WHITE PAPER

# IBM DB2 for SAP: A No-Compromise Transactional and Analytic Database Platform

Sponsored by: IBM

Carl W. Olofson
Dan Vesset
Matthew Eastwood

April 2014

## EXECUTIVE SUMMARY

This white paper is written for SAP customers evaluating their infrastructure choices, discussing database technology evolution and options available. It is not easy to put forth a black-and-white choice as the SAP workloads straddle both real-time analytics and extreme transaction processing, and infrastructure choices can now be vast, given technology advancements around in-memory and faster processing speeds.

IBM offers DB2 with BLU Acceleration to provide extreme performance without limits for analytics and DB2 PureScale for flexible and comprehensive scaling capabilities for SAP ERP. According to IBM, these DB2 configurations running on Power Systems form the complete solution for all SAP customers. Considering all options in the market, and their maturity, IDC would recommend that SAP users consider adopting IBM DB2 as the data platform for the full range of SAP applications, both transactional and analytics.

## IDC OPINION

A vast number of businesses across a wide variety of industries around the world use SAP applications as the backbone for their mission-critical processes. IDC research indicates that these firms face continual pressure to plan for and deliver a computing platform that keeps pace with the processing demand that these SAP workloads exert. That is, internal business process owners and the customers, partners, and suppliers all relying on these applications continuously demand greater responsiveness and more access to critical data. SAP's industry-leading mobile extensions such as MEM ERP have further driven the volume, variability, and characteristics (e.g., changing read-to-update ratios) of these workloads. Then the business needs ever more rapid and wider synthesis of burgeoning data coming from both SAP applications and other sources (CRM, Web traffic, etc.) for analytics. The heaviest part of both these demands lodges directly against the heart of the system: the SAP ERP Central Component (SAP ECC) and Business Information Warehouse (BW)] *and* the database that runs them. To keep ahead and succeed, IT departments need to run these critical core database elements in the most effective, cost efficient, and streamlined manner possible. In a word, these transactional and analytical demands require more powerful, easier-to-administer databases than ever before.

To ensure that the database foundation of an SAP application handles these growing and variable transaction and analytic demands, without overburdening datacenter staff or database administrators (DBAs), organizations need to implement a database platform that does the following:

- Provides data management that is enhanced by a compression scheme that saves space, simplifies operation, and speeds processing in comparison with classic disk optimization
- Provides columnar organization for analytic data
- Supports dynamic scaling of transactional data with high availability through clustering
- Fits well into the overall datacenter management scheme, ideally belonging to the same product family as is used elsewhere for strategic database management

IBM DB2, with its optimization technology for transactions and analytics, presents as a leading candidate in this regard, and additional value can be gained by running it on IBM Power Systems.

This white paper discusses the database requirements for SAP applications as the accelerating pace of business drives enterprises to seek support for extreme transaction processing and real-time analytics. It outlines major developments in database management systems (DBMSs) technology that address these requirements. It shows how DB2, with PureScale, delivers extreme transaction processing and how DB2 with BLU Acceleration delivers the power for real-time analytics. It also shows how DB2 on Power Systems overall addresses the needs of SAP customers without forcing them to compromise in the datacenter organization and operation.

## SITUATION OVERVIEW

By now, everyone is well familiar with the issues involved in the accelerating pace of business: the 24 x 7 business availability requirement, the speed of transaction processing caused by the digitization of practically everything, and the hyper-competitiveness of businesses caused by the availability of customer tracking and sentiment data coming from mobile devices, Internet click streams, and various Big Data sources (including, but not limited to, social media). The pressure to make smarter decisions in the moment, to transact business faster, and to eliminate downtime has resulted in two key areas of functional enhancement over business applications of the past. These are extreme transaction processing and real-time analytics.

## Required Database Technology for the Real-Time Enterprise

### *Extreme Transaction Processing*

Formerly, businesses paced their activity, in part, on the rate at which their computer systems could process transactions. This is no longer acceptable. Instead, computer systems must process transactions at a rate that keeps up with the pace of business. As a result, the database, in combination with an architecture designed for performance, must enable the enterprise to do business at the fastest practical rate from the perspective of customers, partners, and suppliers.

