THE CLIPPER GROUP
Navigator™
Navigating Information Technology Horizons℠
Published Since 1993 Report #TCG2012030 December 21, 2012

Addressing New Business Analytics Challenges — When the IBM zEnterprise Really Makes Sense

Analyst: Stephen D. Bartlett

Management Summary

Putting aside the issue of affordability, wouldn't it be nice if we could have an automobile, more generically a motor vehicle, for every day of the week, or for every occasion. On pleasant sunny days when the daily cares of everyday life can be put aside, it might be nice to have a two-seat roadster with a sonorous exhaust note that loves curves and to explore the countryside. When it comes to hauling the soccer team or delivering your offsprings' household effects to their college dorms, moms and dads, as much as they might prefer a bit of "wind-in-the-air", perhaps by necessity must choose a roomy, safe, and convenient minivan. On the other hand, a vehicle used primarily for business could range between a full-size, somewhat plush, four-door sedan to a half-ton pickup with four-wheel drive. The point is that each of these examples might be considered to be too specialized and, thus, each is, in some ways, a compromise. What you really may want is a vehicle that does many things very well and offers few compromises.

As IT professionals, we have to deal with this issue repeatedly. Should we have a separate and unique server platform for each application class – web serving, online transactions, enterprise database repository, business analytics, high-performance computing, etc.? Or do we endeavor to force-fit all workloads onto a single server class, thus possibly accepting compromises in performance, manageability, security, and flexibility, and maybe even cost, to name a few. To examine this issue, let's take a look at the "next big thing" – *mission-critical business analytics*. (Cloud computing may have come to mind first, but this concept already is a reality.)

In this paper, we will use a definition of business analytics (BA) paraphrased from *Wikipedia*, to wit: *the process of developing optimal or realistic business decision recommendations insights derived through the application of statistical modeling and analysis of an existing or simulated database.*[1] The analytics may be simple or extremely complex, and may range from addressing immediate operational (tactical) needs to the less urgent development of a strategic plan. **The ultimate goal of BA is to gain competitive advantage and to optimize business outcomes.** This is a significantly broader definition of what has been traditionally known in the past as *business intelligence*, most notably data mining or data warehousing, as these concepts now are subsumed into business analytics.

You might be asking yourself: *So what do we want our business analytics platform to be?* Hey, wait a moment. Shouldn't you be asking: *What problem(s) are we trying to solve?* (This always is best couched in terms of a business need.) Then you might ask: *What are the essential requirements for a modern BA solution to address this need?*

There are many ways to skin this cat; ask any IT vendor, including IBM, whose answers are the focal point of this bulletin. My goal here is to describe and illustrate that there is a server platform, the *zEnterprise* mainframe in its most current iteration, which does not compromise in any of the dimensions required of a comprehensive BA solution, now or in the future. Please read on to learn how and why.

IN THIS ISSUE

[1] See http://en.wikipedia.org/wiki/Business_analytics.

The Evolution of Needs

For a start, let's imagine that the first business analytics application (way back when) was a suite of programs written in COBOL that sifted through a file containing all the financial records for the last month that had been stored in serial sequence on magnetic tape. This application had three fundamental attributes.

- *Volume* – there were many records to read and sort.
- *Variety* – the reports were becoming more complex because of changes in the data and new demands from management and all the reports had to be available within three days of the last business day of the prior month.
- *Velocity* – as was common in the past, time had to be pre-scheduled with the data center operations manager, with the hope that no reruns would be required.

This is a decades-old example, but clearly more modern business analytics solutions share the same dimensional aspects, frequently referred to as the "3Vs"; only the scale of the metrics really has changed.

Where had BA solutions evolved by the mid-1990s? To protect availability, response time, and security of the enterprise's mission-critical transaction-driven systems from the negative impacts of query- and report-driven production applications, periodic, maybe daily, batch processes were used to create a centralized *enterprise data warehouse*.[2] In those batch processes, various indices and summarizations would have been created to facilitate its usability. To speed responsiveness to their particular needs and to avoid sustained dependence on the data center personnel, individual business units, such as finance, manufacturing, sales, marketing, and development, needed to create their own *data marts*.[3] Quite frequently, these resided on platforms independently operated and controlled outside the purview of enterprise data center.

