HRG
HARVARD RESEARCH GROUP
Experience - Expertise - Insight - Results

# HRG Assessment:
## IBM DB2 10.5 with BLU Acceleration

Industry leading executives and IT professionals know that the effective use of Business Intelligence (BI) and Analytics are critical for continued business success. They recognize that a business needs to identify its own unique BI requirements and data sources as part of their periodic business planning and competitive differentiation activities. Information management, BI, and business analytics software gather, store, access and analyze corporate data. Industry leaders are increasingly finding that bringing Big Data into the mix dramatically impacts the effectiveness of Analytics as both a strategic and tactical business planning and positioning tool.

As businesses accumulate increasing volumes of data from their day to day operations, and a wide variety of other sources, transforming that data into actionable information is essential. With today's data management technologies and techniques no data is beyond integration.

When using DB2 with BLU Acceleration can be used for real-time data analysis to grow the business and investigate additional sources of revenue. "What-if" scenarios and ad hoc queries can be run on the data to quickly examine trends, provide analysis, and gain competitive advantage.

IBM DB2 10.5 with BLU acceleration (DB2 BLU Acceleration) takes In-Memory data management and analytics to the next level by optimizing these three key system components: CPU, Memory, and I/O. This holistic approach to optimization is a critical and necessary next step in the evolution of always on, always available, real time data management, analytics, reporting, and OLTP solutions. Before the release of DB2 10.5 with BLU Acceleration the best available in-memory solutions either were limited to specific hardware configurations (appliances or data base machines) or to the amount of active compressed data that could fit in memory.

The benefits of adopting a near real-time in-memory data analytics solution like IBM DB2 10.5 with BLU Acceleration include:

- Faster and more accurate identification of new business opportunities
- Better product innovation, shorter development cycles, and reduced time to market
- Better focused / more effective sales and marketing initiatives
- Improved customer satisfaction and retention
- More accurate and timely risk assessment and risk avoidance

## Use Cases

BLU is targeted at analytics and reporting workloads. SPSS, SAP BW, Cognos, SAS, Micro Strategy, and Business Objects are examples of software solutions that provide analytics and reporting functionality. These applications benefit significantly in terms of performance when used in conjunction with BLU.

The four targeted use cases for DB2 BLU are:

1) General analytics and reporting
2) Mixed workload environments where the customer has an OLTP system in addition to needing some analysis and reporting and where the analytics and reporting can be accelerated using BLU acceleration
3) SAP BW
4) Cognos ROLAP dynamic cube type applications

## Customer Interviews

The following customer interviews illustrate how DB2 BLU is being used and the value it delivers.

### **Interview:** Global Financial Services Company

One customer HRG interviewed, a large US based global financial services company, said they had worked with IBM DB2 BLU developers from the beginning and offered suggestions and insights regarding requirements, features, functions, and design.

This customer told HRG that they "have seen a significant improvement in query performance over traditional data stores." Depending on the specific query they have seen at least 10x faster query performance and in many cases better than that. For analytical queries that are in the sweet spot of BLU, BLU has according to this customer been "extremely fast." The applications they are running on top of BLU are analytics where the data is simply loaded and the queries are run without having to develop any indexes. The customer referred to this as "Load and Go" as a way of describing the simplicity and ease of use of IBM DB2 BLU.

Before using BLU they would run queries and then depending on the result and time of completion they would have to go back and tune the system to optimize performance. With BLU this is no longer a requirement and they are able to get a 10x or better performance increase without tuning and without having to develop indexes for each column in the data base.

They will be deploying BLU globally and today they are running BLU on thousands of systems. They started using BLU in production at the end of last year (2013).

Because BLU does not require DBAs to create and manage indexes the development and implementation of new applications is greatly simplified resulting in lower costs and improved productivity.

### **Interview:** US Based Health Care Insurance Company

They have been using DB2 BLU since late spring 2013 for analytics. BLU is not in production yet but they are testing it with live production data and doing production work. They will move BLU into production once they

have automated all of their data loads. They are using BLU for analytics and reporting related to bio statistical research, risk analysis, claims analysis, and trend analysis.

They load BLU with billions of rows of data which comprise roughly 1.5 TB of compressed data which represents 13 to 15 TB of uncompressed data. They are loading BLU with data from DB2 on the mainframe and from DB2 LUW.

Today users will pull data out of the warehouse and then use SAS to analyze these data sets. If they pull those same data sets into BLU and run the same queries the improvement in performance is amazing and on top of that they can use the same skills developed for DB2 LUW with DB2 10.5 with BLU acceleration. They are actively discouraging their users from pulling data out of BLU for analysis and are encouraging them to run their queries directly in BLU.

