

Why Users Call IBM DB2 Analytics Accelerator the Magic Box

Scott Davis, Senior Systems Engineer/DBA
Blue Cross and Blue Shield of Alabama

Who we are...



BlueCross BlueShield of Alabama

An Independent Licensee of the Blue Cross and Blue Shield Association.

- 19 Brand Excellence Awards
- 6 J.D. Power Awards
- 17 FEP Plan Excellence Awards

Agenda

- Problem returning large amounts of data
- Possible solutions for the problem
- Proof of Concept process
- Observations and Issues
- Implementation
- Results

What's the issue?

Data Performance

- Business areas run ad-hoc queries against databases, but results are often slow to return or don't return at all
- Time should be spent on analyzing the data rather than gathering the data

Why are query speeds slow?

Large amounts of data being analyzed

Joining multiple tables/views to accomplish required analysis

Non-technical analysts using software to create queries/reports

WHERE

PRD.ENDDT < PRD.SNPSHT_DT GROUP BY PRD.SNPSHT_DT , PRD.UNQ_MEM_ID , PRD.CNTRCT_PFX ,
PRD.CNTRCT_ID , PRD.MEM_SEQ_NR) G1 JOIN SUBSME.CNTRCT_MEM_EFF_CNDNSD_SNPSHT G2 ON
G2.SNPSHT_DT = G1.SNPSHT_DT AND G2.UNQ_MEM_ID = G1.UNQ_MEM_ID AND G2.CNTRCT_PFX =
G1.CNTRCT_PFX AND G2.CNTRCT_ID = G1.CNTRCT_ID AND G2.MEM_SEQ_NR = G1.MEM_SEQ_NR AND
COALESCE(G2.CNTUS_COVG_PRD_NR, 0) = G1.COVGPRD GROUP BY G1.SNPSHT_DT , G2.UNQ_MEM_ID ,
G2.CNTRCT_PFX , G2.CNTRCT_ID , G2.MEM_SEQ_NR , G1.COVGPRD) G3 JOIN
SUBSME.CNTRCT_MEM_EFF_CNDNSD_SNPSHT G4 ON G4.SNPSHT_DT = G3.SNPSHT_DT AND G4.UNQ_MEM_ID =
G3.UNQ_MEM_ID AND G4.CNTRCT_PFX = G3.CNTRCT_PFX AND G4.CNTRCT_ID = G3.CNTRCT_ID AND
G4.MEM_SEQ_NR = G3.MEM_SEQ_NR AND G4.CNTUS_COVG_PRD_NR = G3.COVGPRD AND G4.EFF_TO_DT =
G3.G2_ENDDT JOIN SUBSME.UNQ_MEM_ID_SUMM_SNPSHT AE1 ON AE1.SNPSHT_DT = '04/27/2012' AND
AE1.UNQ_MEM_ID = G4.UNQ_MEM_ID AND AE1.CNTRCT_PFX = G4.CNTRCT_PFX AND AE1.CNTRCT_ID =
G4.CNTRCT_ID AND AE1.MEM_SEQ_NR = G4.MEM_SEQ_NR LEFT JOIN REF.DIV_CAN_CD DC ON G4.DIV_CAN_CD =
DC.DIV_CAN_CD AND G4.DIV_CAN_CD IS NOT NULL LEFT JOIN REF.CAN_RSN_CD CC ON
G4.CNTRCT_CAN_RSN_CD = CC.CAN_RSN_CD AND G4.CNTRCT_CAN_RSN_CD IS NOT NULL LEFT JOIN