In creating these departmental data marts, further summarization (extraction and transformation) would take place, consuming ever-increasing resources which, accordingly, frequently were relegated to off-peak hours.

However, the window of opportunity to do this preprocessing became smaller and smaller as the demands on the mission-critical systems continued to increase and the number and size of data marts proliferated. Users typically would access their respective data marts from their desktop and, though predictable standardized reporting remained essential, ad hoc query became the dominant load on the underlying data warehouse and data mart infrastructures. At the same time (the 1990s), the numbers of users, maybe only a hundred or less employees, were easy to identify and authenticate.

There were a number of consequences of this model, most of them not good. For instance, each function probably had a different data model because it was designed solely to meet the needs of that department. Frequently, different and incompatible hardware may have been used, and they were likely to have different analytic tool sets for their respective reporting and query activities. Most notably, the data mart (by definition) was out of date soon after it was loaded and, thus, lacked consistency with other data marts. From the overall enterprise perspective, costs were uncontrolled, redundancies abounded, "data fiefdoms"[4] were created, and there was no assurance that the business processes were being optimized to meet the overall goals of the enterprise. Predictably, this model quickly became over-stressed and suffered from runaway costs that were not easy to reduce.

There is a strong case indicating that creating and replicating more database extracts does not solve the problems of slow performance, load balancing, and poor utilization effectively, often exemplified by hot and cold spots on compute and storage servers and on demands for network bandwidth. More images may also mean moving and managing data across several diverse database management technologies. Users are increasing their demands for responsiveness and quicker turnaround time, and the flexibility to acquire additional resources quickly to meet new business initiatives.

Where Are We Now?

Jump ahead to the current decade, *what are we now likely to be observing?* Data volumes are growing at high rates; in larger enterprises, 10TB data stores are a thing of the past, 100TB ($100x10^{12}$) is more likely, and petabyte ($1x10^{15}$)

[2] A data warehouse is a database used for reporting and data analysis. It frequently is an extraction from multiple sources, including operational data stores, that have been transformed (normalized) and uploaded for user access.

[3] A data mart is most frequently a subset of a data warehouse which has undergone further extraction and dimensionalization for a specialized purpose or particular business unit, e.g., sales.

[4] Even within these fiefdoms there also would be competing constituencies: the users, the DBAs, and operations that worry about arcane issues such as a back-up system for disaster recovery.

volumes will become more typical within this decade.[5] They are being driven by applications growing in size, complexity, and their "reach" into the enterprise, plus the need to be prepared for spot and disaster recoveries via backup, snapshots, etc.

Operational transaction data, i.e., data that will eventually need to be analyzed, often is scattered throughout the enterprise's systems. Although 70% or so is still at the primary data center, frequently some of it is dispersed geographically and, at best, may be loosely connected, both physically and logically, with other data. Even if centralized, these transaction-driven systems and their databases likely are on different platforms, for example: Oracle's *MySQL* or Microsoft's *Sequel Server* on x86 servers, *Oracle DBMS* on *UNIX*-based servers, and *DB2* or other DBMSs on a mainframe for the larger application loads.

Additionally, volumes are being driven by a user base that not only is increasing in size and scope but also in its diversity as the Internet and social media expands the reach of the enterprise and the reach of others into the enterprise, such as suppliers and, most especially, customers.[6] With all of this comes the demand for access through many types of mobile devices, those outside the enterprise's private network, thus presenting a number of new security issues. It gets worse, as services must be available 24x7 in all time zones, as a result of globalization. Thus, it is easy to conclude the level of concurrent activity has increased many times.

The enterprise's data also is becoming more diverse; no longer is critical business data wrapped in a nice, neat package ready for loading into structured tables. New data types must be recognized and stored from automated processes (machines), sensors of all types, graphics, audio, and video streams, web services interactions, and, of course, messy social networks.[7]

With this diversity there emerges a fourth "V" – *Veracity* – to be added to the three listed previously. *Is the data "true"? How will your system deal with data from your customer-facing Web portal where the user may not be bounded by many of your enterprise's data rules?* Some data must be analyzed for its level of validity, some level of standardization and normalization may need to be set, and exceptions need to be managed *before* they are accepted into the database and, of course, this should be automated and continuous.