To meet this requirement, the database must execute even complex transactions at a speed that ensures that it creates no drag on the business process. Such a rate may require transaction speeds that range from sub-second to milliseconds. For most DBMSs, this has meant a radical redesign in how transactional data is handled to ensure scalability with maximum availability and throughput.

### Real-Time Analytics

In addition to the need for fast transactions, businesses also require smarter transactions. Speeding up transactions can result in multiplying mistakes unless more intelligence is introduced into the process. Executing relevant complex queries as business is being transacted can enable more efficient business operations and more sales. For instance, if a sales associate can find purchasing patterns (through high-speed complex query) based on the behavior of customers matching the profile of the customer standing right there, that sales associate can make a decision in the moment that leads to a more successful sales interaction. The current collection of data can yield business intelligence that results in richer, context-sensitive interactions with customers, partners, and suppliers, resulting in greater operational efficiency and better exploitation of business opportunity.

The ability to deliver timely business intelligence as quickly as tactical decisions need to be made, and with the latest data, is called "real-time analytics" because the analytics are so time sensitive that they must be delivered within the effective time interval of actual business activity. They require specific DBMS technologies as discussed in the sections that follow.

## Shifts in the Technology Landscape

These business requirements, combined with changes in technology from a capability and cost perspective, result in the development and delivery of a substantially transformed DBMS, which is intended to enable both extreme transaction processing and real-time analytics. To understand how this is possible, one must consider recent developments in the evolution of computer technology.

### Clustering for Scalability and Failover

The ability of a database to scale with flexibility, according to the volume of data processed, is key to enabling enterprises to keep up with the pace of business. The essential approach to cost-manageable transactional database scalability has involved clusters. Database cluster architecture has been around for quite some time, but the techniques required to enable cluster-based transaction processing while efficiently managing locks to avoid delays are available in only a few DBMS products, yet such an architecture is critical to the success of a dynamically scalable transactional database. A shared storage cluster that handles locks and shares buffer memory efficiently across nodes without any single point of failure is essential to delivering a transactional database server that can both scale and ensure continuous availability.

A number of major DBMS vendors claim to deliver this technology. In evaluating these claims, it is important to consider how they are achieved and the extent to which vendors make good on their claims. Key factors to look for include:

- All nodes active and able to contribute to the problem at hand while also acting as failover nodes if called upon
- No bottlenecks or internode conflicts that can slow or stop the whole cluster
- Simplicity in administration including DDL composition, application SQL, and adding or removing nodes
- Ability to add extra resources to handle "spikes" in demand and scale back later

### Large Memory Models and Fast Processors

Formerly, relatively limited amounts of available direct random access memory (DRAM) and the comparatively slower speed of processors constrained DBMS performance. Over the past 10 years, processors have increased their power substantially and can now use 64 bits to address memory. This means they can handle much larger memory spaces. In addition, each processor carries multiple cores, ensuring that each processor can execute as many purely parallel processes as they have cores. At the same time, the cost of both processors and DRAM has been coming down, main bus architectures have improved to deliver better access speeds against large amounts of memory, and enhanced processor instruction sets enable more efficient use of the processor cache, thereby reducing the number of DRAM accesses. All of this has dramatically increased the size and speed of in-memory processing.

### *Moving from Disk-Optimized to Memory-Optimized Database Technology*

Because of the inherent historic limits on memory, processor, and address space in previous years, DBMSs have concentrated data management on manipulating spinning disks to optimize their operations. As a result, traditional DBMS technology exerts a tremendous amount of its processing activity on managing data on disk, sustaining disk-mapped buffers in memory, and addressing data based on disk access. It has now become practical to manage data based on its organization in-memory rather than on disk. The former approach to data management is "disk optimized" database management, while the latter approach is "memory optimized." Memory optimization does not just mean large or cleverly managed caching or buffering but a fundamental change to database architecture from the storage layer on up that is optimized for in-memory operations. The benefits of memory-optimized database management include simpler administration and much faster database operations, especially query processing.