REF.MEM_RMVL_CD MC ON G4.MEM_RMVL_CD = MC.MEM_RMVL_CD AND G4.MEM_RMVL_CD IS NOT NULL) G
LEFT JOIN REF.ACTURL_ENRL_SUBC AS1 ON AS1.ACTURL_ENRL_SUBC = G.OLDAE LEFT JOIN
REF.ACTURL_ENRL_CATGY AC1 ON AC1.ACTURL_ENRL_CATGY = AS1.ACTURL_ENRL_CATGY JOIN
SUBS.SNPSHT_DT_TBL_XREF SX ON SX.TBL_NM = 'COVD_LVS_SNPSHT' AND SX.SNPSHT_DT_1 >= G.SNPSHT_DT
AND SX.SNPSHT_DT_1 < DATE(G.SNPSHT_DT + 7 MONTHS) LEFT JOIN
SUBSME.CNTRCT_MEM_EFF_CNDNSD_SNPSHT OTH ON OTH.SNPSHT_DT = SX.SNPSHT_DT_1 AND
OTH.UNQ_MEM_ID = G.UNQ_MEM_ID AND (OTH.CNTRCT_PFX = G.CNTRCT_PFX OR OTH.CNTRCT_ID =
G.CNTRCT_ID) AND (OTH.EFF_TO_DT IS NULL OR (OTH.EFF_TO_DT > OTH.EFF_DT AND OTH.EFF_TO_DT >
G.OLDEND)) AND OTH.CNTRCT_PFX = 'WRI' LEFT JOIN SUBSME.UNQ_MEM_ID_SUMM_SNPSHT AE2 ON
AE2.SNPSHT_DT = OTH.SNPSHT_DT AND AE2.UNQ_MEM_ID = OTH.UNQ_MEM_ID AND AE2.CNTRCT_PFX =
OTH.CNTRCT_PFX AND AE2.CNTRCT_NR = OTH.CNTRCT_NR LEFT JOIN
SUBSME.CNTRCT_COVG_CNDNSD_SNPSHT CCH ON CCH.SNPSHT_DT = OTH.SNPSHT_DT AND CCH.CNTRCT_PFX
= OTH.CNTRCT_PFX AND CCH.CNTRCT_NR = OTH.CNTRCT_NR AND (CCH.EFF_TO_DT IS NULL OR
(CCH.EFF_TO_DT > CCH.EFF_DT AND CCH.EFF_TO_DT > G.OLDEND)) AND AE2.CNTRCT_PFX IS NULL AND
CCH.TYP_BUS IN ('1', '2', '4') LEFT JOIN SUBSME.CNTRCT_COVG_CNDNSD_SNPSHT CCD ON CCD.SNPSHT_DT =
OTH.SNPSHT_DT AND CCD.CNTRCT_PFX = OTH.CNTRCT_PFX AND CCD.CNTRCT_NR = OTH.CNTRCT_NR AND
(CCD.EFF_TO_DT IS NULL OR (CCD.EFF_TO_DT > CCD.EFF_DT AND CCD.EFF_TO_DT > G.OLDEND)) AND
AE2.CNTRCT_PFX IS NULL AND CCD.TYP_BUS = '7' LEFT JOIN SUBSME.CNTRCT_COVG_CNDNSD_SNPSHT CCR ON
CCR.SNPSHT_DT = OTH.SNPSHT_DT AND CCR.CNTRCT_PFX = OTH.CNTRCT_PFX AND CCR.CNTRCT_NR =
OTH.CNTRCT_NR AND (CCR.EFF_TO_DT IS NULL OR (CCR.EFF_TO_DT > CCR.EFF_DT AND CCR.EFF_TO_DT >
G.OLDEND)) AND AE2.CNTRCT_PFX IS NULL AND CCR.MAJ_LN_BUS = '6' LEFT JOIN
SUBSME.CNTRCT_COVG_CNDNSD_SNPSHT CCX ON CCX.SNPSHT_DT = OTH.SNPSHT_DT AND CCX.CNTRCT_PFX
= OTH.CNTRCT_PFX AND CCX.CNTRCT_NR = OTH.CNTRCT_NR AND (CCX.EFF_TO_DT IS NULL OR
(CCX.EFF_TO_DT > CCX.EFF_DT AND CCX.EFF_TO_DT > G.OLDEND)) AND AE2.CNTRCT_PFX IS NULL AND
CCX.TYP_BUS NOT IN ('1', '2', '4', '7') AND CCX.MAJ_LN_BUS = '6'

