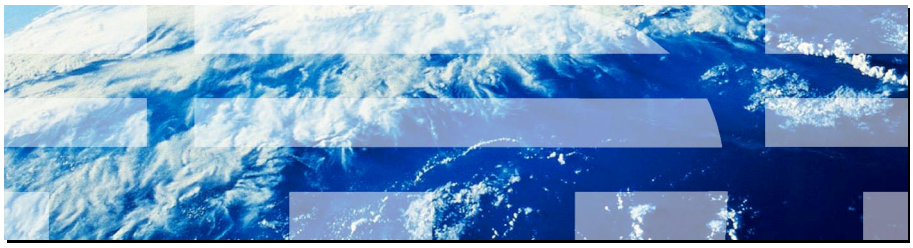
An IBM Presentation

SmarterAnalytics

IBM

IBM DB2 10.5 BLU Acceleration



Version 1.3 September 13th, 2013
Vikram S Khatri
Maria N Schwenger

SmarterAnalytics

IBM

TLA BLU

Does "BLU" stands for anything?

- **Outside IBM – Some examples**
  - Blue Chip Value Fund
  - Bulk Loading Unloading
  - Boston Linux Unix
  - Bande Latérale Unique
  - Basic Link Unit
  - Basic Logic Unit
  - Builders League United
  - Border Liaison Unit
  - Bomb Live Unit
  - Better Left Unsaid
  - Boys Like Us

However, we can make **BLU** stand for the following:

```
B – Big Data
L – Lightening Fast
U – Ultra Easy
```

- **The 'BLU' does not stand for anything. It was an IBM Research project 'Blink'.**

http://researcher.watson.ibm.com/researcher/files/us-ipandis/vldb13db2blu.pdf

SmarterAnalytics IBM

DB2 for Linux UNIX and Windows Packaging
DB2 editions

DB2 10.5 Editions

- **Express-C (A no-charge community edition)**
  - X64 based system. 16 GB memory, 2 cores, 15TB (No Fix Pack)
- **Express Server Edition**
  - X64 based system. 64 GB memory, 8 cores, 15TB (includes HADR and Fix Packs)
- **Developer Edition**
  - Support all the features of DB2. Use for learning. Cannot be used for production systems.
- **Workgroup Server Edition**
  - Windows®, Linux, Solaris®, IBM AIX®, HP-UX®, Linux on System z
  - 128 GB memory, 4 sockets / 16 cores, 15TB
- **Advanced Workgroup Server Edition**
  - DB2 DPF, BLU Acceleration, DB2 pureScale feature, SolidDB®, Compression, 5 User license of IBM Cognos BI
- **Enterprise Server Edition**
  - Windows, Linux, Solaris, IBM AIX, HP-UX, Linux on system z
  - No limit on memory and CPU
- **Advanced Enterprise Server Edition**
  - Same as DB2 ESE but with many more features licensed including DB2 pureScale feature, BLU Acceleration, WLM, DPF, SolidDB, Compression, 5 User license of IBM Cognos BI

SmarterAnalytics IBM

DB2 for Linux UNIX and Windows Packaging
Optimized for your business

Pick the DB2 that fits your needs

1. A single machine with a dedicated database
- Linux, Unix and Windows
- Grow by adding more processors and more memory (scale up)
- Start small and grow bigger by simply using more disk space
- Best suited for the majority of business requirements, from small to the large scale
- **Use BLU Acceleration for the analytics workload** NEW

2. A multiple logical or physical machine with partitioned database (DPF)
- Linux, Unix and Windows
- Grow by adding more logical or physical nodes (scale out)
- Best suited for data warehousing needs
- PureData for Operational Analytics is a pre-configured appliance

3. A multiple logical or physical machines with a shared database (pureScale)
- AIX and Linux
- Optimized for the OLTP operations with a focus on the continuous availability
- A shared database used by all DB2 members
- Grow by adding logical or physical DB2 members
- PureData for Transactions is a pre-configured appliance

SmarterAnalytics

IBM PureData – Overview
Different Workloads – Optimized Solutions

PureData for Transactions	Database services that handle large volumes of transactions with high availability, scalability and integrity
PureData for Operational Analytics	Data Warehouse services for complex analytics and reporting on data up to petabyte scale - with minimal administration
PureData for Analytics	Operational Warehouse services for continuous ingest of operational data, complex analytics, and a large volume of concurrent operational queries
PureData for Hadoop	Big data exploration easy. Built-in archiving tools, enterprise class appliance with built-in security and up to 8 times performance than custom built clusters.

SmarterAnalytics

DB2 10.5 Features
What is new?

- BLU Acceleration
- Support for HADR between two DB2 pureScale cluster
- Add member online in a DB2 pureScale cluster
- Explicit hierarchical locking (reduce CF traffic)
- WLB on subset of members
- Restore DB2 10.5 backup image to a DB2 pureScale cluster
- In-place table reorg in DB2 pureScale
- Member specific STMM in pureScale
- Simplified fix pack install in DB2 ESE and pureScale
- DB2 ACS allows custom scripts for snapshot backup and restore for HADR
- Function (expression) based indexes
- Text Search Enhancements
- Extended Row size support (`extended_row_size` must be set to enable)
- Exclude NULL keys in CREATE INDEX
- Not enforced primary keys / unique constraints
- Enhanced tooling : Data Studio, OQWT, OPM, OCM etc.

SmarterAnalytics IBM

BLU Acceleration
Driving Principles for the Design

- **Fast**

Unprecedented performance for analytical workloads, often 8x to 25x faster

- **Small**

Stronger compression and less space required for auxiliary data structures

- **Simple**

Much less tuning needed, more predictable and reliable performance

SmarterAnalytics IBM

BLU Acceleration Design Principles
The 7 Big Ideas

BLU Acceleration utilizes 7 big ideas:

- **Simplicity and Ease of Use**
  - It's just DB2 using standard SQL - with minimal set up requirements
- **Compute Friendly Compression**
  - "Always on" for column organized tables on disk
- **Column Organized Tables**
  - Data storage and retrieval is optimal for queries that touch fewer columns
- **Data Skipping**
  - Efficiently ignores data in a column that is not required
- **Core-friendly Parallelism**
  - Understands and exploits physical attributes of the server cores
- **Parallel Vector Processing**
  - Multi-core and SIMD (Single Instruction Multiple Data) parallelism
- **Scan-friendly Memory Caching**
  - High percentage of interesting data fits into and then stays in memory