### *Memory-Based Data Organization*

Memory-resident database management enables the database kernel to perform basic database operations with fewer instructions, often fewer than one-tenth the number required for a disk-optimized database. For example, consider a hypothetical simple data retrieval problem. To retrieve data, a disk-optimized system must maintain database keys that enable the DBMS to determine the page location of the data; determine whether the data is in buffer, and if not, identify where the page resides on disk; retrieve the page and swap out another page to make room for it; and then traverse the page to find the associated record. By contrast, the memory-optimized DBMS retrieves the data by simply converting the database key (actual physical address) to a memory pointer and goes directly to the data location in memory.

This means complex queries complete in one-tenth the time of traditional disk-optimized systems, and each server node can handle six times the load that traditional disk-optimized systems could. Organizations implementing this technology save as they speed transactions that have embedded queries, increase query capacity, and reduce infrastructure footprint, enabling the enterprise to do more business and make smarter, timelier decisions.

In addition to the improved efficiency of the DBMS kernel, the memory-optimized DBMS reduces the IT staff labor involved in managing the database. With a disk-optimized DBMS, the database administrator must develop a storage management strategy that usually involves such tasks as careful assignment of tablespaces to files on volumes, distribution of table data across volumes using hash key partitions, using data striping and mirroring for better availability, and maintaining global and partition indexes for faster random data retrieval. In addition, there is a regular need to perform checkpoint backups, unload and reload operations when table definitions change, defragment the database, and index rebuilds. A memory-optimized database eliminates all of these tasks. A simple highly compressed checkpoint backup (or snapshot save) to disk is generally the only operational task, and often this can be done without quiescing the database. The only other disk writes required, in general, are writes to a change log or logs, used for rollback, roll forward, and recovery.

For an analytical database, a memory-optimized architecture offers considerable advantages. For a large transactional database, however, a disk-optimized approach is often preferred because it can offer better overall transaction performance, reliability, and availability than other options. So it is necessary to choose the right tool for the job.

## Transactions Versus Analytics

Transactions generally require retrieval and insertion of whole rows of tables, sometimes in the context of fairly simple queries. Also, transactions generally require insertion or update of a small number but query of a larger number of tables. Analytics, on the other hand, involve complex queries involving selected columns, sometimes applying mathematical selection criteria across multiple tables, and nested to multiple levels of depth. Because of the differing requirements of transaction processing and analytical processing, the DBMS must organize the data differently for these two classes of processing. In the transactional case, the data is generally kept in row format. For analytical processing, a columnar organization is preferred, and in a memory-optimized database, this format can be further optimized, using compression and careful placement of sequentially accessed data on specific memory address boundaries for additional performance benefits.

### *Transaction Optimization Technology*

Row-oriented table management with memory optimization in the database server is the best approach for transaction optimization. Because most transactional database applications tend not to reference the same data once written from the same stage in processing, managing that data on disk is the most common approach to holding the data rather than keeping it in memory. To deliver flexible scalability — so that as transaction volume increases, the server can be expanded to keep up with demand — a shared disk clustering technique is called for.

### *Analytics Acceleration Technologies*

For analytics, a variety of techniques are available to further optimize data retrieval in a memory-optimized database.

#### Columnar Organization and Compression

As previously mentioned, a columnar data organization is optimal for most analytic processing. This is for two reasons: most analytic queries involve criteria that reference only a small subset of the columns for the tables involved in their operations, and by looking at a column of data organized together, the DBMS can take advantage of organization and compression schemes that reduce query time and eliminate the need for indexes.