Initial Solutions

Network traces to check for bottlenecks or bandwidth issues

High powered workstations

Database reorgs/runstats run "as needed" based on real-time statistics

Query tuning

BCBSAL Environment

2 EC12 CPCs (1 model 705, 1 model 706)

z/OS 1.13

DB2 V10 NFM

3 main DB2 z/OS subsystems – Processing,
Warehouse and Development

Enterprise Data Warehouse utilizes 3 way data
sharing with one member hosting a majority of
the workload

Information Management

Information Management program established to manage information/data cohesively and comprehensively

'Data Performance' identified as a major area for improvement

Information Management Strategy developed with the assistance of IBM

DB2 Analytics Accelerator Solution

IBM DB2 Analytics Accelerator

- Netezza (IBM PureData for Analytics) based solution

Netezza Data Warehouse appliance connected to IBM mainframe through 10 gig dedicated fiber cable

DB2 Analytics Accelerator software on mainframe integrates with DB2 to determine which queries to accelerate

DB2 Analytics Accelerator Solution (continued)

Netezza appliance built on Massive Parallel Processing (MPP) architecture

- Several times faster than native DB2 processing

IBM DB2 Analytics Accelerator for z/OS V2.1

What it is / fit – validate agenda assumption

What is it?

The IBM DB2 Analytics Accelerator is a workload optimized, appliance add-on to a DB2 z/OS environment that services long-running, complex queries.

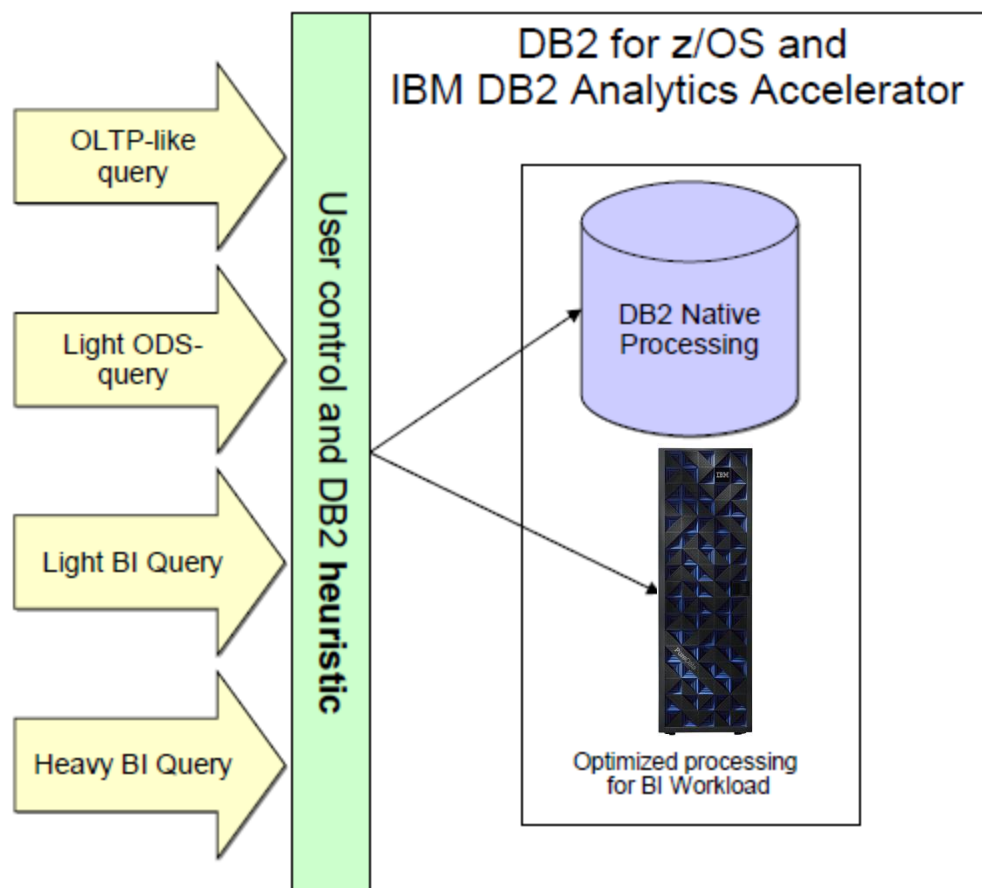
z196 or z114
OLTP and Transactional Analytics