SmarterAnalytics IBM

BLU Acceleration – Big Idea # 1
Simplicity and Ease of Use

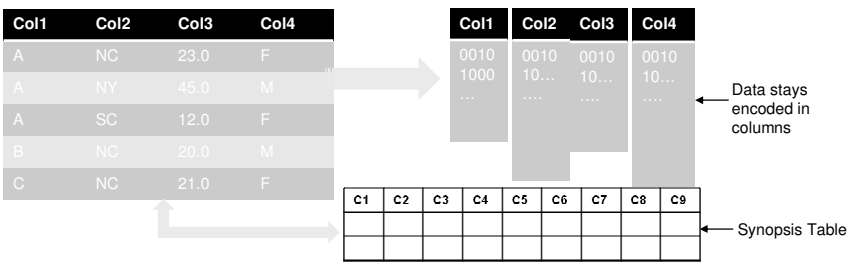
- **One registry variable:**
  DB2_WORKLOAD=ANALYTICS
  - Create the database to get optimal settings for an analytic environment and optimized to your hardware
  - It's still just DB2: use traditional storage, utilities, SQL, application interfaces, and so on
- **LOAD and GO**
  - Compression automatically done and statistics automatically collected
  - No need for indexes, partitions, MQTs (materialized views), MDCs, hints, statistical views, etc.
  - Underlying structures that support BLU are created and updated automatically
- **Maintenance Free**
  - REORGs (for space reclaim) and workload management is all automatic

```
db2set DB2_WORKLOAD=ANALYTICS
db2 CREATE DATABASE COLDB
db2 CREATE TABLE "BLU"."FACT_RX" (
    "MONTH_ID"           DECIMAL(6,0) ,
    "DATE_OF_SERVICE"    DATE ,
    "PROVIDER_ID"        DECIMAL(10,0) ,
    "PRODUCT_ID"         DECIMAL(10,0) ,
    "PERSON_ID"          DECIMAL(10,0) ,
    "PERSON_ZIP3_CD"     VARCHAR(3) ,
    "CID"                DECIMAL(14,0) NOT NULL ,
    "FILL_NBR"           DECIMAL(10,0) ,
    "DAYS_SUPPLY"        DECIMAL(10,0) ,
    "QUANTITY_DISPENSED" DECIMAL(10,0) ,
    "PERSON_OPC"         DECIMAL(10,0) ,
    "TOTAL_AMT_PAID"     DECIMAL(14,4) ,
    "PAYER_ID"           DECIMAL(10,0)
)
db2 LOAD FROM /tmp/fact_rx.dat OF DEL REPLACE
INTO BLU.FACT_RX
db2 "SELECT COUNT(DISTINCT RX.PROVIDER_ID)
  FROM BLU.FACT_RX RX INNER JOIN BLU.PERSON PT
  ON RX.PERSON_ID=PT.PERSON_ID"
```


SmarterAnalytics IBM

BLU Acceleration – Big Idea # 3
Column Store

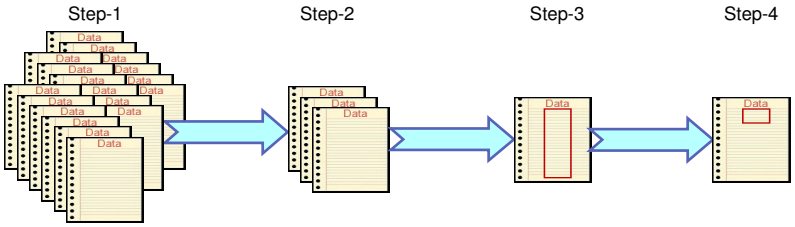
- **Analysis phase of LOAD**
  - Frequency determined for Huffman encoding and tree saved for decompression
  - SIMD processing is used for encoding to reduce CPU cycles
- **Rows converted to Columnar Storage**
  - Synopsis table built for minimum and maximum values of columns for every 1024 rows
  - Encoded data is stored in column store
- **Subsequent Inserts**
  - Local Huffman encoding tree saved at page level to provide adaptive compression for inserts
  - Synopsis table maintained automatically for IUD operations



SmarterAnalytics IBM

BLU Acceleration – Big Idea # 4
Data skipping

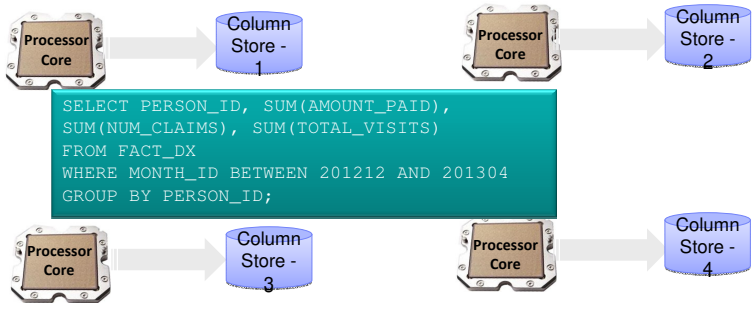
- **Step - 1**
  - A large quantity of data is sitting in file system - (Say 10 TB)
- **Step – 2**
  - Data is loaded in column store and encoding reduces data size - (Say 2 TB)
- **Step – 3**
  - Data is accessed for a column - (Say 500 MB)
- **Step – 4**
  - Pages skipped for the range that do not qualify as per the synopsis table
  - Actual data read is very small - (Say 32 MB)



SmarterAnalytics IBM

BLU Acceleration – Big Idea # 5
Multi-Core parallelism

- **Multi-Cores are used by BLU Acceleration in parallel on multiple columns**
  - A separate agent per core can fetch individual column data
  - Each agent works on a different query functions
  - Before CTQ operator is reached, data is processed per core
  - REBAL access plan operator is an indication that multi cores are used



SmarterAnalytics IBM

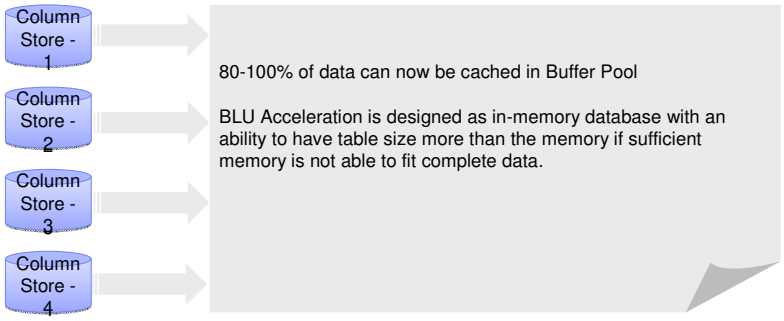
BLU Acceleration – Big Idea # 6
SIMD CPU exploitation

- **SIMD (Single Instruction Multiple Data) are designed to exploit parallelism at data level**
  - Same operation is performed on multiple data elements simultaneously. For example, encoding of data in column store
  - SIMD processing for Scans, Joins, Groupings and Arithmetic
- **Modern CPUs have built-in capability to perform SIMD**
  - For example, *Intel® Streaming SIMD Extensions* (Intel® SSE) Version 4.2 can be found on Intel Core i7 processor
  - BLU Acceleration has been tested to perform on:
    - AIX on Power - best performance on Power 7
    - Linux 64-bit on Intel / AMD - best performance on Intel Nehalem or better