#### Vector Processing

Vector processing, in general, is a programming technique that involves setting up vectors (one-dimensional arrays, or serial lists of contiguous values) and enabling single instructions to operate on those vectors using an approach called SIMD (single instruction, multiple data). With SIMD, work can be spread across multiple threads and processor cores to greatly accelerate performance, helping decision makers get the answers they need quickly. The more cores, and threads per core, the greater the performance. For example, IBM POWER8 processor-based technology offers eight threads per core, twice that of other offerings. This larger number of threads should have the ability to provide even greater performance.

When combined with columns, this means organizing the column data in memory so when the processors act on the data, it is loaded into the processor cache in such a way that each process can act on the maximum number of values before going back to DRAM. This technique can speed up searches by 10 times or more. When combined with optimal compression using a strategy that ensures the data need not be decompressed before performing comparative evaluations, the performance gain can be much greater.

### The Idea of a Single Database

An idea that is being promoted by some vendors is that it would be ideal to keep all enterprise data for both transactional and analytical work in one database. At first blush, this seems to make sense: it seems simpler, tidier, and easier to manage. When we look at it more closely, however, we see that the vision of a single transactional and analytical database for the entire enterprise is really highly impractical.

To begin with, as we have already seen, the technologies involved in optimizing transactions and complex analytical queries are very different. We have seen that scalable clusters of nodes that share resources and *manage the data in rows* make the most sense for transaction processing, while keeping *the data in optimally compressed columns* delivers great benefits for analytics. The reasons for managing analytical and transactional data separately are simpler and more fundamental, however. They include the following:

- The data needed to execute a transaction and the data needed for analytics are not the same, even for the same domain. Transactional data includes details not important for analysis but vital for transactions, while analyses need contextual data that is irrelevant to transaction processing. Combining them results in a messy and excessively complex schema.
- Analysts are constantly tweaking the data they analyze and need the freedom to change the schema without fear that those changes may adversely impact transaction processing.
- For SAP application users, there is currently no overlap, at least in the first instance, between the data managed by SAP Business Suite and the data handled by SAP Business Warehouse or the SAP analytical tools. Although SAP projects some transactional data into relational tables for analytics, other relevant transactional data must still be converted into a form that makes it useful for analytical queries and operational reporting (and visualizations).

There are also operational concerns. Even if a database can keep the mainly transactional data in rows and the mainly analytical data in columns, governing both with the same schema to support mixed queries, there are obvious compromises in performance and efficiency that arise from such an approach, and some operational interdependencies that can develop between mainly transactional and mainly analytical data from, for instance, foreign keys, could defeat the optimizations designed to make each mode of operation effective.

Finally, this is a solution in search of a problem. Those who promote the single database approach say that the batch-oriented extract, transform, and load (ETL) approach to moving data from transactional systems to the analytical system results in unacceptably long latency that prevents real-time analytics. This is true, but batch ETL is not the only option. Streaming data movement, driven by technology such as change data capture (CDC), can ensure that the analytical database is current enough for the analyses required.

### *CDC Versus ETL*

There is also a difference between analytical data used for real-time analytics that are tied to transactions and decisions "at the point of impact" and analytical data that is used for analysis of a period of time or other key dimension. This latter form of analysis is key to longer-term, strategic decision making. While the former case may be covered with a high-speed analytical data store, the latter may require a data warehouse. The former may call for CDC-based dynamic data movement from the transactional database to the analytical, while the latter calls for a scheduled ETL operation involving data from multiple sources for loading into a data warehouse.

All this having been said, it does remain true that managing all the relevant data under a single DBMS product family does make eminent good sense, especially if the vendor can offer means for managing the data in both classes of database in a coherent way.

## IBM DB2 for SAP

In considering the database requirements for successive versions of SAP Business Suite, SAP Business Warehouse, and SAP BusinessObjects applications, the user has a number of possible options, all of which are reasonable. The challenge is to choose the right option for your needs. DB2 for Linux, Unix, or Windows (LUW) is one such option and has several differentiators that should be taken into consideration.

