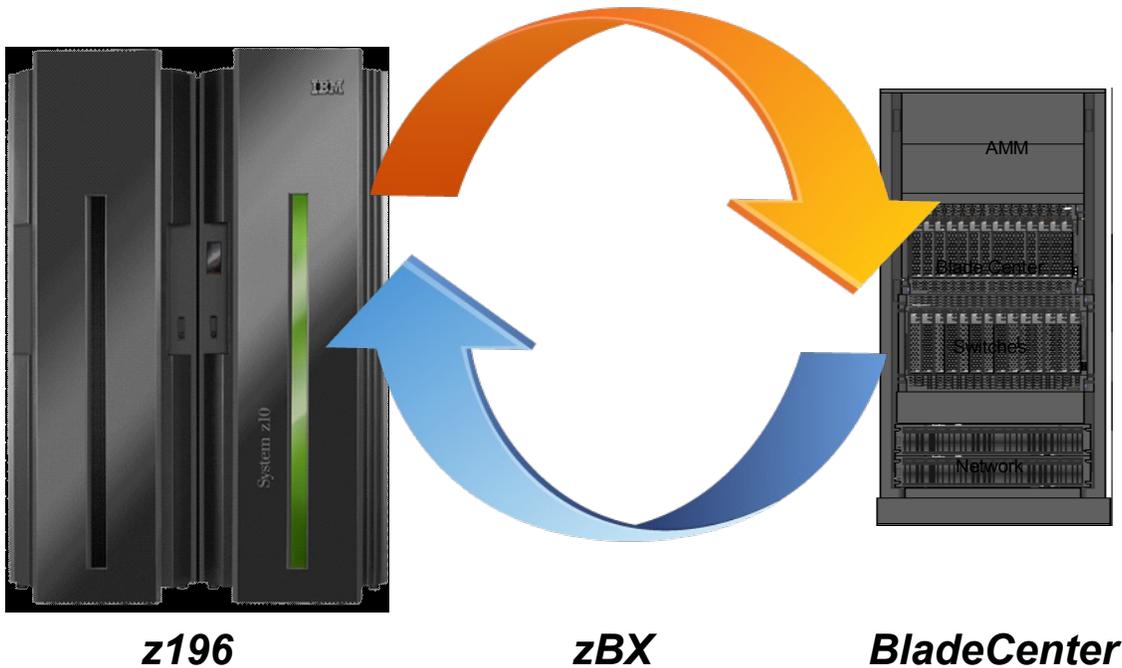


query where no query has gone before...



IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1

James Stutzman, Luis Cruz, Lisa Dodaro, Tammy McAllister

IBM Systems & Technology Group, Development Ops & Tech Support

Introduction

As a data warehouse/BI DBA, how many times in the past did you come across a request from senior management to provide business analytical information in a timely manner (like right now). Does this sound familiar?

“Jim, I need you to compute the average quantity, list price, discount, and sales price for promotional items sold in stores where the promotion is not offered by mail or a special event. Restrict the results to a specific gender, marital, and educational status. I need to have this information before my 10 AM meeting with our advertising vendor.”

Of course, in the back of your mind, you are saying, “Well, yeah, you could have asked me for this information yesterday. I could have scheduled it last night and not affect our prime time operation. I might have to create an index or two to get the results faster. Wait, do I have to reorg my tables? *Groan...*”

In today's business world where a tremendous amount of data is generated and stored in data warehouses, more and more companies are grappling with these types of business requirements. Decision makers and I/T personnel realize the huge potential value available in mining this data and the benefits of providing critical information in a timely manner. However, performing business analytics against this data often conflicts with critical real time production workloads. Service level agreements, continuous availability, and performance consideration relegate business analytics to off-shift processing, which is becoming non-existent. In addition, results from business analytics may become stale by the time it reaches the decision makers.

Now, with IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1, you can query where no queries have gone before. “IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1, together with related hardware, is a high-performance, integrated hardware/software solution for IBM System z®, delivering dramatically faster analytic query responses transparently to users. It integrates into an IBM DB2 9 for z/OS data warehouse environment, providing high-performance query software that is based on advanced data in-memory technology and executed on an IBM zEnterprise™ BladeCenter® Extension (zBX) attached to and managed by the System z server. This offering helps enable a new class of high-speed business intelligence (BI) queries never before available on System z. It extends System z qualities of service of manageability, security, and availability to analytic applications, while seamlessly combining IBM's hardware and software worlds.”¹

The IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1, software solution for IBM System z®, can be the answer to solving the stakeholders' business analytical questions in a timely manner. The IBM Smart Analytical Optimizer Studio, V1.1, is a client interface used to interact with the accelerator. In this paper, we will show the overall process to define, deploy, and load the data. We will then compare the CPU time and elapsed time of a single analytical query on DB2 for z/OS versus running it on the accelerator. We will be using a small zBX configuration. For more information on the supported configurations and their capacities, see *IBM® zEnterprise™ BladeCenter® Extension (zBX) 2458-002 Systems Assurance Product*

Figure 1 illustrates the Smart Analytics Optimizer configuration that we used.