SmarterAnalytics IBM

BLU Acceleration – Big Idea # 7
Scan-friendly memory caching

- **Data effectively cached in memory using new algorithm**
  - LRU algorithm to evict data and MRU to keep data used in regular DB2 row organized tables is not a best fit for in-memory caching for column organized tables and instead new algorithm was designed.
  - New scan-friendly memory caching is an automatically triggered cache-replacement algorithm that provides egalitarian access
  - High percentage of data can now fit in memory with new algorithm 80-100% as opposed to 15-50% for row organized tables



SmarterAnalytics IBM

DB2 BLU Acceleration
Hardware and Platform

- **Supported Hardware Platform**
  - Intel Nehalem (e.g. Core i7 for desktop and Xeon for servers) or better
  - POWER7 – Superscalar Symmetric Multiprocessor or better
  - Create a new UNICODE Database
- **Supported Platform and Operating Systems**
  - Linux 64-bit on Intel/AMD
    - RHEL 6 or higher, SLES 10 SP2, SLES 11 SP2
  - AIX on POWER7 hardware
    - AIX 6.1 TL7 SP6, AIX 7.1 TL1 SP6

SmarterAnalytics IBM

DB2 BLU Acceleration
Workload Recommendations

	Small	Medium	Large
Raw Data Size	1 TB	5 TB	10 TB
	Minimum required for performance		
Cores	8	16	32
Memory	64 GB	256 GB	512 GB
	For high-end performance		
Cores	16	32	64
Memory	128-256 GB	384-512 GB	1024-2048 GB

SmarterAnalytics IBM

DB2 BLU Acceleration
Workloads that will benefit or not benefit

Workloads that will benefit	Workloads that will not benefit
Analytic workloads, data marts	Singleton selects with few insert/update/delete
Star or dimensional schemas	Fully Normalized database for OLTP
SAP Business Warehouse	Insert, update or delete of few rows per transaction
Grouping, aggregation, range scans, joins	Queries accessing most or all columns in a table
Queries that access only a subset of columns in a table	Heavy use of LOBS, XML, structured data types, temporal data, generated columns
Queries that touch more than 1% of the data	HADR, LBAC and RCAC
Data sizes as shown above	OS is not AIX or Linux
Moderate amount of data (< 20 TB)	

SmarterAnalytics IBM

Compare Database
Row vs. Column Organized Database

DB2 10.1	DB2 10.5
CREATE DATABASE command	Same
CREATE BUFFERPOOL statement	Same
CREATE TABLESPACE statement	Same
CREATE TABLE statement	Use the ORGANIZE BY COLUMN clause; otherwise, the syntax is unchanged.
LOAD command for each table	Same
CREATE INDEX statement	Not required
Define constraints	Same
Define or refresh MQTs	Not required because performance is already optimized
RUNSTATS command on all tables and MQTs	Not required, because table runstats operations are performed as part of the data load process, and MQTs are not created
Create statistical views and invoke the RUNSTATS command against those views	Not required because performance is already optimized

SmarterAnalytics IBM

Leverage DB2 BLU Acceleration
Old Version Databases

- **Upgrading from the previous version of DB2 release to DB2 10.5**
  - You can upgrade databases created in DB2 versions 9.7, 9.8, and DB2 10.1 to DB2 version 10.5
  - A database that was created prior to DB2 9.7 must be upgraded to DB2 versions 9.7, 9.8 or 10.1 before it can be upgraded to DB2 10.5
  - The existing database must use UNICODE with IDENTITY or IDENTITY_16BIT collation to leverage BLU acceleration
  - There is no option to change collation of an existing database. Only option is to recreate the database
  - Run `db2chkupgrade` tool to verify if existing database can be upgraded
  - Upgrade DB2 instance by using `db2iupgrade` utility
  - Upgrade Databases – Use one of the option
    - In-place upgrade using the `UPGRADE DATABASE` command
    - Side-by-side upgrade using the `RESTORE` command for the backup image that was taken in older DB2 release

SmarterAnalytics IBM

Leverage DB2 BLU Acceleration
Existing Database

- **Existing database in DB2 10.5**
  - An existing UNICODE database with IDENTITY or IDENTITY_16BIT collation has few tables used in analytics queries. Consider converting those tables to column organized tables
  - Set `db2_workload=ANALYTICS`
  - Restart the DB2 instance `(db2stop/db2start)`
  - Run `AUTOCONFIGURE APPLY DB AND DBM` command to configure the database for analytics
  - Enable AUTOMATIC STORAGE
    - If you are not using automatic storage, use `CREATE STOGROUP` to create a storage group

      ```
      CREATE STOGROUP ibmcolstogrp ON '/data1' SET AS DEFAULT
      ```

    - Convert existing DMS table spaces to use automatic storage

      ```
      ALTER TABLESPACE tbsp MANAGED BY AUTOMATIC STORAGE
      ALTER TABLESPACE tbsp REBALANCE
      ```

    - Create new table spaces using 32 KB page size with extent size of 4 pages

      ```
      CREATE TABLESPACE coltbsp PAGESIZE 32 K EXTENTSIZE 4
      ```


SmarterAnalytics IBM

DB2_WORKLOAD=ANALYTICS
Automatically set DBM and DB Parameters

	Customized Setting	Description
Instance Parameter	INTRA_PARALLEL=YES	Enables intra-partition parallelism
Database Parameter	DFT_TABLE_ORG=COLUMN	Tables are created column organized by default unless otherwise specified
	DFT_DEGREE=ANY	Enables intra-partition parallelism to use all detected cores
	PAGESIZE=32768	Default page size for table space or buffer pool if not specified
	DFT_EXTENT_SZ=4	The default extent size for a table space
	SORTHEAP=[default+n]	Private sort heap [set higher than the default]
	SHEAPTHRES_SHR=[default+n]	Shared sort heap [set higher than the default]
	UTIL_HEAP_SZ=[default+n]	Utility heap [set higher than the default]
	CATALOGCACHE_SZ=[default+n]	System catalog cache usage of the dbheap [set higher than the default]
	AUTO_REORG=ON	Enables automatic REORGs for space reclamation
WLM objects	• Work Action Set • Service Subclass • Threshold	These objects are created and set to maximize throughput in your database on your hardware when many large analytic type queries are submitted

SmarterAnalytics IBM

Leverage DB2 BLU Acceleration
New Database

- **New database in DB2 10.5**
  - Set `db2_workload=ANALYTICS`
  - Restart the DB2 instance `(db2stop/db2start)`
  - Create a new UNICODE Database

```
CREATE DATABASE COLDB USING CODESET UTF-8
 TERRITORY US COLLATE USING IDENTITY
```

  - Create Automatic Table Space

```
CREATE TABLESPACE coltbsp
```


SmarterAnalytics IBM

DB2 BLU Acceleration
Creating a column-organized table

- **Example**

```
CREATE TABLE COL_TAB
(
   c1 INT NOT NULL,
   c2 INT,
   ……
   PRIMARY KEY(C1)
) ORGANIZE BY COLUMN;
```

- If `DFT_TABLE_ORG = COLUMN` `ORGANIZE BY COLUMN` is the default and can be omitted
- Use `ORGANIZE BY ROW` to create row-organized tables
- Do not specify compression, MDC, or partitioning for BLU tables
- Do not create indexes or MQTs
- Columnar tables are always compressed by default
- The `TABLEORG=R` or to `TABLEORG=C` in `syscat.tables` indicates if a table is row or column organized.