### *IBM DB2 PureScale for SAP Business Suite*

IBM DB2 PureScale is designed to deliver the kind of flexible scalability required by a comprehensive ERP application system such as SAP Business Suite. It runs on a shared storage cluster that allows users to adjust the number of nodes to match processing demands with almost no effort at all. Its key differentiating features include:

- Scaling that has no inherent limitation, enabling the addition of nodes with a few simple commands
- Centralized lock management that avoids the problem some shared storage database clusters have with lock management collisions that can freeze the whole cluster
- Centralized buffer management that enables all relevant nodes to have current data in memory, thanks to a technique that refreshes their buffers without affecting their operations or performance in any way — a technique called remote direct memory access (RDMA), which enables the update of a remote system's memory without causing an interrupt
- The DB2 approach to managing shared data that obviates the need for any kind of partitioning to achieve better performance, unlike some alternative technologies

These benefits are, in part, the result of decades of IBM experience in delivering parallel processing technology to customers and are made available to applications with no changes required to the application SQL.

### *IBM DB2 with BLU Acceleration for SAP Analytics*

As previously discussed, transactional workloads call for one set of optimizations, centered on scalability and maximum transaction throughput. Analytical workloads call for another set of optimizations, aimed at delivering very short response times for large and complex queries. The DB2 technologies that address the needs of SAP analytical application users, including SAP Business Warehouse, are DB2 with BLU Acceleration.

### Full Acceleration for Analytics

IBM DB2 with BLU Acceleration for analytics uses a compression technique that goes beyond standard dictionary-style compress (i.e., substituting a token for each value and storing each value exactly once in a compressed form based on the range of values and value types that exist in that column) by establishing an evaluative order for the tokens so that the actual values need not be referenced and sizing the tokens based on the frequency of the values they represent, where the most common values correspond to the shortest token (1 bit). This enables the DBMS to hold much more data in memory than is possible for most DBMSs while enhancing query performance.

Performance is further boosted by using vector processing, as outlined previously. While vector processing is not unique to DB2 with BLU Acceleration, the combination of this technique with a compression approach that includes columnar organization, data skipping, and cache optimization is compelling.

### In-Memory Speed Without the Limits of Memory

Some memory-optimized DBMSs are designed to execute queries only for data in memory. Since a large fraction of the data in a given database is rarely accessed, this is cumbersome to manage. IBM DB2 with BLU Acceleration is designed to optimize the use of both memory and disk, so that the entire database need not be in memory in order to realize full in-memory speed. Also, some memory-optimized DBMSs can only operate on a table if the whole table is in memory. IBM DB2 with BLU Acceleration manages table data, ensuring that full memory-optimized speed is achieved without requiring all the queried data to be loaded in memory.

## Optimized for IBM Power Systems

DB2 PureScale and DB2 with BLU Acceleration are designed for use on either x86 or Power Systems. To add extra benefit, however, DB2 designers have collaborated with Power Systems hardware engineers for years to ensure that Power Systems, especially POWER7, POWER8, and above, can take full advantage of DB2 and vice versa.

These designers have tuned the hardware/software environment to ensure that Power Systems' immense number of cores operate at maximum capability to complete the BLU Acceleration query at the speed your business demands. Built on a heritage of strong resiliency, availability, and security, Power Systems, with the newest POWER8 processor-based technology, provides larger cache sizes, greater memory bandwidth, and more parallelism (up to 12 cores per socket, with 8 simultaneous threads per core) to satisfy your most demanding transactional and analytic workloads. The POWER8-balanced design optimizes core utilization and reduces or eliminates CPU core wait times, resulting in faster query to solution times.

### *The Benefits of Software-Hardware Optimization*

DB2 with BLU Acceleration for Power Systems tuning means that those enterprises with Power Systems can leverage their existing hardware infrastructure to add BLU-accelerated analytics to their SAP data stores. This can help organizations add more analytics without expanding server infrastructure footprint.