DB2 Analytics Accelerator
Deep Analytics



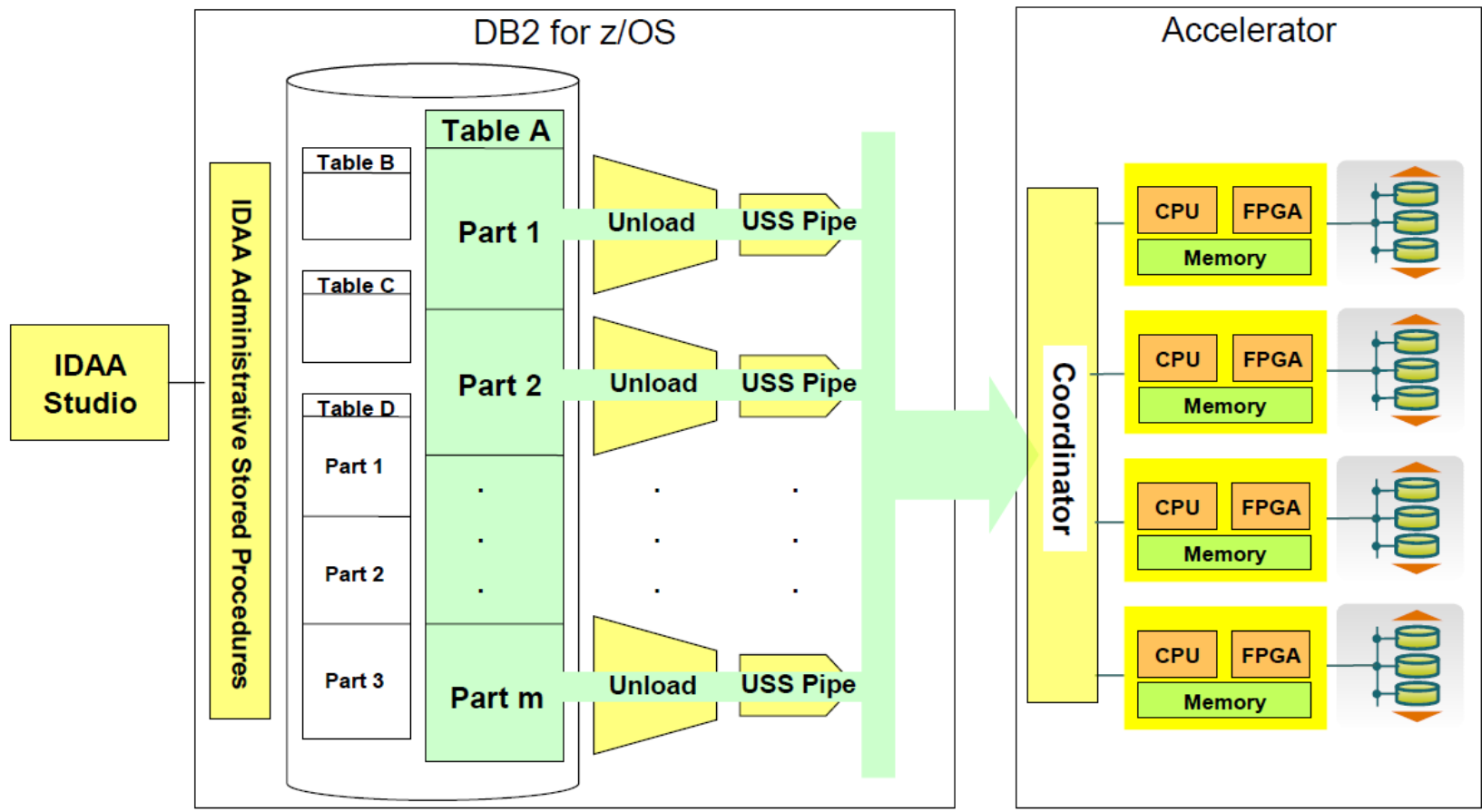
- Breakthrough technology enabling new opportunities
- Extreme performance for complex analytics (aka Train of Thought Analysis)
- Integrated with DB2 for z/OS V9 and V10 as a dedicated appliance exclusive to the System z environment
- Transparent to DB2 applications and users

Workload-Optimized Query Execution



- Single and unique system for mixed query workloads
- Dynamic decision for most efficient execution platform
- New special register QUERY ACCELERATION
 - NONE
 - ENABLE
 - ENABLE WITH FAILBACK
- New heuristic in DB2 optimizer
- **Combines the strengths of both System z and Netezza**
- **Merging operational and data warehouse into a single optimized environment**

Accelerator Data Load



- Load speed up to 1 TB/hr are common
- Trickle-feed update under development

Workload Analysis

Data Warehousing workload captured for 3 days across all 3 data sharing subsystems

Information analyzed by the IBM Germany labs

- A majority of the workload had the potential to utilize DB2 Analytics Accelerator solution
- Results presented to IT leadership, Technology Support, Application Development and Enterprise Architecture

Workload Analysis (continued)

DB2 Analytics Accelerator concept also socialized with Business Areas

Management approval to proceed with the Proof of Concept (PoC)

Site Survey

Technical feasibility of PoC

Data Center site survey

- Mainframe did not have the networking card required for making the 10 gig fiber optic connection to the Netezza appliance

Stating the Proof of Concept

IBM shipped Netezza hardware and assisted with the installation and configuration of the DB2 Analytics Accelerator software

DB2 Analytics Accelerator only connected to one of the lesser used Data Warehousing members of the data sharing environment