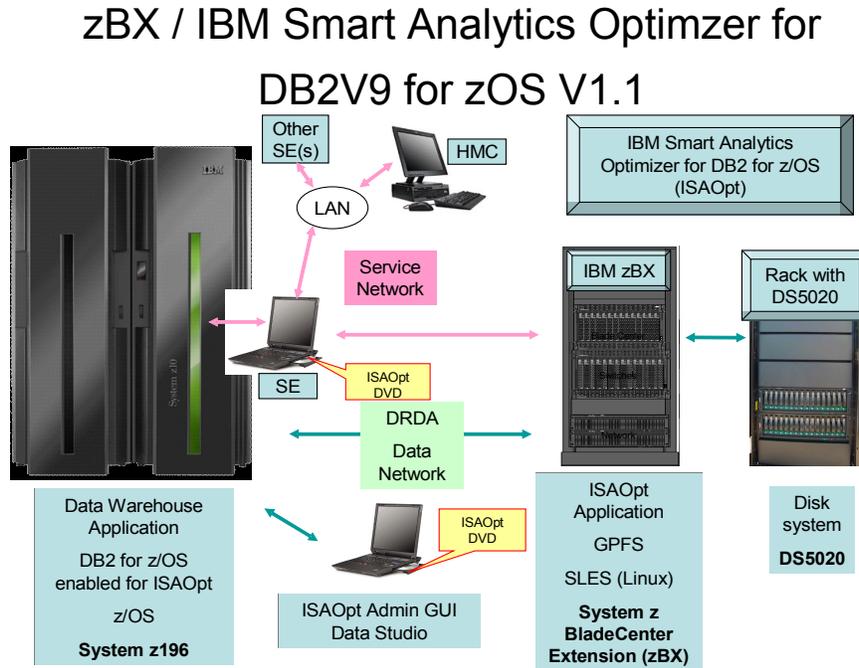


Figure 1: Overview of the zBX configuration used in this solution

The remainder of this paper discusses the following steps that we followed to implement our solution in greater detail:

1. Identifying the data warehouses to accelerate
2. Creating an accelerator project
3. Defining the data marts
4. Validating, deploying, and loading the data mart
5. Determining if the query is eligible for acceleration
6. Comparing query execution
7. Refreshing the data mart

1. Identifying the data warehouses to accelerate

With our IBM zEnterprise™ BladeCenter® Extension (zBX) up and running, we identified five of our DB2 data warehouses that would fit on our small configuration, which can handle up to 1TB of decompressed DB2 data. Table 1 lists the data warehouses and their sizes in DB2.

Table 1: Our DB2 data warehouses

DB2 data warehouse name	DB2 database size compressed(GB)
Retail	61
Message	32
Beverage	36
Bookstore1	316
Bookstore2	233

The data warehouses listed in Table 1 are used as follows:

- Retail warehouse contains information about stores, sales, inventory, and customers for a parts retailer.
- Message warehouse contains messages that were loaded from our environments OPERLOG.
- Beverage warehouse contains information such as employee, customer and inventory to represent a beverage company.
- Bookstore warehouses contain information about customers and their book purchases through an online bookstore retailer.

2. Creating an accelerator project

Now that we have identified the data warehouses that we want to accelerate, we used IBM Smart Analytics Optimizer Studio, V1.1, to create the accelerator project. As shown in Figure 2, we defined a project named PET.

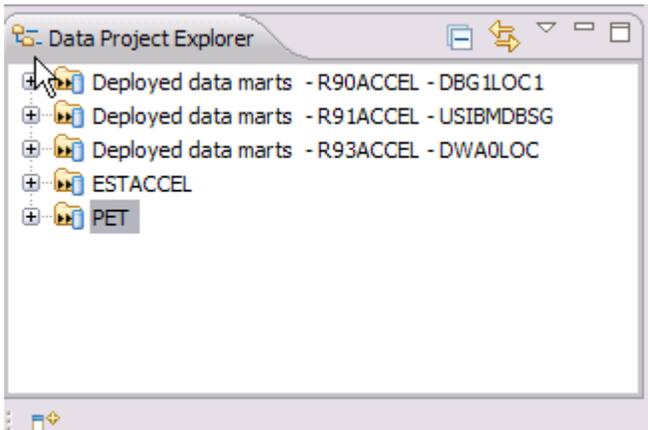


Figure 2: Data Project Explorer view

3. Defining the data marts

Within the PET project folder, we defined our data marts. A data mart is a collection of tables to be loaded to the accelerator. For the purpose of this white paper, we used our Retail and

Bookstore data marts to demonstrate the process we used to enable query acceleration.