### *Non-Disruptive for an Existing IBM Power Environment*

In addition to the benefits outlined previously, it should be noted that, unlike other options for managing SAP application data, the DB2-based approach fits seamlessly into an environment that is otherwise based on IBM technology. For instance, if one is considering the IBM Power platform, this choice makes perfect sense, because one can manage the entire environment holistically, rather than treating the SAP application resources collectively as a special case. BLU Acceleration is not simply an add-on; support for this key feature is built into the kernel and storage management layers, which means that administering a database on DB2 with BLU Acceleration is a straightforward matter for DB2 DBAs. Also, DB2 with BLU Acceleration delivers memory-optimized database speed and efficiency without requiring the user to move to another platform, such as Intel Xeon, as another approach might mandate.

### *A Cost Effective No-Compromise Alternative to Other Mixed Workload DBMS Options*

Some DBMS approaches that support both analytical and transactional SAP applications involve emerging technology that is rapidly evolving and requires special expertise to manage. Others prove prohibitively expensive when compared with the DB2 on Power Systems platform and require a great deal of tuning. Not only does the DB2 on Power Systems platform offer ease of administration, speed, and scalability, but it is a proven technology supported by a broad talent pool around the world.

### *In Line with IBM's Future Database Platform Direction*

Committing to DB2 on Power Systems also ensures that the datacenter is well aligned with IBM's future database platform direction. Assuming other applications and databases are involved besides SAP, the ability to coordinate database operations and concentrate on DB2 skills is an important benefit for future interoperability and decision making.

In addition to the ability to link different DB2 databases together directly, IBM offers a range of options for database coordination using products in the InfoSphere product line.

## FUTURE OUTLOOK

There can be little question that memory-optimized technology represents the future of DBMS. Many vendors have some form of this technology today, and many more have plans to bring it out shortly. Few of these, however, are suitable data platforms for SAP applications. Some of them can operate on IBM hardware today but may not represent the best choice for such a future environment. It is a certainty that IBM DB2 on Power Systems will always be an optimal choice.

## CHALLENGES AND OPPORTUNITIES

Although IBM's memory-optimized technology, including its compression, vector processing, and columnar organization, is among today's industry-leading technologies, a number of leading DBMS vendors are continuing to develop their technology. IBM must continue to innovate in these areas to maintain its competitive position.

## CONCLUSION

The world of business computing is undergoing radical changes, driven by commercial opportunities and competitive pressures and enabled by transformative changes in the power and reduced cost of processors and memory. Trends in areas such as Big Data, mobile computing, social media, and instant gratification ecommerce are creating a greater need for applications to support fast analytics against large data collections, smart extreme transaction processing, and real-time analytics. SAP is responding to these pressures by evolving its applications appropriately. The DBMS that supports these applications must do likewise.

The DBMS platform that succeeds in doing so must satisfy these key requirements:

- Deliver performance levels that are achievable only through optimal parallel processing architecture, augmented by the flexible scalability of a shared storage cluster.
- Support more complex queries with less administrative overhead while yielding greater responsiveness, made possible only through a columnar-compressed organization with optimized execution by techniques such as vector processing.
- Enable the coordination of transactional and analytical data in real time, or near real time, in a simple and easy-to-manage way.
- Fit seamlessly into the overall strategic organization of the datacenter, without introducing exotic physical systems that require special care and attention, thereby implementing systems that can be managed as part of a general approach to systems and storage management.

DB2, with its two configurations — the DB2 PureScale (for scalable transaction processing) and DB2 with BLU Acceleration (for real-time analytics) — meets all these requirements. When you add in the Power Systems dimension, you include technology engineered to give DB2 an extra performance boost.

For SAP users looking for a leading data persistence platform, IBM DB2 PureScale coupled with IBM DB2 with BLU Acceleration on Power Systems meets all these requirements. For SAP users looking for scalable, robust, proven DBMS capability with all the speed of in-memory processing, IBM DB2 with BLU Acceleration on Power Systems is a strong alternative to any other approach available.

## About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

## Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com

---



This document was developed with IBM funding. Although the document may utilize publicly available material from various vendors, including IBM, it does not necessarily reflect the positions of such vendors on the issues addressed in this document.