PoC Process

Reporting environment

- Business Objects Crystal Reports
- Business Objects Web Intelligence
- Business Objects Enterprise InfoView 3.1
- Tableau
- Toad Data Point
- AQT
- WinSQL
- SPUFI
- DB2 Connect 9.1/DB2 Connect 9.5

PoC Process (continued)

Participants included “power users” representing all major business areas

Problem queries captured

- Referenced tables identified, defined and loaded to DB2 Analytics Accelerator
- Same queries executed in both DB2 Analytics Accelerator and DB2 native environments
 - Runtime and CPU usage captured and compared

PoC Process (continued)

Various query and reporting tools used

Time needed to load data to DB2 Analytics

Accelerator also captured

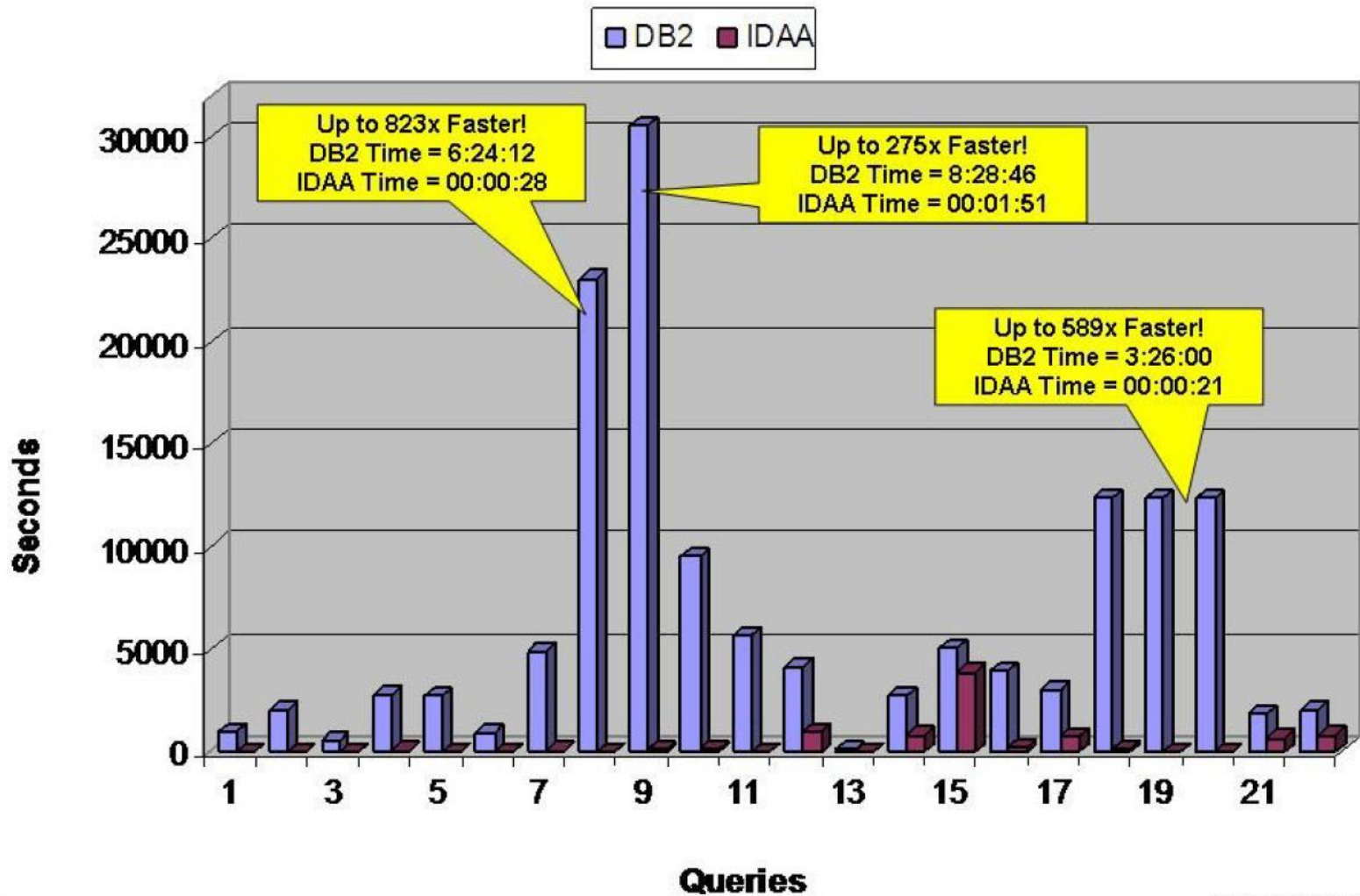
PoC Results

One problem query that took 24+ hours in native DB2 completed within 2 minutes in DB2 Analytics Accelerator