Figure 3 shows the relationship between the tables in our Retail data mart.

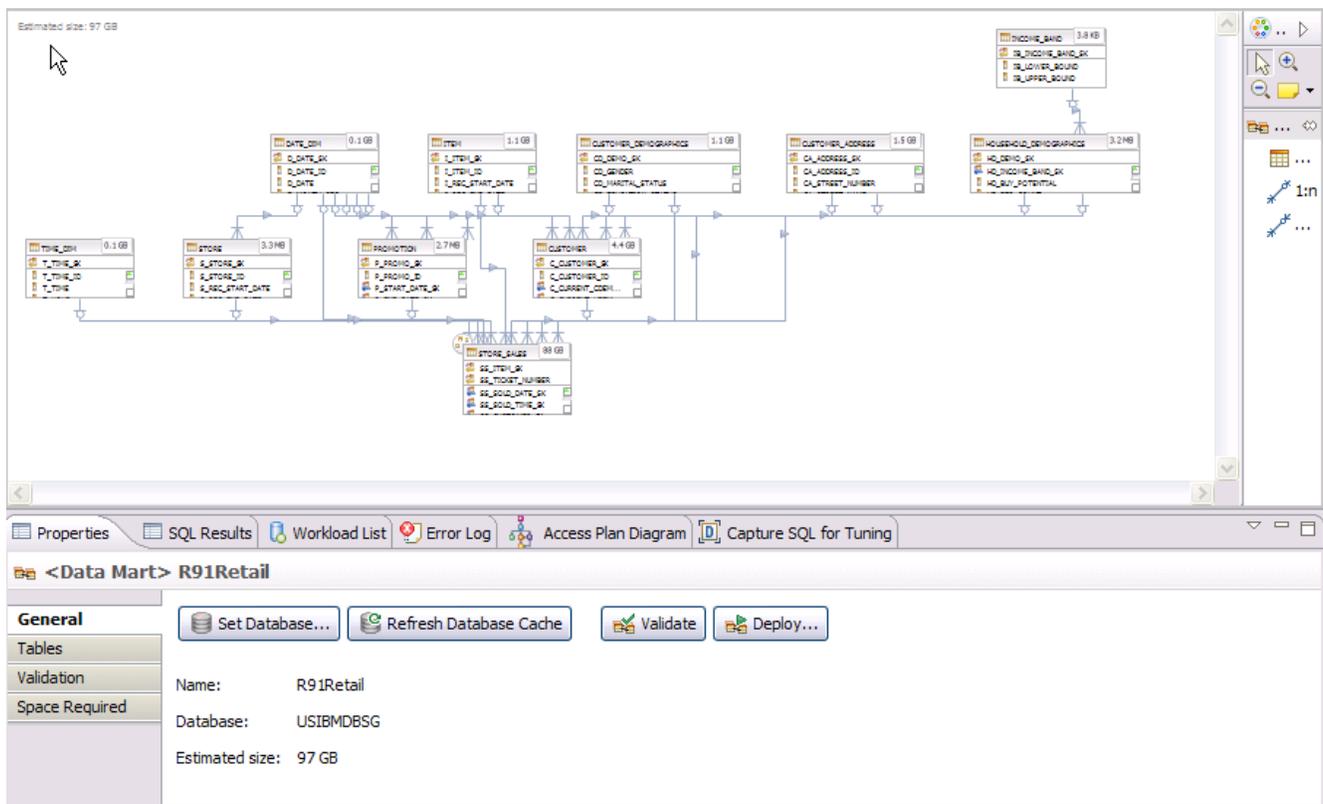


Figure 3: Graphical view of the Retail data mart

The **Estimated size** value, 97 GB as shown in Figure 3, is a good representation of how much space is required for this data mart to be loaded to the accelerator.

As shown in Figure 4, the estimated size of our Bookstore data mart, 0.3 TB, would exceed the capacity of our small configuration.