Using BLU the results from queries and the reports the business requires take significantly less resources and time to produce.

They cannot tune for every contingency that comes up and this has in part driven the need for creating, for example, SAS data sets for analysis. When this same analysis is done using BLU the flexibility and performance are impressive. They are working with users to get them to do analytics directly in BLU rather than doing data extracts for analysis.

One query for actuarials used to run for 20 hours before producing a result. By running this query in BLU it now completes in 1 minute. In another instance a query that ran in Cognos took 6 hours to complete and now when run on BLU it completes in ½ minute.

This positive impact on time, DBAs, and costs results in significant business value. With BLU users can now run more queries, ask more questions, and more effectively tune the business. They plan to use blue primarily for unknown or unspecified workloads driven in part by "what if" analysis, BLU will really benefit the data scientists and heavy duty research analysts. BLU will help them dig through massive amounts of data to find the valuable nuggets of information.

## **Interview:** Handelsbanken in Sweden

Since the financial crash in 2008 regulations for all banks have gotten more stringent and in Europe Basil II and Basil III have been implemented. As a result Handelsbanken had to build a solution that would be able to represent risk on a more detailed level and so the bank had to consolidate all data on a transactional level into a single risk analysis and assessment solution. The bank picked DB2 for data loading and data warehousing and SAS institute for the risk algorithms. This solution has evolved over the past 3 years. Approximately 2 years ago they entered the alpha program for DB2 10.5 with BLU Acceleration. BLU was chosen to help address some performance issues they were having as a result of the increased volumes of data they needed to analyze in order to stay in compliance with regulations.

They are using standard row store on DB2 to capture transactional data and then they are extracting one transaction day to calculate the risk and the result of that calculation will be stored in another DB2 table on BLU for analysis. Currently they are using SAS for analysis but they are working to use more DB2 BLU analytics functions instead of SAS in order to optimize performance. Everything is stored in DB2. Data staging is stored as row based and the risk calculation results are stored in BLU column store.

Using BLU in this manner allows highly competent people to focus on the right things rather than spending most of their time on lower level repetitive DBA tasks such as managing and tuning indexes. They no longer have to understand on a detailed basis the types of indexes the users are running. They can now focus on helping users get better analytical performance from the system. The DBAs role will change if you are using BLU to a higher level focus on risk. Once you put the data into a BLU table queries are automatically optimized and normal optimization of queries is no longer something that the DBA has to manage. The only way to improve DB2 BLU performance is to add more CPUs or to add more and faster storage. They have seen some queries that used to take hours to run now complete in seconds with BLU. This allows the reporting department to spend more time making sure that the reports are correct and to do more in-depth data exploration which in turn limits the banks' exposure to risk, fines, and regulatory sanctions. Using BLU also limits exposure to both as yet unidentified and unquantified risk.

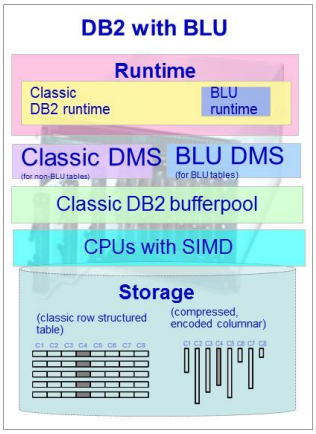
Last year the risk department was struggling to meet regulatory requirements. Today they have more breathing room and can focus on higher value things which will positively impact the banks internal strategy as they can now provide more frequent and detailed reports so that the management team can be more proactive. Now with BLU the bank is better able to address risk as it is expressed rather than being blind-sided by unidentified risk.

Using BLU is also quite beneficial when the bank is asked for information by one of the rating agencies. Occasionally in the past they would have to focus their efforts on satisfying such a request to the exclusion of other normal activities. Now with DB2 BLU the bank can easily meet such information requests with no disruption to their normal risk assessment and management activities. Missing the deadline for supplying such information to a rating agency could potentially result in a rating downgrade which could damage the bank's reputation. With DB2 BLU the risk department can now spend more time exploring their data in-depth and doing tightly focused "what if" analysis of potential risk scenarios.

For each day the bank calculates all of the risk for each of the transactions on that day. The result is that performing risk analysis on 100 to 300 MB of data each day will typically result in 50 GB or more of reports, and the results are then stored as active data for one month per regulatory requirements.