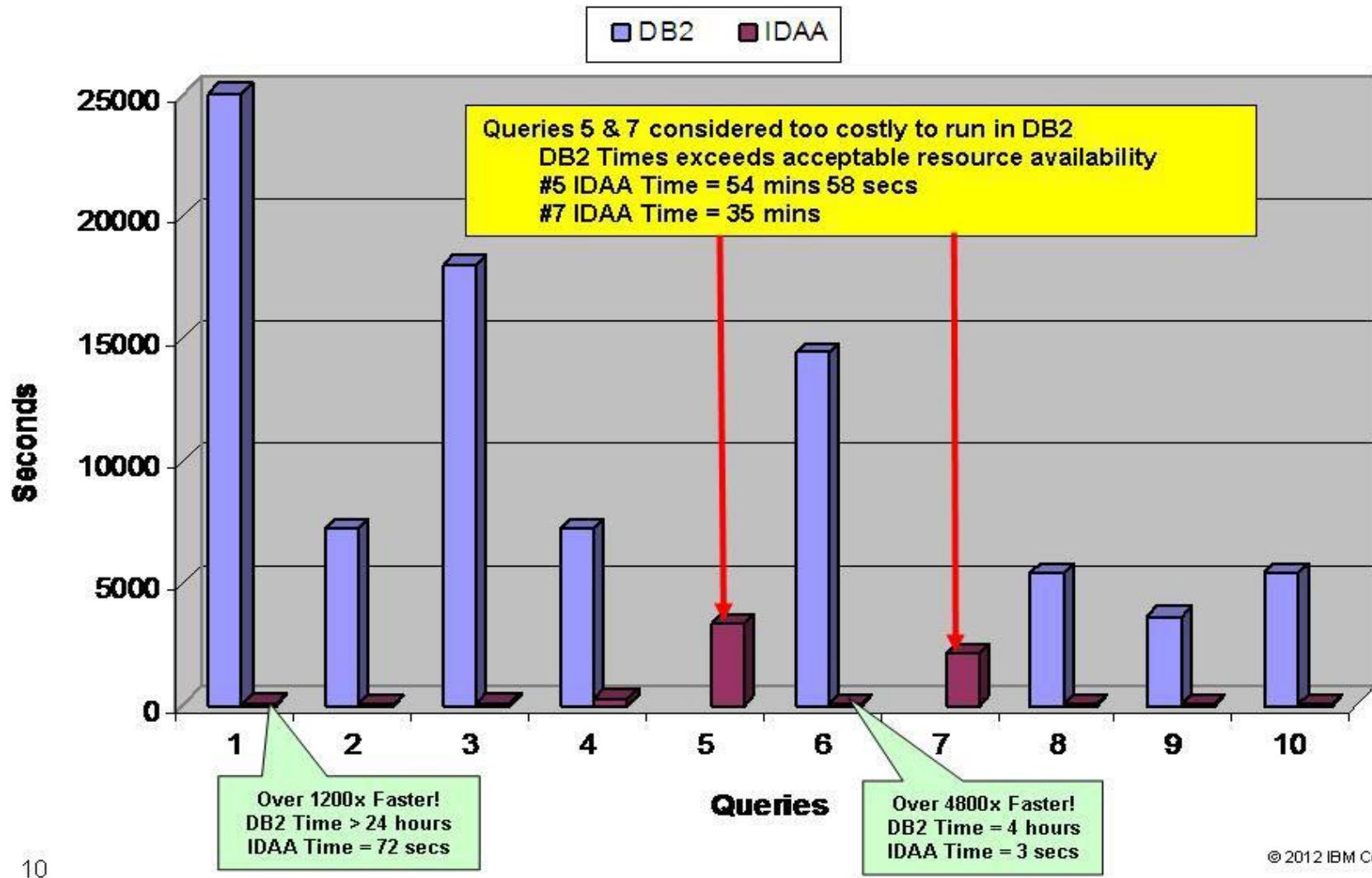
Minimal changes required by business users

- Point to correct data sharing subsystem
- Manually issue the 'Enable Acceleration' query

Comparison of Query Elapsed Time



Elapsed Time Comparison for Exception Queries



PoC Results (continued)

Magic???