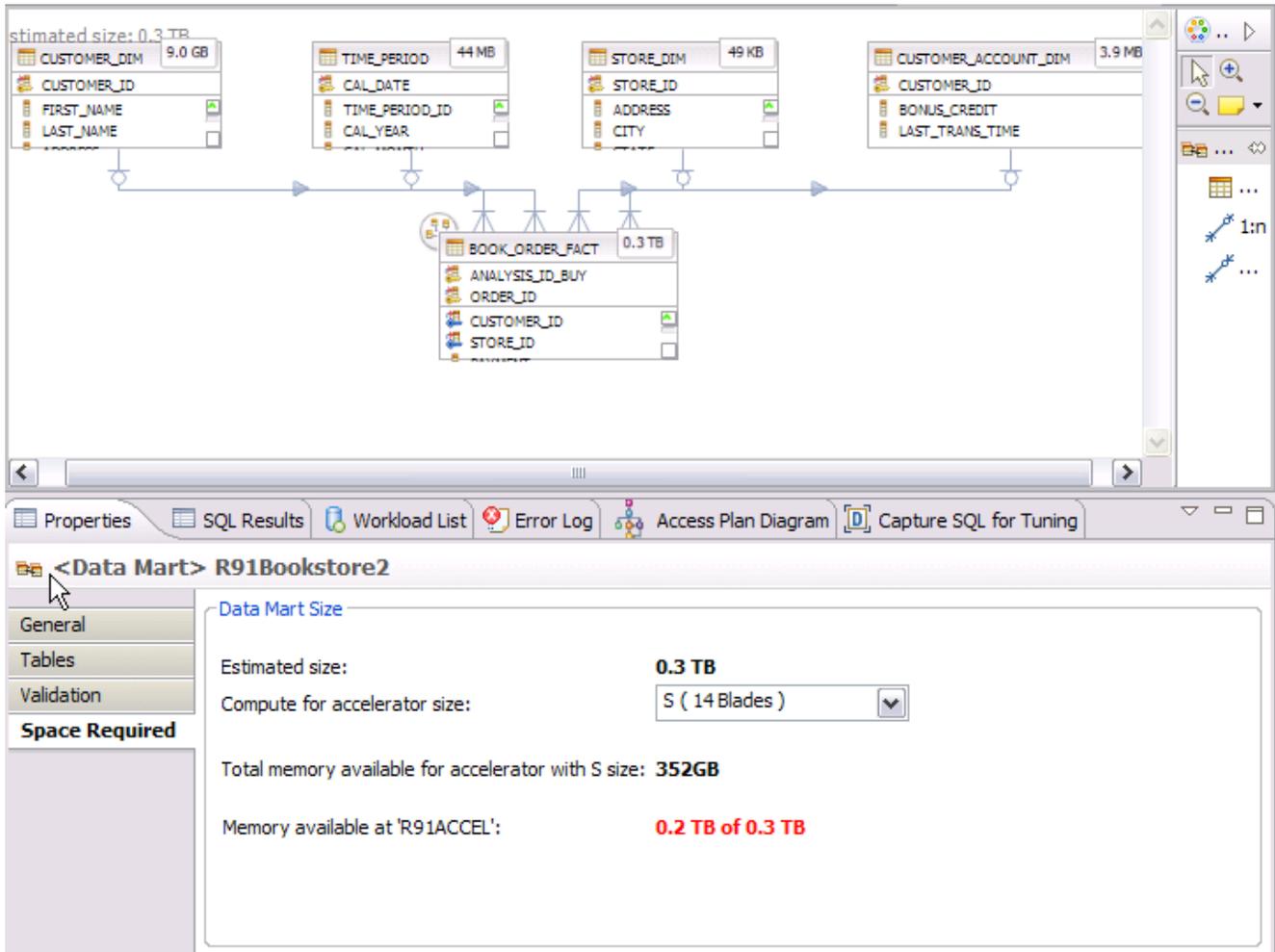


Figure 4: Space requirements for our Bookstore data mart before removing columns

To reduce the size of the data mart, we chose to select just the relevant columns that will be referenced by the queries. This allowed us to dramatically reduce the size of our data mart.

Figure 5 shows the columns that we selected to keep for our query.

Select the columns to accelerate.

Filter: type filter text

Name	Primary Key	Data Type	Length	Scale	Not Null	Estimated Size
<input type="checkbox"/> ANALYSIS_ID_BUY	Yes	BIGINT	-	-	Yes	8.2 GB
<input type="checkbox"/> BOOK_TITLE	No	CHAR	255	-	Yes	0.3 TB
<input checked="" type="checkbox"/> CUSTOMER_ID	No	INTEGER	-	-	Yes	3.7 GB
<input checked="" type="checkbox"/> LAST_TRANS_TIME	No	DATE	-	-	Yes	1.4 GB
<input type="checkbox"/> ORDER_ID	Yes	INTEGER	-	-	Yes	4.1 GB
<input checked="" type="checkbox"/> PAYMENT	No	DECIMAL	7	2	No	2.1 GB
<input type="checkbox"/> SERVER_ID	No	INTEGER	-	-	Yes	0.1 GB
<input type="checkbox"/> SHIPMENT_CHARGE	No	DECIMAL	7	2	No	0.1 GB
<input checked="" type="checkbox"/> STORE_ID	No	INTEGER	-	-	Yes	0.6 GB
<input type="checkbox"/> TAX	No	DECIMAL	7	2	No	1.7 GB
<input checked="" type="checkbox"/> TSTAMP	No	TIMEST...	-	-	Yes	10 GB