SmarterAnalytics IBM

DB2 BLU Acceleration
Non-enforced PK / FK constraints

- **Only non-enforced foreign keys are supported**

```
ALTER TABLE INVENTORY_FACT
ADD CONSTRAINT FK_INVENTORY FOREIGN KEY ( BRANCH_KEY )
REFERENCES BRANCH_DIM( BRANCH_KEY) NOT ENFORCED;
```

- Primary keys and unique constraints can be enforced or not enforced
- Enforced Primary Keys - DB2 uses an internal B-tree structure to guarantee uniqueness efficiently

SmarterAnalytics IBM

Storage Consideration
Column Organized Tables

- Data must be in row format to load in column organized tables
- Data gets compressed and converted to columnar format upon LOAD/INSERT
- Separate set of extents and pages for each column
- Generally, column oragnized tables take less space than row-organized tables
- Column organized tables with many columns and less number of rows will take more space than row organized tables (extents being nearly empty)
- Tuple Sequence Number (TSN) is stored with column data on a page
- TSN is the link to stitch columns from a page into a row
- Lot more I/Os compared to a row organized table to get all columns
- Lot less I/Os compared to a row organized table to fetch few columns
- Each column organized table has an auxiliary synopsis table
- The size of synopsis table is 0.1% of the actual size
- No indexes – no worry about space allocation for them.

SmarterAnalytics IBM

Table Organization
Catalog Information

- **Row / Column Organized Tables**
- **New Column `TABLEORG` in `SYSCAT.TABLES`**

```
SELECT tabname, tableorg
FROM   syscat.tables
WHERE  tabname in ('PERSON', 'POLICY')

TABNAME         TABLEORG
--------------- --------
PERSON          C
POLICY          R
```


SmarterAnalytics IBM

Storage Layout
Rows to Columns

- **Row Organized Table - Page**

```
200907 07/27/2009 1192217 1037324 300  78356108 85585838582 5  30  4
200907 07/12/2009 2862550 1037324 704 242047371 34216679536 2  60  8
200907 07/07/2009  985202 1037324 300 163197664 80230203864 0  30  8
200907 07/09/2009 2379287 1037324 841  61976761 34188181271 1 100  0
200907 07/13/2009 2115026 1037324 731 154015058 34212598932 0  30 10
200907 07/22/2009  602222 1037324 125 380807278 34283841044 2  90  9
200907 07/28/2009  129190 1037324 939  85578872 34325381915 2  30 15
200907 07/20/2009 2008190 1037324 530 229825070 85515209633 0  30 10
200907 07/09/2009 1240780 1037324 190   3917470 85496553638 5  30  0
200907 07/13/2009 2162196 1037324 300 291154219 52626842512 2  30 10
```

- **Column Organized Table – Pages (Data stays compressed)**

```
Ox0009292 200907 200907 200907 200907 200907
200907 200907 200907 200907 200907
```

```
Ox0009302 07/27/2009 07/12/2009 07/07/2009
07/09/2009 07/13/2009 07/22/2009 07/28/2009
07/20/2009 07/09/2009 07/13/2009
```

```
Ox0009322 1192217 2862550 985202 2379287
2115026 602222 129190 2008190 1240780
2162196
```

```
Ox0009342 1037324 1037324 1037324 1037324
1037324 1037324 1037324 1037324 1037324
1037324
```

```
Ox0009362 300 704 300 841 731 125 939 530
190  300
```

```
Ox0009382  78356108 242047371 163197664 61976768
154015058 380807278 85578872 229825070 3917470
291154219
```

```
Ox0009402 300 704 300 841 731 125 939 530 190  300
```

```
Ox0009422 85585838582 34216679536 80230203864
34188181271 34212598932 34283841044 34325381915
85515209633 85496553638 52626842512
```

```
Ox0009442 5 2 0 1 0 2 2 0 5 2
```

```
Ox0009462 30 30 30 100 30 90 30 30 30 30
```

```
Ox0009482  4 8 8 10 9 15 10 0 10
```


SmarterAnalytics IBM

Loading Data
LOAD Utility

- **Key LOAD symantics remains unchanged for column organized tables**
- **New "Analyze" Phase for column organized tables**
- **Fully formatted and compressed column organized pages are created from row-organized input data**

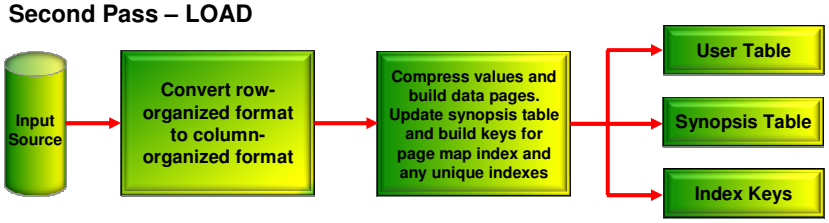
```
load from /tmp/root/fact_rx_50M.dat of del replace into blu.fact_rx statistics use profile
SQL3109N  The utility is beginning to load data from file "/tmp/root/fact_rx_50M.dat".
SQL3500W  The utility is beginning the "ANALYZE" phase at time "08/03/2013 00:30:13.042626".
SQL3519W  Begin Load Consistency Point. Input record count = "0".
SQL3520W  Load Consistency Point was successful.
SQL3515W  The utility has finished the "ANALYZE" phase at time "08/03/2013 00:31:37.724462".
SQL3500W  The utility is beginning the "LOAD" phase at time "08/03/2013 00:31:37.726853".
SQL3110N  The utility has completed processing.  "50000000" rows were read from the input file.
SQL3519W  Begin Load Consistency Point. Input record count = "50000000".
SQL3520W  Load Consistency Point was successful.
SQL3515W  The utility has finished the "LOAD" phase at time "08/03/2013 00:34:47.470123".
SQL3500W  The utility is beginning the "BUILD" phase at time "08/03/2013 00:34:47.474585".
SQL3213I  The indexing mode is "REBUILD".
SQL3515W  The utility has finished the "BUILD" phase at time "08/03/201300:34:47.850447".
Number of rows read         = 50000000
Number of rows skipped      = 0
Number of rows loaded       = 50000000
Number of rows rejected     = 0
Number of rows deleted      = 0
Number of rows committed    = 50000000
```


