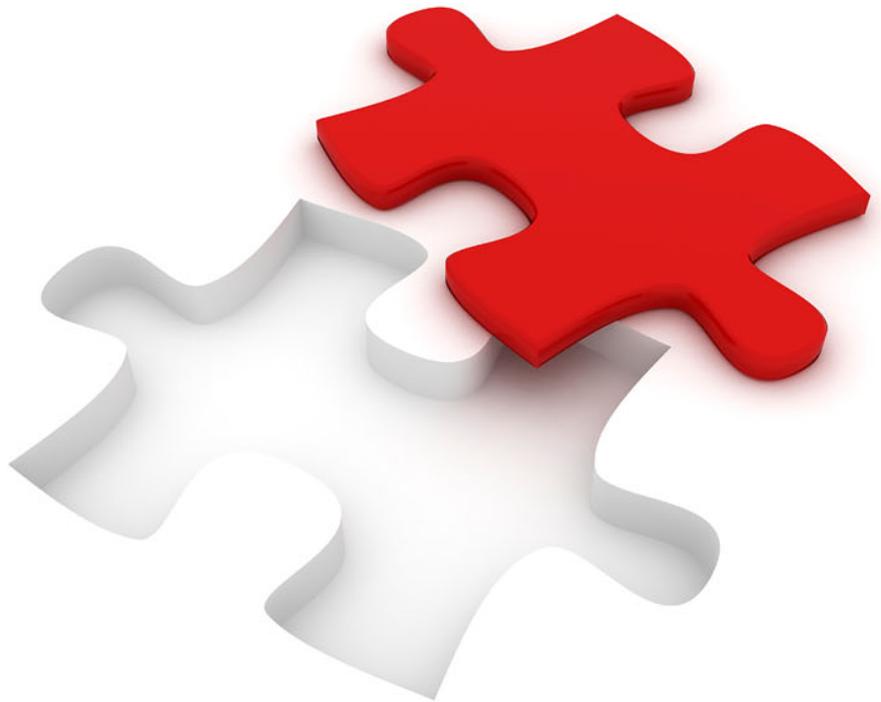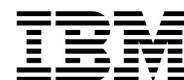
Avijit Chatterjee, Ph. D.
IBM Senior Technical Staff Member

Analytics

z Systems

IBM

Point-of-View

IBM

Moving data costs companies millions

The IBM z Systems platform is the business world's preferred system of record, responsible for more than 70% of the operational data generated by banks, retailers, and other large enterprises around the globe. And each day, much of that data is moved to separate analytics environments to gain new business insights.

Moving data costs money. The additional MIPS that are used by operational systems contribute to software, hardware, storage, and labor costs that can total millions of dollars.

Companies mostly overlook these costs and continue to analyze their operational data, such as online transaction processing (OLTP) records, by copying it to distributed servers. One oft-quoted anecdote states that for every OLTP database, there are seven or more copies being maintained in other locations for analytics or other purposes

No one questions the value of analyzing OLTP data. You cannot overstate the benefit of finding hidden trends in sales reports, insurance claims, and so on. But at what cost?

Conventional thinking has been that off-platform analytics is the best option, and that data movement expenses are insignificant. But conventional thinking is wrong. The IBM zEnterprise® Analytics System 9700 solution lets you analyze data as rapidly as ever while dramatically reducing the expense of moving it.

Quantifying data movement costs

A typical IBM mainframe customer moves multiple terabytes of OLTP data from z Systems to distributed servers every day. A recent IBM analysis found that this activity, often called extract, transform, and load (ETL), can consume 16% to 18% of a customer's total MIPS. For some, the figure approaches 30%.

Table 1 shows the results of the data movement study, which focused on two large banking customers in Europe and Asia that each routinely moved their OLTP data off-platform for analysis.

Table 1   MIPS and core consumption, IBM data movement study

	Distributed core consumption	Total MIPS consumption
European bank	28%	16%
Asian bank	8%	18%

Therefore, moving data to a separate analytical platform clearly consumes a lot of resources. But what does this mean in terms of dollars and cents?

To quantify the cost of this intensive ETL activity, IBM conducted a separate, laboratory-based study that resembled the way that the example banks moved their data from their z Systems environment and onto an x86 server (in this case, a pre-integrated competitor V4 eighth unit single database node). A four-year amortization schedule was used to spread out the cost of the system (hardware, software, maintenance, and support), along with network, storage, and labor expenses.

Conventional thinking has been that off-platform analytics is the best option, and that data movement expenses are insignificant. But conventional thinking is wrong. The IBM zEnterprise Analytics System 9700 solution lets you analyze data as rapidly as ever while dramatically reducing the cost of moving it.