Figure 5: Bookstore data mart showing selected columns to be accelerated

After removing the columns that we did not need for our query, the estimated size of the data mart was reduced to 31 GB, as shown in Figure 6, and was well within the capacity of our small configuration.

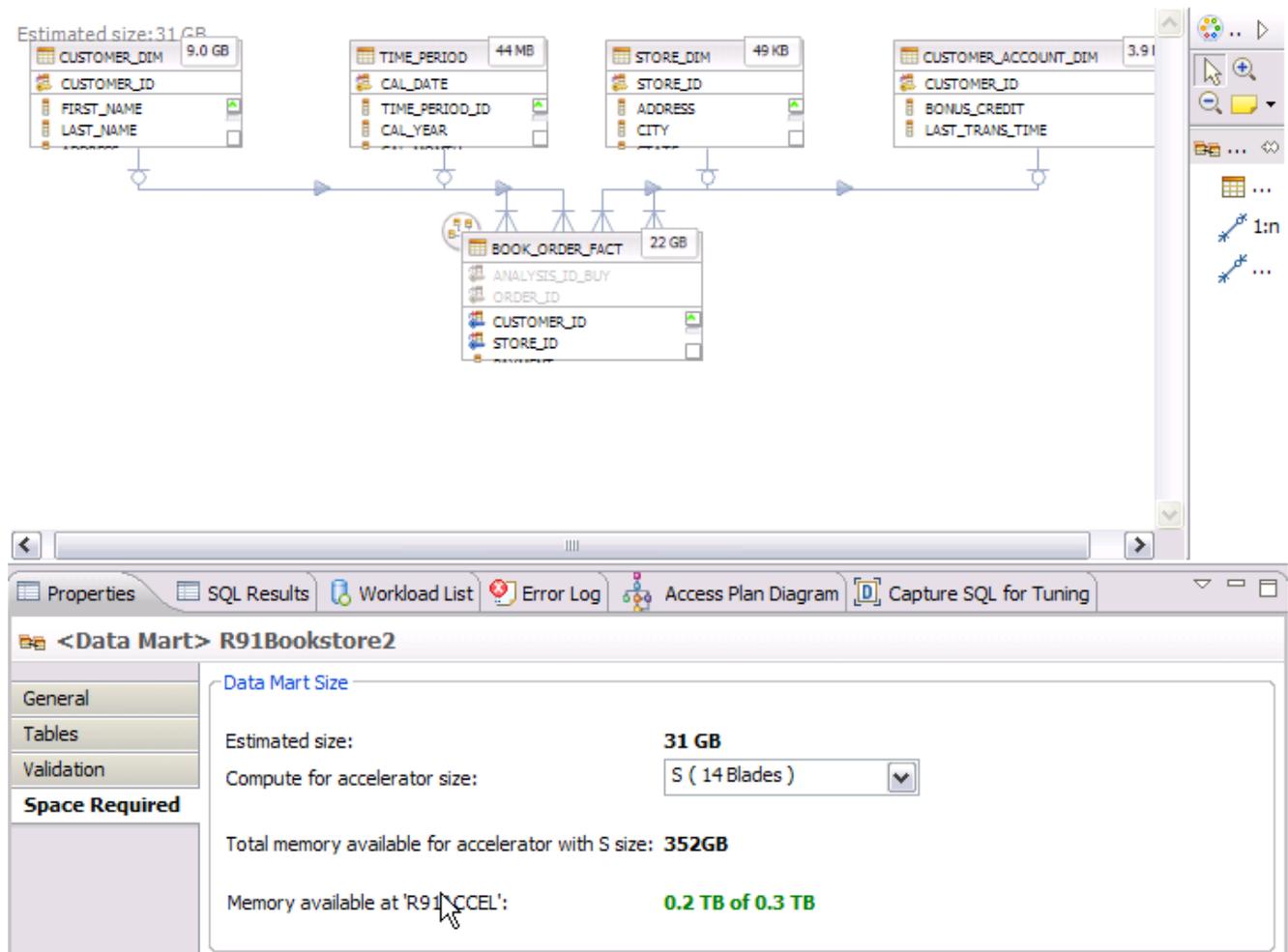


Figure 6: Space requirements of the Bookstore data mart after removing columns

4. Validating, deploying, and loading the data mart

Once our data mart definitions were complete and we had resized, as appropriate, we were ready to:

- Validate the structural integrity of the data mart.
- Deploy.
- Load over to the accelerator.