SmarterAnalytics IBM

LOAD Utility
REPLACE INTO TABLE

- **First Pass – ANALYZE**

- **Second Pass – LOAD**



SmarterAnalytics IBM

Convert Tables Online
Row to Column Organized

- **Column organized table conversion from row organized tables**
  - Command: **`db2convert`**
  - Keeps table fully on-line during the conversion (utilizes **`ADMIN_MOVE_TABLE`**)
  - Convert single tables one at a time, entire schemas of tables or the entire database

```
db2convert  -d  COLDB -z  blu  -t  fact_rx
```


SmarterAnalytics IBM

Efficient Storage Management
Automatically Reclaim Space

- **Column organized table – reclamation of space**
  - If `DB2_WORKLOAD=ANALYTICS`, automatic space reclamation is active for all column-organized tables
  - Enable Automatic Table Maintenance

```
update db cfg using auto_maint ON
auto_tbl_maint ON auto_reorg ON
```

  - Use REORG explicitly to reclaim space

```
db2 REORG TABLE FACT_RX RECLAIM EXTENTS
```


SmarterAnalytics IBM

Columnar Dictionary
Catalog Information

- **Row / Column Organized Tables**

```
SELECT tabname, tableorg, compression
FROM   syscat.tables
WHERE  tabname in ('PERSON', 'POLICY')

TABNAME         TABLEORG COMPRESSION
--------------- -------- -----------
PERSON          C
POLICY          R        N
```

- **For column organized tables, COMPRESSION is always blank as this cannot be changed.**

SmarterAnalytics IBM

Columnar Dictionary
Catalog Information (continued…)

Row Organized Table	Column Organized Table
PCTPAGESSAVED	**PCTPAGESSAVED**
AVGCOMPRESSEDROWSIZE	
AVGROWCOMPRESSIONRATIO	
AVGROWSIZE	
PCTROWCOMPRESSED	

- **Only** PCTPAGESSAVED **applies to column-organized tables**
  - **Approximate percentage of pages saved in the table**
  - **RUNSTATS collects PCTPAGESSAVED by estimating the number of data pages needed to store table in uncompressed row organized**

SmarterAnalytics IBM

Compression Dictionaries
Row / Column Organized Table

	Row Organized Table	Column Organized Table
Table Level Compression	ALTER TABLE … COMPRESS YES STATIC	Table level compression is always on
Page Level Compression	ALTER TABLE … COMPRESS YES ADAPTIVE	Page level compression is always on
Number of Dictionary	One	'N' = Number of columns
Static	Once built - never updated	Once built - never updated
Adaptive	Page level if required	Page level for each column if required
Turn off compression	ALTER TABLE … COMPRESS NO	Cannot be turned off
REORG	Use KEEPDICTIONARY or RESETDICTIONARY	Does not change column level dictionary
LOAD REPLACE	KEEPDICTIONARY is default if dictionary exists	RESETDICTIONARY is default if dictionary exists
LOAD INSERT	KEEPDICTIONARY is default	KEEPDICTIONARY is default

SmarterAnalytics IBM

Compressing Data
Memory Considerations

- **LOAD uses pages from `util_heap_sz` memory pool**
- **`util_heap_sz` - Bigger is better for LOAD**
- **At least 1,000,000 pages**
- **Preferred 4,000,000 pages if server has > 128GB memory**
- **For concurrent LOAD, increase `util_heap_sz`**
- **Synopsis table has one row for each 1024 rows in the user table**
- **Size of the synopsis table is 0.1% of the size of the user table**

- **Row Organized Tables**
  - **Insufficient memory during load will slow load performance**
  - **Compression of the tables is not affected**

- **Column Organized Tables**
  - **Insufficient memory during the LOAD ANALYZE phase could yield less than optimal compressed tables**

SmarterAnalytics IBM

Compressing Data
Best Results

- **First LOAD should be on large quantity of data – Need enough input values to build effective dictionary**

- **`util_heap_sz` should be as big as possible**
- **Though Automatic Dictionary Creation is possible, LOAD ANALYZE can work on large data set and thus can have better compression dictionary**

SmarterAnalytics IBM

Estimating Table Level Compression
Row and Column Organized Tables

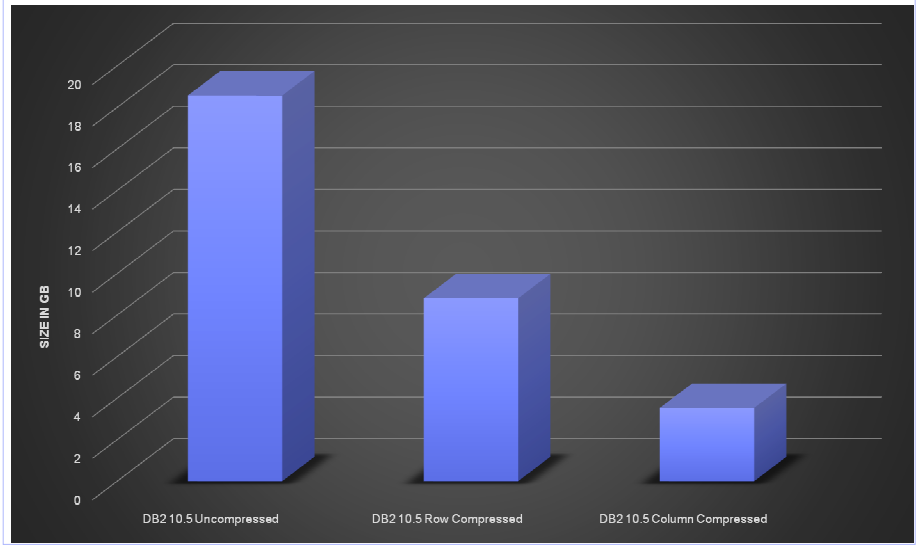
Row Organized Table	Column Organized Table
PCTPAGESSAVED = 50	**PCTPAGESSAVED = 75**
Compression Ratio = 1 / (1 – PCTPAGESSAVED/100)	Compression Ratio = 1 / (1 – PCTPAGESSAVED/100)
Compression Ratio = 1 / (1 – 50/100)	Compression Ratio = 1 / (1 – 75/100)
Compression Ratio = 2	Compression Ratio = 4