The result was a unit cost per GB or ETL job to move data off of the z Systems platform. These metrics were used to compute the cost of moving 1 TB of data each day using a simple z Systems software stack, including the IBM z/OS® operating system, IBM DB2® for z/OS, and various DB2 tools. The data would be moved from an IBM z13 to an operation data store (ODS) and then on to three data marts.

As shown in Figure 1, the study projected total data movement costs of more than $10 million over the four-year period. The study assumed there are four cores on the z13 running at 85% utilization and 12 cores on each of the x86 servers running at 45% utilization. In this scenario, ETL activity burned 519 MIPS and used 10 x86 cores per day.

The primary focus of this ETL study was the cost of extracting and loading data, not transforming it. So the true cost to the banks, or any company, would be substantially higher than is shown here if you added the expense of data transformation using tools such as IBM DataStage®, Ab Initio, Informatica, or others.
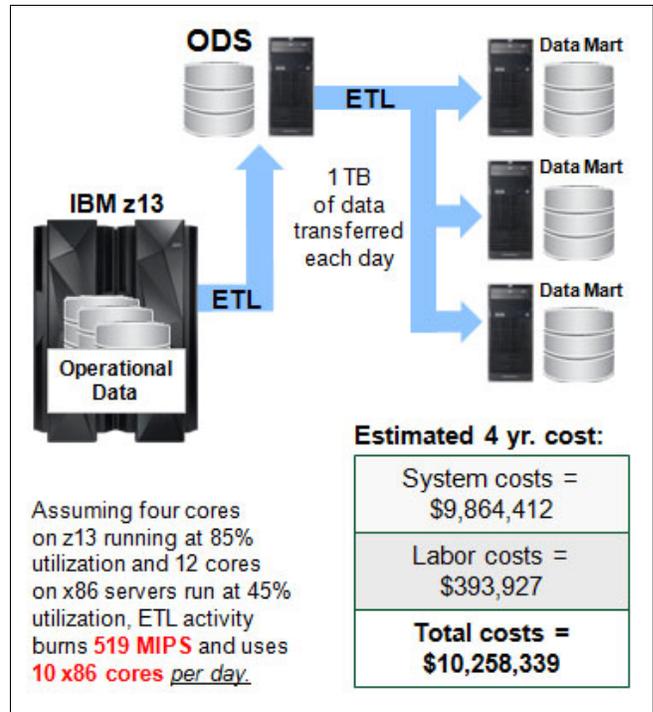

Figure 1   Four-year cost of moving 1 TB of data daily to an ODS and three data marts

Co-locate analytics with OLTP

Running analytics workloads as near as possible to where the data already are can help you avoid the high data duplication costs identified in the studies. And to achieve this collocation, IBM offers the IBM zEnterprise Analytics System 9700 solution, a hybrid analytics architecture powered by IBM Netezza® and its balanced, massively parallel processing capabilities. Your operational data remains in the z Systems environment. Analytics run within the solution using IBM DB2 Analytics Accelerator and its Advanced Loader, which copies data from the operational system to the DB2 z/OS and Analytics Accelerator simultaneously.

Analytics typically comes in two types, which makes the Analytics System 9700 solution's hybrid architecture ideal. Call center operators, for example, often rely on the results of operational analytics (high-volume, short-running, highly concurrent, OLTP-like queries) that run best on z Systems mainframes. Power users, such as statisticians and data scientists, typically perform deep analytics that involve fewer but more complex and long-running queries on multi-TB star-schema databases, which run extremely fast in DB2 Analytics Accelerator.

With the DB2 Analytics Accelerator architecture, the CPUs, field programming gate arrays (FPGAs), and hard disks are engineered to work at the same streaming rate, which helps to achieve extreme parallelism with minimal resource bottlenecks, as the CPU and I/O rates are optimally balanced. Further, DB2 Analytics Accelerator does not use indexes and is designed to run immediately upon installation, with little administration and tuning.

A recent IBM performance study[1] of an analytical workload running 161,166 concurrent reports on a 10 TB star-schema-based data mart found that throughput performance (measured in reports per hour) doubled when using the Analytics System 9700 solution compared to the same competitor system described earlier. Price performance nearly quadrupled, as measured in the cost per report per hour, dropping from $151 to just $40. Table 2 provides the pertinent numbers from the comparison.

Throughput performance... doubled when using the Analytics System 9700 solution compared to the same competitor system described earlier. Price performance... nearly quadrupled, dropping from $151 per report per hour to just $40.