As shown in Figure 7, we first validated our Retail data mart prior to deployment.

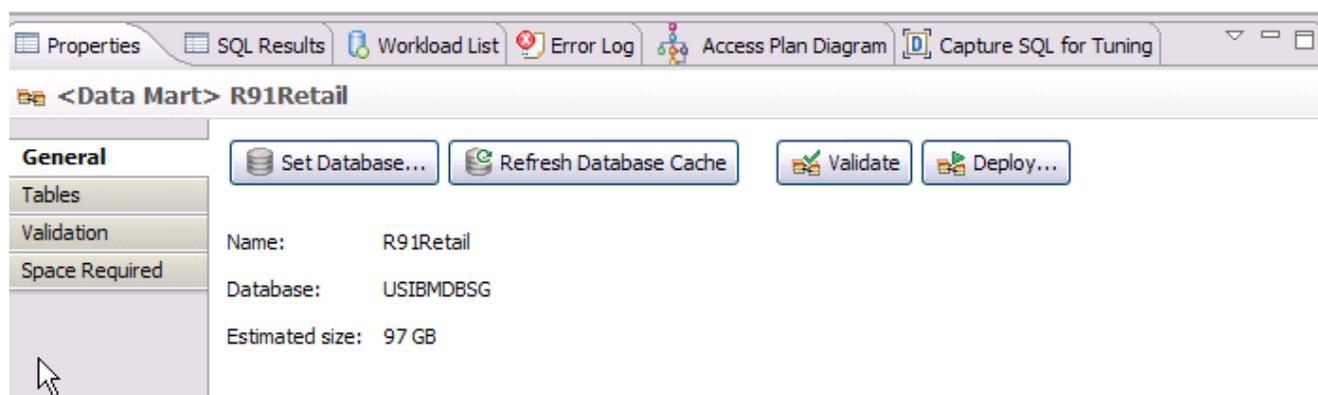


Figure 7: Validate and Deploy screen

Figure 8 shows the successful validation.

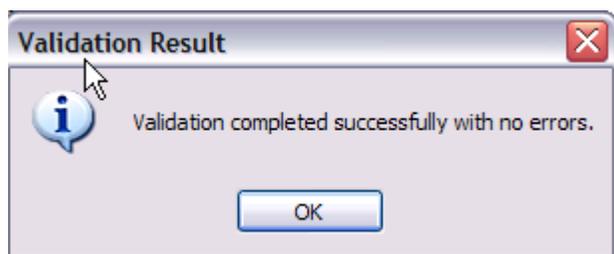


Figure 8: Validation of the data mart

Next, we moved on to the deployment and the initial data load which will be performed by selecting the **Load Data Mart** check box, as shown in Figure 9.