## IBM DB2 10.5 with BLU Acceleration

IBM DB2 BLU Acceleration delivers in-memory performance for queries with better TCO and faster ROI than other in-memory solutions and DB2 BLU Acceleration requires no proprietary hardware. The core DB2 BLU Acceleration technologies provide significant storage compression, extremely fast query speed, and real performance advantages when compared to traditional row oriented databases. Column oriented DBMS have been proven to outperform traditional row based database management systems on average by a factor of 100 times and DB2 BLU Acceleration delivers even better results with its column based BLU tables.




DB2 BLU optimizes and exploits the entire hardware stack (CPU, Memory, and I/O) through a combination of compression algorithms and enhanced storage management functionality.

DB2 BLU Acceleration does not require that all active data be in memory. This positions DB2 for use in the world of Big Data analytics where customers will not be able to store all data of interest in memory because the volume and variety of data will continue to increase and change particularly in light of the rapidly evolving Internet Of Things.

**BLU exploits memory** – to leverage BLU you do not need to make any hardware or memory acquisitions. BLU will exploit your current hardware (CPU registers, cache, memory, and disk) and make the best use of the available system resources.

**BLU exploits CPUs** – with DB2 10.1 DB2 can detect how many cores it has access to and map the multithreaded architecture of the DB2 engine to best utilize all available cores. This same technology is leveraged by BLU acceleration.

Row based tables and column based tables can coexist in the same DB2 BLU database with the same level of referential integrity that customers expect with DB2. From a query perspective such that customers can do joins across row based and column based tables as if they are traditional row oriented tables. From a compiler and optimizer perspective the optimizer is aware that there is a mix of BLU column oriented tables and normal row based tables and takes that into consideration in building an access plan, There is no significant change to the way that the DB2 process model works. There are no changes to the storage layer. DB2 uses normal tables for storage allocation and supports all of the normal page sizes from 4 to 32 k. DB2 utilities all work as expected regardless of whether you are loading data in a column or row store.

The continual data ingest, data streaming, and real time ware housing of data was introduced with the release of DB2 10.1 and is supported in the current release DB2 10.5 and DB2 with BLU Acceleration.

In the past the recommendation was that if you are using a 1TB database you should probably partition the database. With BLU that is unnecessary. Now with BLU those DB2 databases that were running IBM DPF for partitioning can now use BLU and run in a single server single partition environment.

BLU is completely integrated with the DB2 run-time, storage layer, and buffer pool services. Applications, management techniques, workload management, utilities, and data loads work as they would in previous versions of DB2 regardless of whether columnar or row oriented tables are used in the DB2 BLU database. This is DB2 so previous skill sets and experience apply and there is little or no learning curve to deal with when moving to BLU.

## DB2 BLU Compression

All of the Data in DB2 BLU is kept compressed in the database and in memory. When data is loaded into buffer pools or into memory it is kept compressed. Therefore BLU gets significantly more value out of memory than other in-memory databases. There is no physical or coded limitation as to how large a BLU environment can be.

DB2 BLU's order-preserving compression capability is one of its more important features and enables BLU to operate on compressed data. BLU maintains order in the compressed values which allows BLU to perform a range of operations on the compressed data where as other in-memory data bases and data base machines cannot do this. BLU reduces and in most cases eliminates the need for indexes. BLU lets you do predicate evaluation (>, <. or =) on compressed data because BLU compression preserves order in the compressed data.

On BLU you can run queries on compressed data without ever having to decompress the data. However, there are a few queries that will require decompressing the data such as running a substring comparison.

## MPP

IBM recently told HRG that support of MPP environments for full scale-out is on IBM's DB2 BLU roadmap. Currently because BLU does not support MPP you cannot parallelize data loads or backups. HRG expects to see

MPP support for BLU in the near future by leveraging DPF functionality. When MPP is available for BLU data load and backup times will be significantly reduced thereby facilitating much larger BLU environments without incurring any performance degradation. MPP is a shared nothing architecture that allows you to partition your data across multiple partitions in an SMP system or across multiple physical servers. For the initial release of DB2 BLU you cannot scale out by using an MPP cluster however HRG has confidence that IBM will address this constraint.

## Cache Local Processing and SIMD

BLU is designed to do query processing in cache and in the CPU register. Ram is actually the slowest part of the memory hierarchy of modern servers and the fastest part is the register of the CPU. IBM DB2 BLU does as much processing as possible in cache and the CPU register and minimizes access to Ram giving DB2 BLU a performance advantage over other in-memory databases that have to access all of the data in RAM. When you first load BLU on a system it identifies what the system resources are such as CPU, Cache, Ram, etc. and adjusts its internal algorithms to take the fullest advantage of those resources. The goal is to do as much as possible "in the chip" rather than in memory. BLU is dynamically adaptive and it will continue to adapt as CPUs evolve in the future from Power 7, 8, to 9 and the same holds true for Intel CPUs. Because BLU operates on data while it is compressed and because BLU works with data in cache it is able to outperform other in-memory databases.