Increasing in importance and intensity are issues related to regulatory compliance, whether they are imposed by government agencies, often at several levels, or mandatory industry standards. Financial institutions, insurance companies, and public utilities are familiar with these needs but no sector is immune from these requirements. On many agendas are personal data privacy laws, e.g., HIPAA privacy and security rules, which will pervade all of an enterprise's data management and administrative requirements. More recently, but of no lesser importance, is the emergence of environmental regulations, particularly at local levels.[8] There are more than just cost constraints to unbridled expansion. The common thread here is that regulatory compliance must be demonstrated and reported continuously and may be subject to frequent audits.

The Business Dilemma

How has this evolution affected the ways that business is conducted in the enterprise? It's not as if there is an information vacuum – there are many DBMS in place, one or more copies; useful tool sets are available on user's desktops; all functions and operations as well as executive management are being served; and value is being derived from the enterprise's operational data. Nonetheless, disaggregated structures have led to several problems that are poorly addressed, if indeed they are acknowledged at all.

Performance Problems

The most visible of these is timely execution (performance, as we tend to call it in the IT world) of which there are multiple aspects. It would be short-sighted to limit the view simply to response time or turn-around time for queries. Also, it embodies consistency of response times, and optimal and of balanced use of all resources, which always is a valid IT concern. Moreover, all enterprise business processes are not equal. *How is optimal*

[5] Is this *Big Data*? It's a fuzzy construct. We always have had big data, i.e., the foreshadowing of greater and greater growth rates of our data volume. The current manifestation is not so much about pure byte volumes as it is about the diversity of data types, the diversity of the users, the complexity of the needs, the speed of delivery required, and the need to "cut the wheat from the chafe" – *finding the data with value*.

[6] The difference between the number of enterprise employees and the number of its customer clients is likely to be several orders of magnitude.

[7] If someone says something negative or derogatory on Facebook, LinkedIn, or any other social media platform about your business or its products, you might want to know about it; the quicker the better.

[8] Plopping a new 50 MW data center and, for backup power, a dozen or so diesel motor-generators in the backyard of some little town can be expected to result in more than a little concern from the town overseers and residents.

performance assured for those processes that are most critical and valuable to the enterprise? In summary, large, heterogeneous, decentralized structures inherently are more difficult to manage and control in the quest to *consistently* meet performance goals without outstripping the budget.

Data Proximity Problems

Another quite common issue is information latency – which might be called *the data proximity dilemma*. There exists a significant time lag between the collection of information, be it through customer-facing transaction-driven processes, vendor/supplier feeds, logistics, or similar systems. To be accessible for analytics applications and their users, these data have to be extracted, transformed, and loaded periodically into a data warehouse or something similar. For some uses, monthly updates might be adequate, but more likely weekly or daily, or even more frequently describe today's requirements. The most obvious consequence is that in the second following the extraction of the data, it is "old" and stale and, therefore, its value is diminished.[9] **From the IT management perspective, the issue is that there now are at least two different copies of the data, a duplicate that requires additional storage and "tending".** Now, the boss wants answers to questions on why, how big, and for how long?

Security Problems

This then begets another set of issues. The first is the related to the security of the data. Once the data has been extracted and moved, *how is access to it being managed and protected? Is the data secured by means of encryption or other means? And is the link between the original database and the data warehouse secured as well?* Related to these are the issues of data governance – usage, quality, and consistency, and lifecycle management – which, in turn, becomes about archiving, managing, and retention in conformance with business and compliance policies. Overzealous decentralization of active business data exacerbates these problems and burns up (or out) the enterprise's technical resources.