- **Load data in row organized and column organized table**
- **Compare** `PCTPAGESSAVED`

```
SELECT a.tabname,
       DEC(1.0/(1.0-(PCTPAGESSAVED*1.0)/100.0),31,2) as
       COMPRESSION_RATIO,
       PCTPAGESSAVED
FROM   syscat.tables a, SYSIBMADM.ADMINTABINFO b
WHERE  a.tabname = b.tabname
AND    a.tabname like 'MYTABLE%'
ORDER BY a.tabname
```


SmarterAnalytics IBM

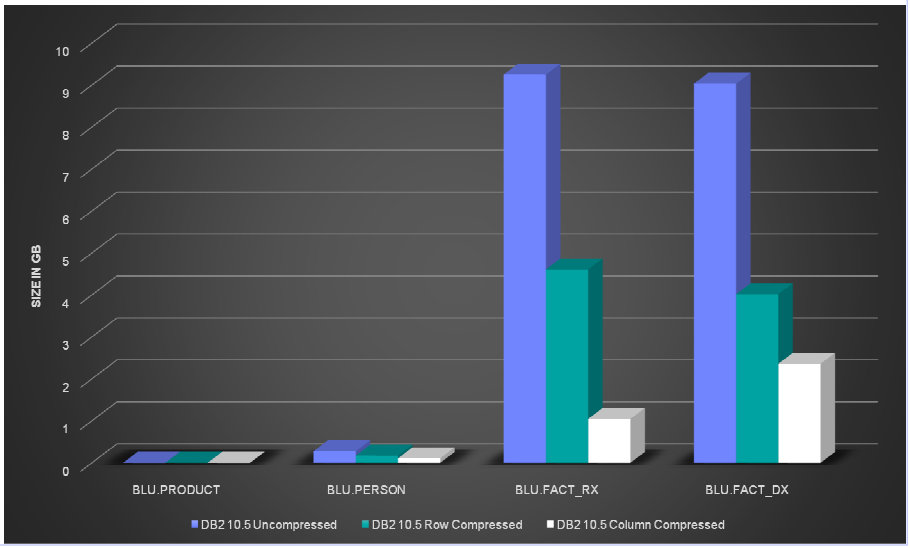
Row/Column Organized Tables
Compare Compression at Database Level



SmarterAnalytics

IBM

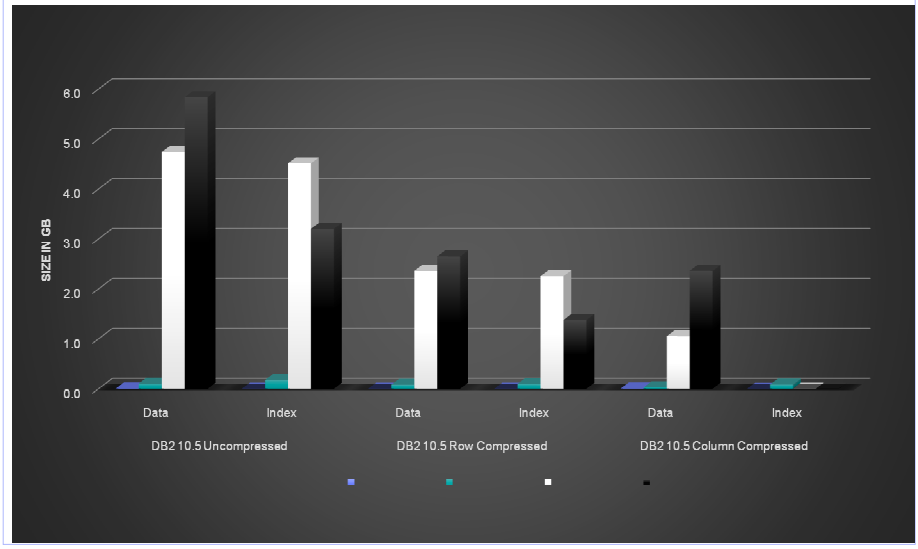
Row/Column Organized Tables
Compare Compression at Table Level



SmarterAnalytics

IBM

Row/Column Organized Tables
Compare Compression at Table Level for Data and Index



SmarterAnalytics IBM

Data Skipping
Concepts

- BLU acceleration is ***not*** for the operational queries that access a single row or a few rows (likely by using an index)
- Operational queries against row organized tables using indexes jump to the data directly. When indexes cannot be used, a full table scan is performed
- DB2 uses MRU (Most Recently Used) algorithm to keep pages in the buffer pool for row organized tables
- DB2 uses LRU (Least Recently Used) algorithm to evict pages in the buffer pool to make room for other pages that need to be read
- BLU acceleration is ***for*** data mart like analytic workloads that use activities such as grouping, aggregation, range scans
- DB2 uses scan friendly memory-caching algorithm to keep maximum data
- In absence of indexes for column organized tables, DB2 achieves the data skipping by using data from the synopsis table for the column being accessed
- The synopsis tables is automatically maintained during INSERT, UDATE, and DELETE
- Each row in the synopsis table is tied to a certain chunk of data records (usually 1024) and index map relates these to the physical blocks on the disk

SmarterAnalytics IBM

Catalog Information
Synopsis Table

- **Each columnar table has a synopsis table**
- **Synopsis table is created and updated automatically**

```
SELECT bschema, bname, tabschema, tabname
FROM   syscat.tabdep
WHERE  bname = 'PRODUCT'

BSCHEMA BNAME   TABSCHEMA TABNAME
------- ------- --------- ----------------------------------
BLU     PRODUCT SYSIBM    SYN130831232022509126_PRODUCT
GOSALES PRODUCT SYSIBM    SYN130831233146744370_PRODUCT
```

- **Synopsis table has one row for each 1024 rows in the user table**
- **Size of the synopsis table is 0.1% of the size of the user table**
- **Enables DB2 to skip portions of a table while scanning data for a query**
- **Loading pre-sorted data helps to cluster data**
- **It inherits same storage format as regular BLU tables**

SmarterAnalytics IBM

Catalog Information
Synopsis Table (…continued)

- **Metadata describing range of values of the user table**

```
$ db2 describe table SYSIBM.SYN130803002948442521_FACT_RX
 Data type                 Column
Column name              schema    Data type name  Length     Scale Nulls
------------------------ --------- --------------- ---------- ----- ------
MONTH_IDMIN              SYSIBM    DECIMAL                  6     0 Yes
MONTH_IDMAX              SYSIBM    DECIMAL                  6     0 Yes
DATE_OF_SERVICEMIN       SYSIBM    DATE                     4     0 Yes
DATE_OF_SERVICEMAX       SYSIBM    DATE                     4     0 Yes
PROVIDER_IDMIN           SYSIBM    DECIMAL                 10     0 Yes
PROVIDER_IDMAX           SYSIBM    DECIMAL                 10     0 Yes
PRODUCT_IDMIN            SYSIBM    DECIMAL                 10     0 Yes
PRODUCT_IDMAX            SYSIBM    DECIMAL                 10     0 Yes
……
……
PAYER_IDMIN              SYSIBM    DECIMAL                 10     0 Yes
PAYER_IDMAX              SYSIBM    DECIMAL                 10     0 Yes
TSNMIN                   SYSIBM    BIGINT                   8     0 No
TSNMAX                   SYSIBM    BIGINT                   8     0 No
```