Table 2   Comparison of report costs per hour

	Stand-alone pre-integrated competitor V4 (full unit)	IBM zEnterprise Analytics System 9700 solution
3-yr. total cost	$6,451,161*	$3,652,131**
Workload time	226 minutes***	105 minutes
Reports per hour	42,787	92,095 **(2.1x more)**
Cost per report per hour	$151	$40 **(3.8x fewer)**

* Cost of hardware, software, and storage based on 75% software discount and 50% hardware discount.
** Cost of hardware, software, and storage with no discount.
*** Workload time for competitor system full unit is estimated from measurements on eighth unit using perfect linearity.

The tested implementation of the Analytics System 9700 solution included a single z13 with one general purpose (GP) processor, three z Systems Integrated Information Processors (zIIPs), and full rack of DB2 Analytics Accelerator N3001-010 hardware. The DB2 Analytics Accelerator Advanced Loader software copies data from the operational system to DB2 for z/OS and the Accelerator simultaneously.

IBM conducted another test[2] to measure the cost of copying 1 TB of operational data from DB2 11 for z/OS (running on the z13) into the Analytics System 9700 solution. The solution was again configured as a full rack of Analytics Accelerator N3001-010 hardware) using the Advanced Loader. Co-locating in this way using the Analytics System 9700 solution consolidates the workloads that run on the three data marts in the competitive environment, thus reducing the amount of data that must be copied.

[1] Competitor Full Unit workload time estimated from Eighth Unit measurements assuming perfect linearity. Actual results will vary. Comparing test results of an IBM zEnterprise Analytics System 9700 with an estimated performance on competitor full unit configuration (version available as of 12/31/2014), for a materially identical 10 TB BIDAY "Fixed Execution" workload in a controlled laboratory environment. BIDAY "Fixed Execution" workload measures elapsed time for running 161,166 concurrent reports using 80 concurrent users. Intermediate and complex reports are automatically redirected to IBM DB2 Analytics Accelerator for z/OS (powered by N3001-010 hardware or Mako). Price comparison of 3YR Total Cost of Acquisition (TCA) based on US prices current as of December 31, 2014, including hardware, software, and maintenance. Used discounted pricing for competitor with 50% hardware discount and 75% software discount. Compared prices exclude applicable taxes, and are subject to change without notice. Competitor configuration: Full Unit including competitor recommended software options and features. IBM configuration: z13 platform with 1CP and 3 zIIPs with 128 GB memory and DB2 Analytics Accelerator Full Rack (N3001-10) with 7 S-blades (140 Intel E5-2680v2 2.8-GHz cores and 128 GB RAM), 2 Hosts (1 active - 1 passive) with 20 Intel E5-4650v2 2.4-GHz cores each and 12 disk enclosures, each with 24 600-GB SAS drives. Results might not be typical and will vary based on actual workload, configuration, applications, queries, and other variables in a production environment. Users of this document should verify the applicable data for their specific environment.

[2] This is based on an IBM internal study designed to replicate a typical IBM customer workload usage in the marketplace. The test involved measuring in a controlled laboratory environment elapsed time for system and administrator to extract, send, and receive a 1,118 GB file from z13 to DB2 Analytics Accelerator N3001-010 Mako Full Rack. Prices, where applicable, are based on US prices as of 12/31/2014 for both IBM and competitor. Estimated amortized cost from 4 Year TCA that includes all hardware, software (OS, DB, and tools), and 4 years of service and support. For Labor costs, used annual burdened rate of $159,600 for IT Administrator for z Systems and x86. Results might not be typical and will vary based on actual workload, configuration, applications, queries, and other variables in a production environment. Users of this document should verify the applicable data for their specific environment.

3

As shown in Figure 2, this load study projected that over a four-year period, it would cost just $1.2 million to copy 1 TB of data daily from the operational system to the Analytics System 9700 solution, compared with more than $10.2 million to regularly copy the same data in the competitor environment described earlier. This amounts to a savings of more than $9 million, or 88%.
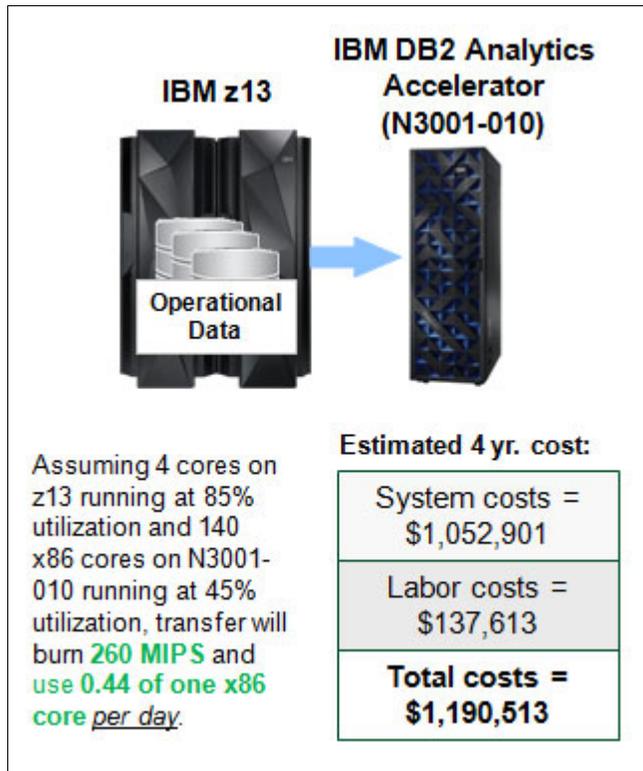