Inflexibility Problems

Another important problem is the challenge of to meet new requirements with a high degree of flexibility. *How well and how quickly can the infrastructure supporting the enterprise's BA applications adapt to support a new or greatly expanded business process? How easily and quickly can new resources – servers, storage, network, and supporting software, and technical support – be marshaled to address these new requirements? What is the turnaround time from acceptance of a new BA requirement to the delivery of value to its constituency?* All the while, IT must keep in mind that each new interconnection increases the complexity of the infrastructure geometrically and, hence increases the difficulty of managing it. **Unfortunately, properly managing growth may require re-engineering the infrastructure. Thus, a reasonable goal of the enterprise might then be to eliminate or avoid BA solutions that compound complexity.**

Then There is Always the Cost

Whatever your choice of a BA solution that will meet your business needs effectively and flexibly, it necessarily must be subjected to close examination of what drives its costs. A close examination of the total cost of operation (TCO) is the best approach to addressing this issue. The most tangible elements are the quantity and qualities of service of servers, storage, and network switches; building in redundancy; multiple copies of operating systems and involved virtualization technology; analytical software and management tools; and the power, cooling, and space required to accommodate the solution. Less tangible, but no less important, are additional costs that must be included in the TCO analysis, including the complexity of the BA solution (driven by the combination and diversity of the tangible elements above) and the personnel required such as technicians, architects, database administrators, policy administrators, etc., and their management.

In summary, the overall effect to the enterprise's IT community of the problems and issues outlined here are:

- The lack of visibility of the total cost of supporting business analytics for the enterprise.
- Investments in BA are segmented and essentially may be invisible to one another.
- Technology and tools are not standardized.
- Development is siloed.
- The effectiveness of disaster recovery, auditability, compliance reporting, and administration may be reduced.
- Best practices and knowledge gained may be shared inadequately.

Fortunately, there is an answer to these serious problems, just ahead.

[9] Personally, I have witnessed two executives arguing at length over exactly what the *real* truth was regarding a particular customer situation. The difficulty arose as the result of the two extracts taken from the same data source at different points in time. But even then, the data was stale.

zEnterprise Meets the BA Challenge

IBM's mainframe *System z*, as embodied in the *zEnterprise EC12 (zEC12)*[10], announced in August this year, represents almost 50 years of continuous research and development in technology, integration, and scale. It now includes a 5.5 GHz processor, fastest in the industry, and incorporates hardware accelerators that enhance the execution of many modern workloads, especially those with lots of I/O. Architectural evolution continues to be driven by goals based on built-in five 9s availability or better and world-class virtualization capabilities. Resources (processors, memory at all levels, and interconnections) are delivered as needed and automatically based on business-oriented goals and policies. New virtual images may be created, or shut down, on-the-fly without affecting ongoing workloads. Multiple virtual images concurrently can share all physical processor and I/O resources ensuring balanced usage and thus avoiding over-commitment.

In addition, System z has been hybrid in nature for over a dozen years, having the ability to simultaneously host both System z's native operating systems and the open *Linux* environments. More recently with the announcement of the *zEnterprise 196 (z196)*[11] in July 2010, and the *zEnterprise 114 (z114)*[12] a year later, System z has extended its reach to encompass the centralized management and control of *IBM Power Systems* blades running applications in the *AIX* environment, and *IBM System x* blades running applications in Linux or *Windows* environments. These architectural enhancements greatly expand the breadth of solution possibilities, all of which will share in the highest qualities of service to which enterprises have become accustomed by their reliance on the mainframe as the core of their IT infrastructure.

I Need the Answer, Yesterday

Time machines are fiction, but the need for timely and accurate information is real. Analytic solutions greatly benefit from the ability to co-locate transactional data, the data warehouse, and the analytics engines on the same platform.[13] Answer the question, *why move 70% of the enterprise data stored typically in transaction-driven systems to another platform that is likely to be less reliable and less flexible?*

The capability to co-locate your computing and your data is facilitated on zEnterprise systems by multiple layers of virtualization (hardware and software-based), a hybrid architecture that accommodates the unique requirements of several computing models, and internal connectivity via *HiperSockets* between tasks at the speed of a memory-to-memory transfer, instead of at external LAN speeds,. The result is reduced data movement, much less redundancy, and negligible latency between the first appearance of data and its availability for analysis. Enormous amounts of computing and technical resources are saved and network bandwidth is utilized more effectively, and all of this at a possibly lower cost than continually offloading and storing redundant data.