- Behind the scenes work performed by Technology Support Team
 - Systems Programmers
 - DBAs
 - Data Warehousing Team
 - Networking Team
 - Data Center Team



Menu Utilities Compilers Help

BROWSE F990.COMP.CNTL(J9I79136) Line 00000014 Col 001 080
Command ==> Scroll ==> CSR

```

/** Load table_into IDAA
/*******
//LOAD0001 EXEC PGM=IKJEFT01,DYNAMNBR=20,COND=(4,LT)
/** parameter #1 for accelerator name
//AQTP1 DD DISP=SHR,DSN=F990.DB2.SQL(AQTP1)
/** parameter #2 for LOCK_MODE
//AQTP2 DD DISP=SHR,DSN=F990.DB2.SQL(LOAAQTP2)
/** parameter #3 for load containing load specification
//AQTP3 DD DISP=SHR,DSN=F990.DB2.SQL(LOAAQTPA)
// DD *
<table name="IPP_STG_PROCESS_DETAIL" schema="ITS"
// DD DISP=SHR,DSN=F990.DB2.SQL(LOAAQTPB)
/** last parameter for message input to control trace
//AQTMMSGIN DD DISP=SHR,DSN=F990.DB2.SQL(LOAMSGIN)
//SYSTSPRT DD SYSOUT=*
//SYSPRINT DD SYSOUT=*
//SYSUDUMP DD SYSOUT=*
//SYSTSIN DD DISP=SHR,DSN=F990.DB2.SQL(LOADTABL)
/*******
/** Enable table in IDAA (set acceleration to ON)

```

Te SSL R 5 C 15 TCP07595

File Edit View Communication Options Transfer Macro Help





Menu Utilities Compilers Help

BROWSE F990.COMP.CNTL (J9I79136) Line 00000034 Col 001 080
Command ==> _ Scroll ==> CSR

//ACCEL01 EXEC PGM=IKJEFT01,DYNAMNBR=20,COND=(4,LT)

//* parameter #1 for accelerator name

//AQTP1 DD DISP=SHR,DSN=F990.DB2.SQL (AQTP1)

//* parameter #2 for ON or OFF

//AQTP2 DD DISP=SHR,DSN=F990.DB2.SQL (ACCAQTP2)

//* parameter #3 for set tables containing tables set

//AQTP3 DD DISP=SHR,DSN=F990.DB2.SQL (ACCAQTPA)

// DD *

<table name="IPP_STG_PROCESS_DETAIL" schema="ITS"

// DD DISP=SHR,DSN=F990.DB2.SQL (ACCAQTPB)

//* last parameter for message input to control trace

//AQTMMSGIN DD DISP=SHR,DSN=F990.DB2.SQL (ACCMMSGIN)

//SYSTSPRT DD SYSOUT=*

//SYSPRINT DD SYSOUT=*

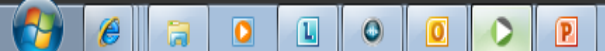
//SYSUDUMP DD SYSOUT=*

//SYSTSIN DD DISP=SHR,DSN=F990.DB2.SQL (SETTABLE)

//* If the following step fails, please mark the job complete and

//* send an email message to DataServics HIS. This step does not

Te SSL R 4 C 15 TCP07595



Administration E... Data Project Expl...

- tdb2.bcbsal.org
- tdb2vipa.bcbsal.org
- tdb3.bcbsal.org
- tdb4.bcbsal.org
- wdb2.bcbsal.org
 - IBM_WDB2 (DB2 for z/OS V10 (New-Func)
 - Accelerators
 - Aliases
 - Application Objects
 - Auxiliary Tables
 - Column Masks
 - Constraints
 - Databases
 - Indexes
 - MQTs
 - Row Permissions
 - Schemas
 - Sequences
 - Storage Groups
 - Synonyms
 - Tables
 - Table Spaces
 - Temporary Tables
 - Triggers
 - Users and Groups
 - VCATs
 - Views
 - XML Schemas
 - wdb3.bcbsal.org
 - wdb4.bcbsal.org
 - Working Sets

Accelerator: IDAABCBS @ IBM_WDB2 (Data Sharing Group)

Refresh: Every minute

Acceleration: Unknown [Start Stop](#) [Change](#) Credentials valid since: 7/18/13 3:29 PM [Update](#)
 Status: Online Trace: DEFAULT / OFF [Configure](#) [Save](#) [Clear](#)
 Used space: 5.77 TB of 32 TB Active queries: 7

- Query Statistics
- About

Tables (1,408 of 1,408 loaded / 1,408 of 1,408 enabled for acceleration) - Refreshing.