SmarterAnalytics IBM

Catalog Information
Page Map Index

- **Automatically created and maintained**
- **Used internally to locate column data in the storage object**
- **Maps TSNs to Pages**

```
SELECT indschema, indname, tabname, colnames, indextype
FROM   syscat.indexes
WHERE  tabname like 'FACT%'

INDSCHEMA INDNAME              TABNAME    COLNAMES                              INDEXTYPE
--------- -------------------- ---------- ------------------------------------- ---------
SYSIBM    SQL130803002949444571 FACT_DX    +SQLNOTAPPLICABLE+SQLNOTAPPLICABLE   CPMA
SYSIBM    SQL130803002948668003 FACT_RX    +SQLNOTAPPLICABLE+SQLNOTAPPLICABLE   CPMA
```


SmarterAnalytics IBM

Synopsis Table
FACT_RX

- **We will examine synopsis table for FACT_RX**

```
SELECT VARCHAR(TABNAME,50) TABNAME
FROM SYSCAT.TABLES
WHERE TABNAME LIKE 'SYN%FACT_RX'
AND TABLEORG='C'

SELECT * FROM sysibm.SYN130803002948442521_FACT_RX
```

MONTH_IDMIN	MONTH_IDMAX	DATE_OF_SERVICEMIN	DATE_OF_SERVICEMAX	PROVIDER_IDMIN	PROVIDER_IDMA
200709.	200709.	09/01/2007	09/30/2007	10171.	5100421
200709.	200709.	09/01/2007	09/30/2007	11045.	5175393
200709.	200709.	09/01/2007	09/30/2007	10585.	5180836
200709.	200709.	09/01/2007	09/30/2007	6415.	638031
200709.	200709.	09/01/2007	09/30/2007	11146.	591530
200709.	200709.	09/01/2007	09/30/2007	10704.	5179285
200709.	200709.	09/01/2007	09/30/2007	10263.	643410
200709.	200709.	09/01/2007	09/30/2007	10990.	6141327
200709.	200709.	09/01/2007	09/30/2007	10126.	536734
200709.	200709.	09/01/2007	09/30/2007	1053.	517928

SmarterAnalytics IBM

Data Skipping
Examine Column Cardinality

- **Table FACT_RX - Column Cardinility**

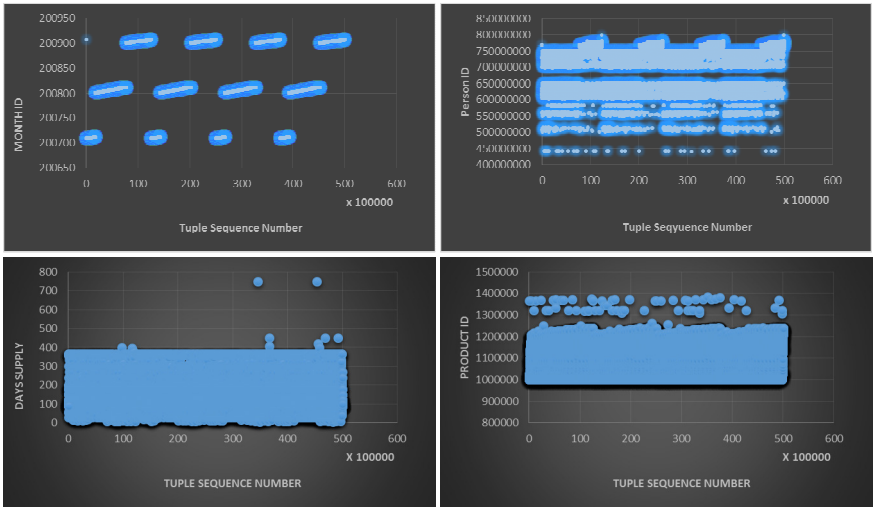
```
SELECT TABNAME, COLNAME, COLCARD
FROM SYSCAT.COLUMNS WHERE TABSCHEMA = 'BLU'
 AND TABNAME = 'FACT_RX' AND TYPENAME != 'VARCHAR'
ORDER BY COLCARD
```

TABNAME	COLNAME	COLCARD
FACT_RX	MONTH_ID	23
FACT_RX	FILL_NBR	101
FACT_RX	DAYS_SUPPLY	316
FACT_RX	DATE_OF_SERVICE	760
FACT_RX	QUANTITY_DISPENSED	1888
FACT_RX	PERSON_OPC	2721
FACT_RX	PAYER_ID	4608
FACT_RX	PRODUCT_ID	33792
FACT_RX	TOTAL_AMT_PAID	123905
FACT_RX	PROVIDER_ID	788280
FACT_RX	PERSON_ID	2657367
FACT_RX	CID	41173561

SmarterAnalytics

IBM

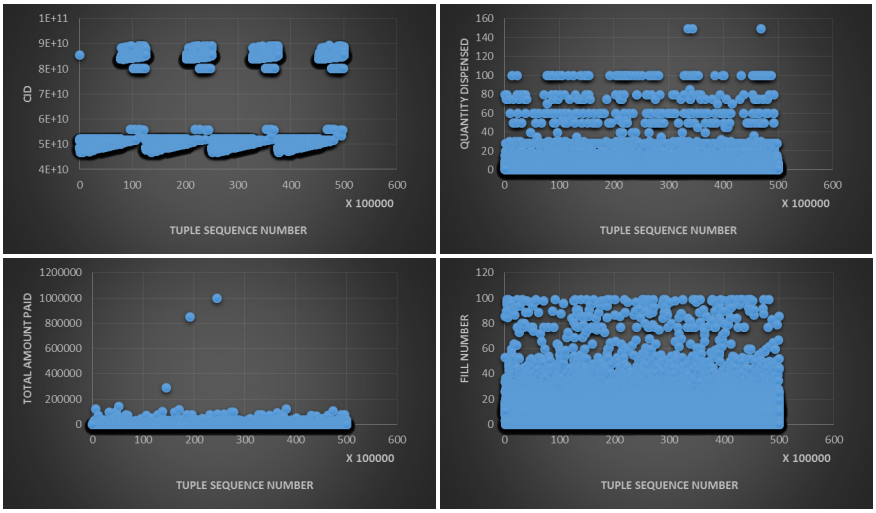
Data Skipping
Examine Column Clustering



SmarterAnalytics

IBM

Data Skipping
Examine Column Clustering (continued…)