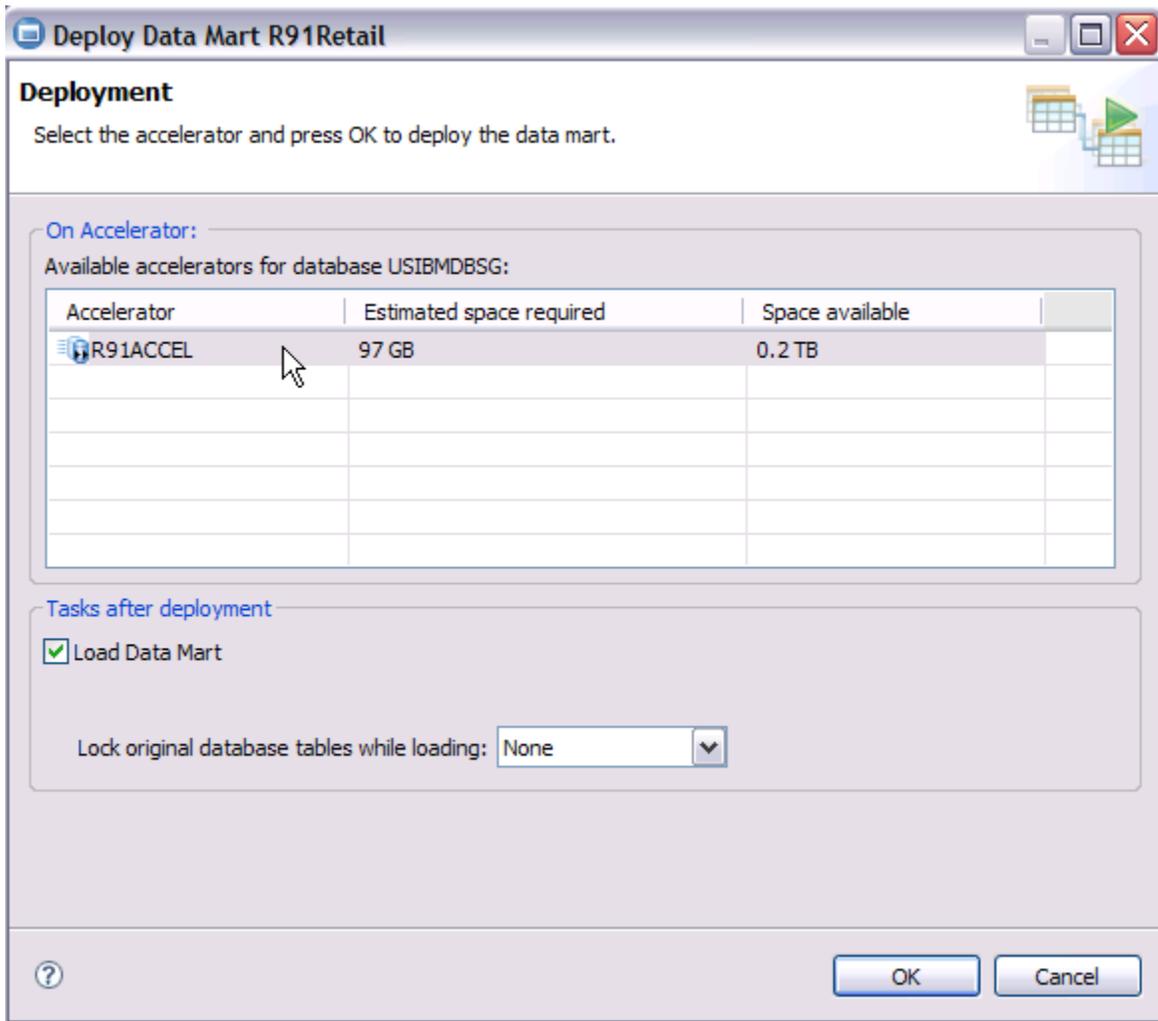


Figure 9: The deployment and load of the data mart

The load will invoke the `accel_load_mart` stored procedure in DB2 for z/OS to call the DSNUTILU utility to unload the data. The data is unloaded from the DB2 tables and loaded into memory over on the accelerator in a compressed format.

5. Determining if the query is eligible for acceleration

After the successful load, we were ready to execute our queries. We used the visual explain option to verify that the query was eligible to be accelerated. The trapezoid-shaped box shown in Figure 10 indicates that the query is a candidate for acceleration.

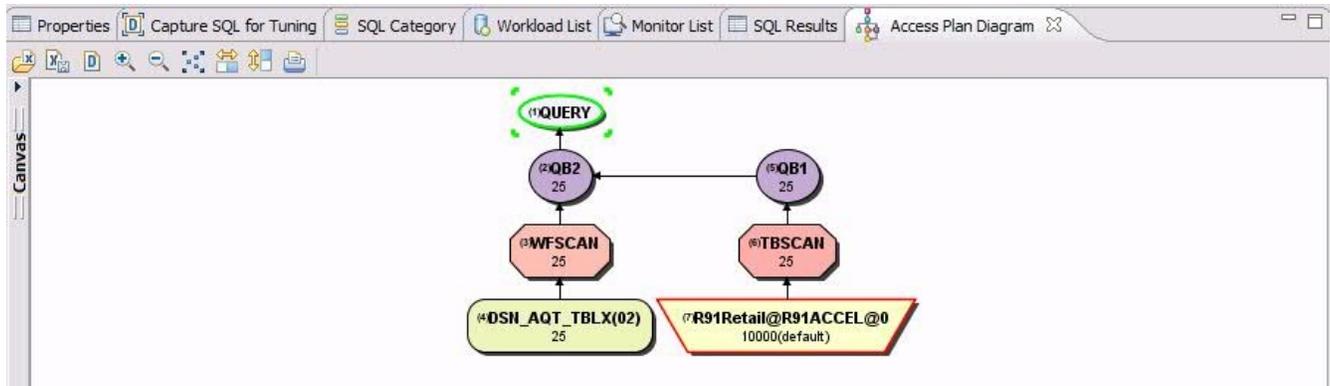


Figure 10: Access plan diagram

6. Comparing query execution

We first executed our complex query on DB2 for z/OS in order to understand the resources utilized, such as CPU utilization and elapsed time. The query we selected ran against our Retail warehouse and it is important to note that we did not tune this query for optimum performance. In addition, other DB2 resources were consumed by running this query on the System z platform that could potentially impact other business goals.