[Add...](#)
[Alter Keys...](#)
[Remove](#)
[Load...](#)
[Acceleration](#)
[Cancel Tasks](#)

Name like: type filter text

Name	Size	Rows	Acceleration	Last Load	Distribution Key	Skew	Organizing Keys
SUBS	1.65 TB	-	179 of 179	179 of 179 tables	-	-	-
ADD_CAND_ANLS	276 MB	12,697,410	Enabled	9/1/13 6:02 AM	Random	0.000	
ADDR_LCTN_ARCH	460 MB	10,623,568	Enabled	9/17/13 5:38 AM	Random	0.000	
ADDR_LCTN_SN			led	8/30/13 9:59 PM	Random	0.002	
AUTO_ENRL_NT			led	9/17/13 5:41 AM	Random	0.000	
AUTO_ENRL_NT			led	9/17/13 5:38 AM	Random	0.000	
BELL_TRNSPLT			led	9/16/13 10:00 PM	Random	0.000	
BEQ_RSP			led	9/17/13 5:37 AM	Random	0.335	
BEQ_RSP_LIS			led	9/17/13 5:38 AM	Random	0.000	
BEQ_RSP_MEDD			led	9/17/13 5:38 AM	Random	0.000	
BEQ_RSP_NCOV			led	9/17/13 5:38 AM	Random	0.000	
BEQ_SUBMTD			led	9/17/13 5:39 AM	Random	0.000	
BILG_CYCL			led	9/17/13 3:25 AM	Random	0.000	
BLUAD_ACPTD	105 MB	1,007,310	Enabled	9/17/13 9:01 AM	Random	0.000	

Query Monitoring

Observations and Issues

No special security required

Java programs accelerated to DB2 Analytics Accelerator

COBOL programs with dynamic SQL accelerated to DB2 Analytics Accelerator

User concurrency

- Query runtime increased with increased workload

Observations and Issues (continued)

Loads to Accelerator initially very slow

- Resolved by allowing more tasks in the WLM environment

ODBC issue (DB2 Connect)

- Accelerator returned 0 rows for date dependent queries
 - Resolved with DB2 Connect 9.7 Fixpack 7

Business Value Proposition

Improved Business Intelligence

- 'Train of Thought' analysis
- Able to analyze more data than ever before

Query acceleration with little or no modifications to SQL

Minimal modifications to existing ETL processes

- Additional step to load data to Accelerator

Business Value Proposition (continued)

Fewer indexes needed in Data Warehouse environment

Short implementation timeline

DB2 Analytics Accelerator Implementation

Purchased Twinfin 12

- PoC was conducted on Twinfin 6

Mainframe connected to new Netezza appliance and DB2 Analytics Accelerator installed on the main Data Warehousing subsystem

DB2 Connect upgraded to version 9.7 Fixpack 7 across all desktops in the enterprise

DB2 Analytics Accelerator Implementation

Batch jobs created to refresh data to DB2 Analytics Accelerator based on the same schedule as the Data Warehouse

Turned on zParm to automatically accelerate all of the eligible workload

Post Implementation Observations

DB2 Analytics Accelerator returned data in different order than native DB2 processing

- Resolved with ORDER BY clause

Accelerated queries run between Data Warehouse load and Accelerator load do not return most current data

DBAs must remove/add/load tables to Accelerator if structure of Data Warehouse table changes

IBM Data Analytics Software updated to version 3.1

Post Implementation Observations (continued)

New technology is a work in progress...

- A true monitoring tool for IDAA does not exist
- Black box
- Upgrades require IBM intervention

However, it quickly becomes a major part of business

- Queries expected to run quickly in business areas
- Developers asking to utilize acceleration

Upcoming Plans

Upgrade IDAA software to version 4.1

Partition existing appliance into
Production/Development environments

Purchase a new appliance with upgraded technology
to use in production environment and utilize existing
appliance in a development environment.

Evaluate newly developed IDAA utility software

In Closing...

"I used to have to run a series of queries (in Crystal) against a pretty big table – the queries had to be split across multiple reports for multiple time periods due to the huge volume of data being queried. Each report took between 30 and 60 minutes to run, and there was a lot of export/cut/paste on the back end to get all the data into a manageable format. Once IDAA was implemented, I was able to convert all those reports into one SQL query that runs in less than 15 minutes and returns over 7 million records. It has eliminated hours of work associated with the old way of preparing this data." – BCBSAL Healthcare Analytics employee

In Closing...

Business customers now rely on the “Magic Box” to retrieve critical data for analysis!!!



Thank You!

sdavis@bcbsal.org