Processing instructions in cache that is local to a CPU is much faster than in-memory (RAM). DB2 BLU loads as much compressed active data as possible into cache in order to gain advantage. When DB2 BLU is running on a SIMD capable CPU, BLU packs the CPU register to further increase query performance and through put resulting in an additional 4x performance advantage for those instructions.

If the CPU is SIMD enabled BLU will take full advantage of it to further increase query performance. However, if a processor is pre Sandy Bridge and not SIMD enabled then BLU can also run in an emulated SIMD environment and in this way at least deliver some of the benefits of SIMD.

Single Instruction Multiple Data (SIMD) lets multiple pieces of data be processed in a single step by packing 128 bits of data into each of the CPU's 4 registers for processing (parallel vector processing) in this way providing a 4x increase in performance for those instructions. BLU uses SIMD to reduce the number of instructions that it needs to execute when performing queries for things like scans, joins, groupings, and other operations which can comprise the bulk of many analytic workloads. Use of SIMD and register packing result in an additional 4x reduction in compute operations for these types of instructions. SIMD is available on all x86 Intel chips from Sandy Bridge going forward and on IBM Power starting with Power 7 running AIX and Linux.

## ETL

With BLU you can either use a standalone ETL (extract, transform, and load) tool or use the native ELT (extract, load, and transform) capabilities that are built into BLU. With ETL the data is extracted, transformed, and then loaded. With BLU ELT the data is extracted, then loaded into the database, and the transformation step takes place inside of BLU so you don't need to purchase a standalone ETL tool and ELT is typically faster than ETL for data loads.

## Mixed Workloads

BLU supports mixed workloads in the same database environment today as long as OLTP (row based) is handled in a separate table from analytics and reporting (column based). However, currently they cannot run concurrently in the same table. IBM is aware of this limitation and for this reason HRG expects it to be addressed later this year in a Fix Pack.

Today DB2 BLU can support a mix of row based and column based (BLU tables) within the same DB2 data store in order to deliver support for mixed operational, analytic, and reporting workloads. HRG expects that IBM will soon provide support for both row and column based DB2 data stores in the same BLU table.

## Data skipping

DB2 loads the compressed data columns required to answer a query and does not need to decompress data to do a predicate evaluation. With BLU data skipping brings only the data needed for a particular query into memory and skips over columns of data that are not needed. This is another benefit of BLU's columnar approach.

## BLU clouds

IBM has integrated DB2 BLU technology into the public cloud platform SmartCloud and announced an early access preview of BLU Acceleration for Cloud (bluforcloud.com) in late 2013. This service offers fast and self-service access for business analysts, data scientists, and other technical professionals by leveraging DB2 BLU enhanced analytics.

## pureScale®

HRG feels that pureScale® should be a key part of the DB2 BLU mixed workload strategy and that it will ultimately be used to enable a subset of members in a pureScale® cluster to run analytics and reporting and another subset to run OLTP and transactional applications. HRG fully expects IBM to offer support for BLU running on a pureScale® cluster to better support mixed workloads.

DB2 pureScale® provides continuous availability in the event of unplanned maintenance or a sudden hardware or software failure. Availability enhancements include; high availability disaster recovery (HADR) where multiple-standby databases provide availability functionality that is transparent to applications.

## Conclusion

Most current generation servers come with expanded memory, high speed storage, and SIMD capable CPUs and when used with DB2 BLU are well suited for big data analytics and BI customer requirements. Customers who are using DB2 BLU report better than 10x query performance improvements with many customers reporting even better results. DB2 BLU delivers excellent ROI at a reduced TCO fully optimizing and exploiting the server hardware stack thereby significantly speeding up queries. DB2 BLU has none of the memory requirement limitations nor hardware specific requirements of less capable in-memory appliances or database machines. DB2 BLU is uniquely well suited for use in cloud based data analysis and BI based service offerings. DB2 BLU deserves serious consideration if you want to fully exploit and benefit from the increasingly competitive and continually changing global business environment.

*Harvard Research Group is an information technology market research and consulting company. The company provides highly focused market research and consulting services to vendors and users of computer hardware, software, and services. For more information please contact Harvard Research Group:*

Harvard Research Group™
PO Box 297
Harvard, MA 01451 USA

Tel. (978) 456-3939
Tel. (978) 925-5187

e-mail: hrg@hrgresearch.com

*http://www.hrgresearch.com*