As for database management software on the mainframe, *DB2 for z/OS* continues its leadership in function, flexibility, and performance. Current releases feature enhancements in highly parallel and in-memory processing, row and columnar store technologies, and compressed data operations. Users have experienced DB2 query performance improvement up to three orders of magnitude, when properly tuned, according to IBM. Security of the data is assured with encryption co-processors built into every System z processor chip and for I/O by means of the *Crypto Express4S coprocessor PCIe* I/O card and supported by the *IBM Security zSecure Suite*. All of this delivers the highest standards for security, Common Criteria (EAL 5+).[14] With the zSecure Suite, access to data can be defined down to the cell level. Built-in security and trace features make end-to-end auditing possible.

Getting the job done quickly involves a number of factors. With zEnterprise, a scale-up or scale-out or a combination of both is possible. The aforementioned 5.5 GHz general-purpose zArchitecture processor is used for the CP (for z's

[10] For more detail on the zEC12, see **The Clipper Group Navigator** entitled *The IBM zEnterprise EC12 - Bigger, Better, Faster*, dated August 28, 2012, and available at http://www.clipper.com/research/TCG2012019.pdf.

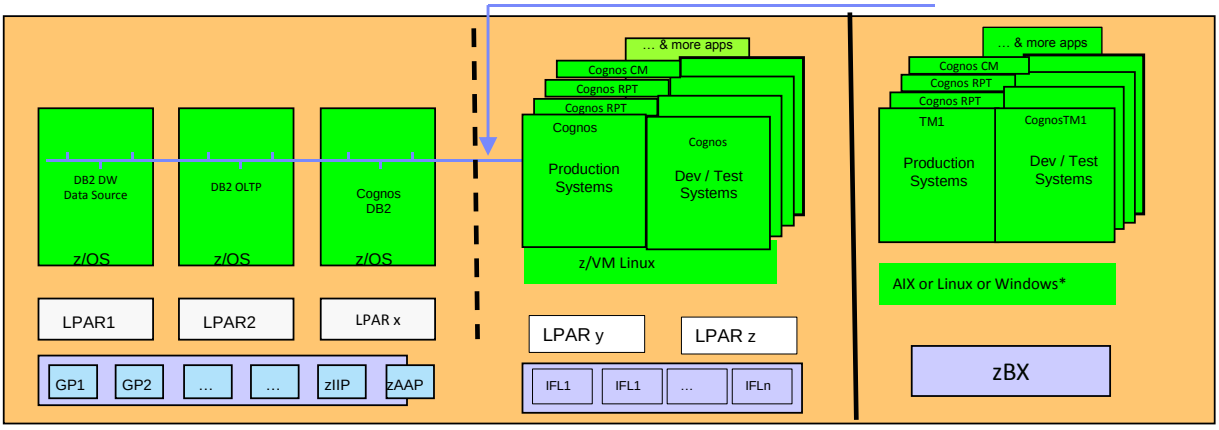
[11] For more detail on the z196, see **The Clipper Group Navigator** entitled *The IBM zEnterprise System Reaches Out - Higher, Wider and Deeper*, dated July 22, 2010, and available at http://www.clipper.com/research/TCG2010033.pdf.

[12] For more detail on the z114, see **The Clipper Group Navigator** entitled *IBM zEnterprise in the Midmarket - Revolution or Evolution?*, dated July 12, 2011, and available at http://www.clipper.com/research/TCG2011024.pdf.

[13] For more on why distance (co-location) matters, see the November 8, 2012, issue of Clipper Notes entitled *Bringing "the Power" Closer to Home - Implications for Future Storage Architectures*, available at http://www.clipper.com/research/TCG2012026.pdf.

[14] Common Criteria for Information Technology Security Evaluation (abbreviated as Common Criteria or CC) is an international standard (ISO/IEC 15408) for computer security certification; EAL5+ is the highest level.

Exhibit 1 — IBM zEnterprise and zEnterprise BladeCenter Extension — An Advanced Hybrid Architecture



Source: IBM

native operating systems), but also is used in all of the special purpose processors.[15] In particular, the *zIntegrated Information Processor (zIIP)* is exploited by z/OS to facilitate its workloads through offloading, such as XML System Services, IPSec offload, certain parts of DB2 DRDA, star schema, and HiperSockets for large messages. In addition, within the last three generations, from the *z10 EC* to the zEC12, the on-chip cache has been increased 800%, total active cores have been increased 150%, and the total processing capacity has increased 260%.