Figure 2   Four-year cost of moving 1 TB of data daily into the consolidated solution

What's next: How IBM can help

To co-locate analytics with OLTP, IBM offers the IBM zEnterprise Analytics System 9700, which is powered by IBM Netezza and its balanced, massively parallel processing architecture.

The bundled offering includes MIPS and memory that are configured on a separate logical partition (LPAR) running DB2 11 for z/OS software that is seamlessly connected to the IBM DB2 Analytics Accelerator appliance. This appliance is available as quarter-rack through a full rack or multiple full racks attached together based on data size and throughput requirements. The appliance's Advanced Loader software copies data from the operational system to DB2 for z/OS and to the Analytics Accelerator. Storage

on an appropriately sized IBM DS8870 subsystem is included, along with the necessary hardware and software, and either three- or five-year maintenance and support plans.

This amounts to a savings of more than $9 million, or 88%.

There are three primary ways in which customers can learn more about what the Analytics System 9700 can do for them:

▶ **Contact an IBM representative**

An IBM representative can perform a high-level study of your ETL costs using an internal iPad app called the IBM z Systems ETL Savings Calculator. The analysis is based on data volume, whether operational data is moved directly from z Systems to data marts or through an operational data store in between, and the complexity of the z/OS operational software stack. The app reveals the estimated cost savings of co-locating analytics with the data using the Analytics System 9700 solution rather than moving it to distributed servers.

▶ **Get an Eagle Total Cost of Ownership Study**

You can get a detailed assessment of your unique analytics environment in a no-cost Eagle Total Cost of Ownership (TCO) study. Learn more about the study here:

`ftp://public.dhe.ibm.com/software/os/systemz/pdf/Need_Help_Understanding_Your_Total_Cost_of_Ownership_-_November_27_2012.pdf`

The Eagle Team, part of the IBM Competitive Project Office, gathers input about your IT environment, data offload requirements, and associated costs. The professionals on the Eagle Team have strong IT, data architecture, and financial backgrounds that they apply to create a tailored and realistic study, quantifying your complete data movement costs.

To initiate an Eagle Total Cost of Ownership study, send email to `eagletco@us.ibm.com`.

▶ **Request a Proof-of-Concept study**

Talk to your IBM representative about evaluating the Analytics System 9700 solution at your location to test the power of co-locating analytics with OLTP. A proof-of-concept study that might take weeks or months to conduct with a competing analytical system can be performed in days using IBM DB2 Analytics Accelerator. This is thanks to its simplified

design that provides superior ready-to-use performance with minimal tuning and administration.

Resources for more information

For more information about the IBM zEnterprise Analytics System 9700 solution and the other topics described in this document, see these resources:

► IBM zEnterprise Analytics System 9700

  http://www-03.ibm.com/software/products/en/zenterprise-smartana9700

► IBM DB2 Analytics Accelerator for z/OS:

  http://www-03.ibm.com/software/products/en/db2analacceforzos

► *IBM PureData System for Analytics Architecture: A Platform for High Performance Data Warehousing and Analytics*:

  http://www.redbooks.ibm.com/abstracts/redp4725.html?Open

► IBM® IT Economics - TCO Studies (includes Eagle Team information):

  http://www.ibm.com/iteconomics/

► *An Eagle Eye* (IBM Systems Magazine):

  http://www.ibmsystemsmagmainframedigital.com/nxtbooks/ibmsystemsmag/mainframe_20100506/index.php?startid=22

Notices

IBM®

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document, REDP-5217-00, was created or updated on May 11, 2015.

Trademarks

Redbooks®

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (or), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

DataStage®
DB2®
IBM®
Redbooks (logo)
z/OS®
zEnterprise®

The following terms are trademarks of other companies:

Netezza, and N logo are trademarks or registered trademarks of IBM International Group B.V., an IBM Company.

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

Other company, product, or service names may be trademarks or service marks of others.