We observed the following performance timing in executing the query on System z:

```
TOTAL TCB CPU TIME=          4.88 TOTAL ELAPSED TIME= 285.5
```

We then ran the same complex query on the accelerator by adding the following SQL statement to our query:

```
set current refresh age any
```

The elapsed time and CPU utilization were within *seconds* and it produced the same result set.

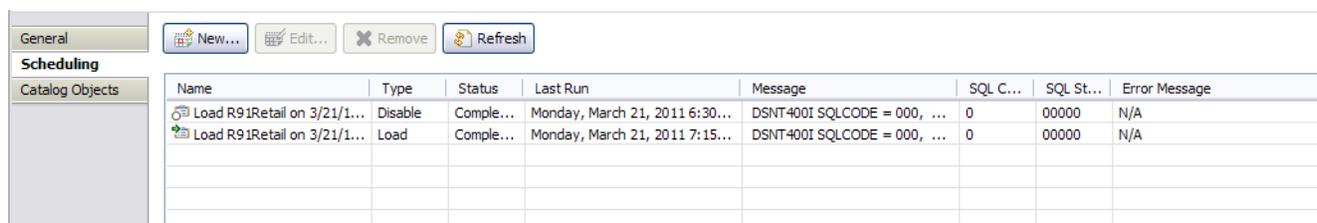
We observed the following performance in executing the same query on the accelerator:

```
TOTAL TCB CPU TIME=          .00 TOTAL ELAPSED TIME=    .09
```

Based on our test, minimal System z resources were used for this query and it did not impact other business goals. We were able to produce the result set for our business stakeholders in 0.09 minutes by running the query on the accelerator, as opposed to 285.5 minutes by running the query on DB2 for z/OS.

7. Refreshing the data mart

We scheduled a full refresh of our Retail data mart to run weekly, as shown in Figure 11.



Name	Type	Status	Last Run	Message	SQL C...	SQL St...	Error Message
Load R91Retail on 3/21/1...	Disable	Comple...	Monday, March 21, 2011 6:30...	DSNT400I SQLCODE = 000, ...	0	00000	N/A
Load R91Retail on 3/21/1...	Load	Comple...	Monday, March 21, 2011 7:15...	DSNT400I SQLCODE = 000, ...	0	00000	N/A

Figure 11: Scheduled loads

You also have the option to run scheduled loads using batch JCL. The product provides sample JCL that you can modify to disable and load your data marts. This would allow you to use your environment's scheduling software to manage your data mart refreshes. In our case, we utilized the batch JCL to schedule weekly refreshes of our Retail data mart during off-shift hours.

Conclusion

IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1, Warehouse Accelerators is an innovative approach to satisfying complex query processing. This can be accomplished without affecting critical real time production workloads. Faster response time means quicker answers, quicker insights, and a business that can adjust faster. This paper only addresses one complex query and the CPU time and elapsed time savings. With V1.1 of this product, it offers fast and predictable query response times on unpredictable query workloads, lowers overall operating costs by reducing database tuning efforts and off-loading query workloads to a platform optimized for price and performance, and dramatically reduces administration efforts by reducing the need for database tuning.

Trademarks

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. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. 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 “Copyright and trademark information” at <http://www.ibm.com/legal/us/en/copytrade.shtml>.

Other product and service names might be trademarks of IBM or other companies.

References

IBM Smart Analytics Optimizer for DB2 for z/OS enables a new class of business intelligence and data warehousing workloads for the IBM System z platform.

1. IBM United States Software Announcement 210-266 July 22, 2010
2. IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1 Quick Start Guide (GH12-6915-00)
3. IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1 Installation Guide (SH12-6916-00)
4. IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1 Getting Started (GH12-6953-00)
5. IBM Smart Analytics Optimizer Studio V1.1, User's Guide (SH12-6919-00)
6. IBM zEnterprise BladeCenter Extension (zBX) 2458-002 Systems Assurance Product Review Guide Technical and Delivery Assessment (SA10-006-01)

Disclaimer

© Copyright IBM Corporation 2011. All rights reserved. U.S. Government Users Restricted Rights -Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

THE INFORMATION CONTAINED IN THIS WHITE PAPER IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, 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 WHITE PAPER OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS WHITE PAPER 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 ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE. THE USE OF THIS INFORMATION IS A CUSTOMER RESPONSIBILITY AND YOUR EXPERIENCE WILL VARY DEPENDING ON YOUR DATA.