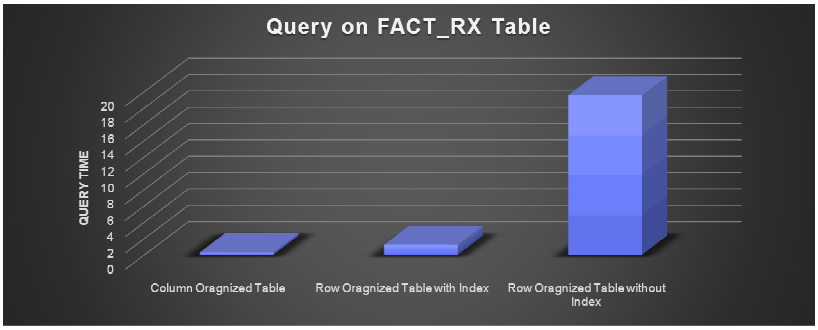


SmarterAnalytics IBM

Data Skipping
Compare Row / Organized Table

- **Query on FACT_RX Table**

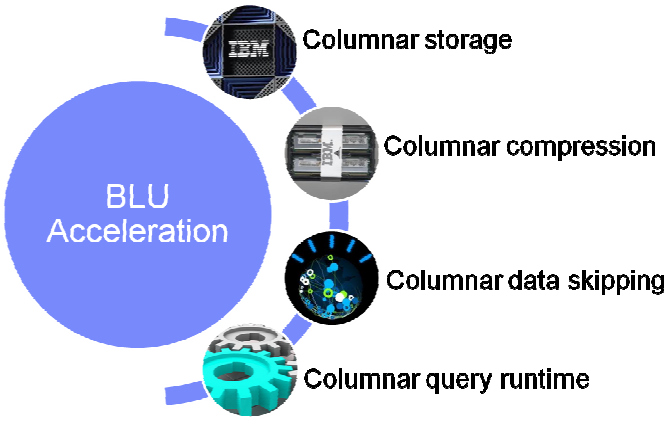
```
SELECT count(*) NUM_ROWS
FROM   blu.fact_rx
WHERE  month_id = 200709
AND    product_id between 1190000 and 12000000 WITH CS;
```



SmarterAnalytics IBM

DB2 BLU Acceleration
Query Optimization

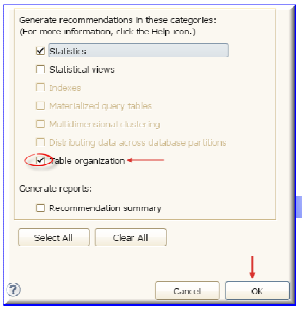
- BLU acceleration is much more than columnar storage

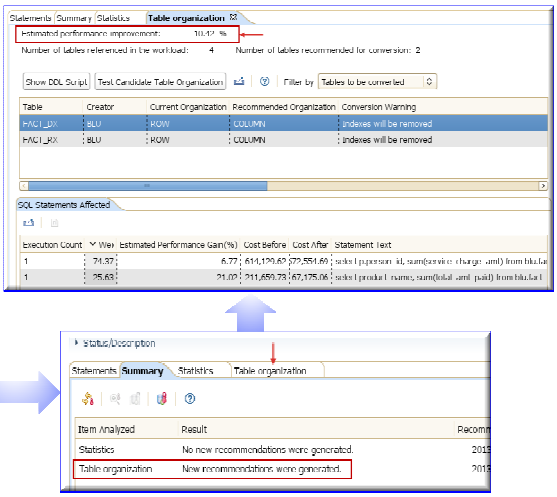


SmarterAnalytics

Query Optimization
Optim Query Workload Tuner

- Optim Query Workload Tuner provides Table Organization Advisor to suggest which tables are good candidates to convert to column organization

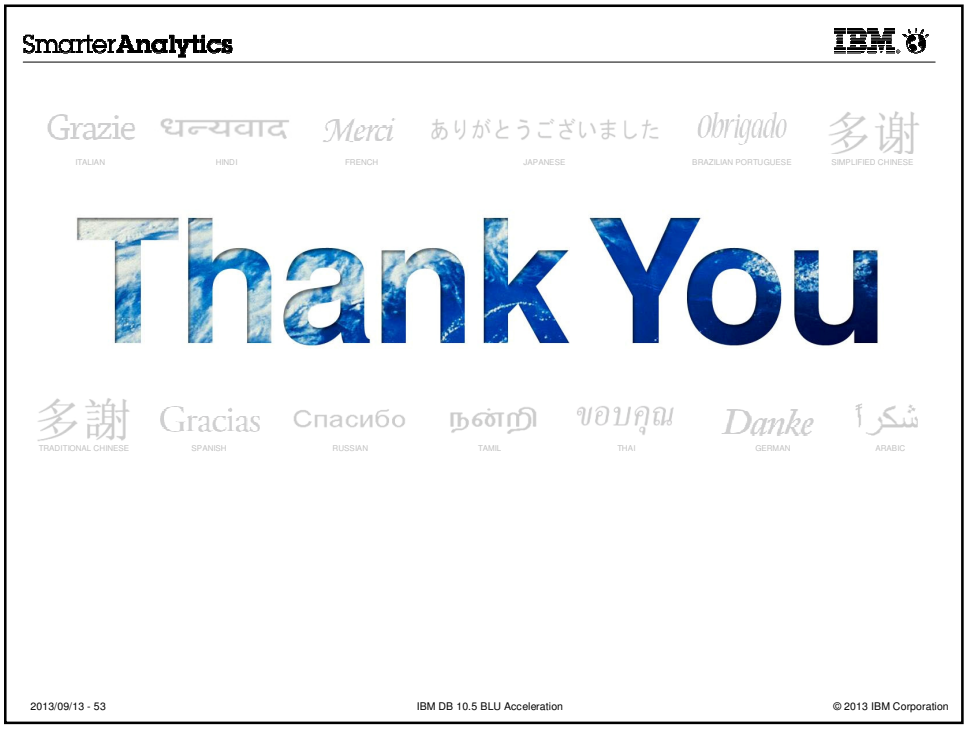




SmarterAnalytics

IBM DB2 with BLU Acceleration site

- **Announcing DB2 with BLU Acceleration microsite** www.ibmBLUhub.com


SmarterAnalytics
IBM
Grazie
ITALIAN
धन्यवाद
HINDI
Merci
FRENCH
ありがとうございました
JAPANESE
Obrigado
BRAZILIAN PORTUGUESE
多谢
SIMPLIFIED CHINESE
Thank You
多謝
TRADITIONAL CHINESE
Gracias
SPANISH
Спасибо
RUSSIAN
நன்றி
TAMIL
ขอบคุณ
THAI
Danke
GERMAN
شكراً
ARABIC
2013/09/13 - 53
IBM DB 10.5 BLU Acceleration
© 2013 IBM Corporation