But what good are these resources if they cannot be managed and utilized to their maximum potential? Advanced, automated workload management can be used to prioritize analytics, as well as all other processes, and user requests by dynamically allocating resources (virtual or real processors and memory) to meet pre-defined, policy-driven priorities. Service level goals are achieved more easily and peak loads and surges are accommodated by the ability to drive processors to 100% capacity. When additional performance and growth are needed, additional processors, memory, and storage devices may be added and brought online very quickly (in minutes) without disrupting ongoing operations. This capability is complemented by industry-leading and long-standing virtualization and workload isolation capabilities that can adapt very rapidly and flexibly to accommodate additional users and data, as needed.

The announcement and availability of the zEnterprise EC12 was highlighted by the introduction of a unique and innovative technology, *IBM zAware*. This new feature is a self-learning, integrated expert solution that analyzes messages in near real-time to provide insight into the behavior of the z/OS systems that it monitors. It incorporates sophisticated analytics and applies expert knowledge and machine learning to alert operations managers of any anomalous conditions that may lead to an interruption of services thus reinforcing System z's reputation for its exceptional availability.

The Flexible and Cost-Effective Hybrid Solution

For over a dozen years, the mainframe has demonstrated its ability to perform brilliantly as a hybrid-architecture platform. Properly configured with IFL engine, *Linux for System z*, and *z/VM*, zEnterprise systems are capable of managing thousands of virtual images while continuing to provide traditional mainframe qualities of service. The introduction of zEnterprise systems, the z196 and the z114 a year later, in combination with the *zEnterprise BladeCenter Extension (zBX)* and the *Unified Resource Manager (zManager)* extended System z's architectural breadth to include POWER7 and x86 architectures. These actions opened up a vast new array of new cost-effective and exceptionally efficient infrastructure solutions, business analytics among them.

The System z infrastructure diagram (in Exhibit 1, on the previous page) illustrates the diversity of business analytics workload potential.

- ***Sharing general-purpose and special-purpose engine accelerators*** are a z/OS DB2 database

[15] They are ICF, IFL, SAP, zAAP, and zIIP. For more detail, see *Exhibit 6* in **The Clipper Group Navigator** entitled *The IBM zEnterprise EC12 — Bigger, Better, Faster*, dated August 28, 2012, and available at http://www.clipper.com/research/TCG2012019.pdf.

Exhibit 2 — IBM DB2 Analytics Accelerator V3 with Netezza Technology

Source: IBM

serving a CICS transaction-based application, a DB2 data warehouse, and a Cognos analytics workload; each is isolated on several LPARs whose resources, processors and memory, are shared and dynamically and flexibly managed.

- ***Sharing one or more IFLs*** are multiple Linux for System z virtual images managed and resourced by control through the z/VM hypervisor. Business analytics production, development and test, and any other application certified for Linux on System z (there are thousands) may be hosted here. Real memory may be shared with the z/OS applications and data interchange is accomplished speedily through the HiperSockets facility.
- ***On the zBX frame(s)*** equipped with IBM POWER7 (running AIX) and/or IBM System x blades (running Linux or Windows Server), business analytics applications and development and test may be hosted in a fully virtualized environment managed by the PowerVM and the KVM-based hypervisor, respectively.

What these workloads have in common is that they share resources and are centrally managed by the zManager on the host processor and are interconnected by internal, private, secure, and redundant management and high-speed data links. As a result, a three-tiered BA solution (or any other business solution for that matter) may defined under essentially one set of covers beginning with a user-facing Web portal, through the application server, and then to the data server (and back), all centrally monitored and managed from a common interface. Hardware resources utilization is maximized, unnecessary redundancy as well as single points of failure are eliminated, and the management and technical staff are less involved on an operational basis.

The IBM DB2 Analytics Accelerator – Turns Hours into Seconds and Dollars into Pennies

In addition to the very broad zEnterprise hybrid structure, IBM offers an opportunity for enterprises to achieve even higher performance and scale through the inclusion of the *IBM DB2 Analytics Accelerator for z/OS Version 3*[16], also known as *the Accelerator*, as shown in Exhibit 2, above. Powered by *Netezza* technology, the Accelerator V3 is a workload-optimized appliance specifically designed to assist enterprises meet the challenge of complex analytical needs. It can lower the turnaround time and predictability of complex queries dramatically; 1000 or more times improvements having been observed in real world applications (but, of course, your results will vary).

The Accelerator is integrated fully with zEnterprise processors (zEC12, z196, and z114)[17] as a logical extension of DB2 10, query tuning is eliminated as well as the need to create and maintain indices. The Accelerator is completely transparent to the user and, therefore no query modifications are required to begin to exploit its performance. Connection from the Accelerator to a zEC12, z196, or a z114 is via a pair of private and secure 10GbE links. Installation and loading can

[16] The IBM DB2 Analytics Accelerator V2 remains fully supported in the market with no price differential to V3.

[17] And, presumably, the next midmarket processor follow-on to the zEC12.

be accomplished quickly and can be production-ready in a few days.

A key new feature is the *High Performance Storage Saver (HPSS)*, which enables users to store historical data[18] only on the Accelerator (i.e., it is not a duplicate of any other data, except for backup, of course), thus lowering the cost of large data warehouses by shifting the storage costs from higher-priced, channel-connected storage technologies to lower-cost storage within the Accelerator.

One ever-present issue to tackle in the management of data warehouses is maintaining its currency to reduce latency. The Accelerator addresses this with a software enhancement known as *Incremental Update.* The Accelerator monitors DB2's logs, to be alerted when its copy of the data has been changed. This change then is duplicated asynchronously so that there is virtually no latency (i.e., very little time) between the new or changed operational data on the mainframe and the data in the data warehouse. As a result, the quality of the data at hand has been improved, the inherent value of the very current data has increased, and a significant amount of processor, storage, and administrative and management resources have been saved.

Another issue related to the maintaining a data warehouse is the labor and other resources devoted to the *extraction-load-transfer (ETL)* functions. The latest version of the Accelerator now includes a feature called *Unload Lite*. This feature essentially transfers the upload workload from the processor's z/OS-DB2 engines to the Accelerator, in a sense operating like a zAAP or zIIP processor. As a result, mainframe MIPS are saved – repeatedly.

Query prioritization also has been addressed. In earlier versions, the z/OS Workload Manager established the priority of the query workload relative to the total current z/OS workload. However, once this query had been passed to the Accelerator, its priority essentially became first-in, first-out (FIFO). With this enhancement, each query's priority is assigned by the z/OS WLM and is passed over to the Accelerator with each request and executed accordingly.

The *Netezza 1000* hardware technology (inside the Accelerator) consists of a blade-based streaming architecture that incorporates custom System x blades, memory, and storage combined with patented data filtering using Netezza's Field Programmable Gate Arrays (FPGAs).[19] Data scales to 1.28 petabytes (in ten connected cabinets) and is partitioned across multiple disks in RAID arrays, thus enabling a massively parallel processing (MPP) execution architecture. IBM has made remarkable progress in bringing this technology forward, by accelerating its ambitious development agenda. This is evidence of the potential for this technology – in concert with System z – to address broader and more demanding needs of emerging BA applications.

The Software that Drives the Solution Home

zEnterprise's business analytics solution structure also is supported by the *InfoSphere Information Server*. This integrated software suite provides the essential functions of information integration and governance (including ensuring data quality, providing a single view of Master Data, and performing lifecycle management, privacy and security functions) that are essential to a complete BA solution. The suite is supported by Linux for System z, AIX on Power System servers, and Linux on x86-64. It offers the following services.

- ***Extracting, transforming, and loading data***, batch or real-time; delivering and replicating changed data; and virtualizing access to disparate information (i.e., it enables federated databases).
- ***Discovering data relationships*** and creating and managing and common enterprise vocabulary across business units and IT.
- ***Analyzing and validation data***; establishing data rules; standardizing, matching, and de-duplicating data; and managing exceptions.
- ***Integrating "Big Data" volumes, velocity, and variety*** through support of *Apache Hadoop Distributed File System (HDFS)* and the ability to generate *MapReduce* jobs.

The InfoSphere Information Server suite is complimented by a set of world-class BA user tool sets. *IBM Cognos Business Intelligence* is an open, enterprise-class platform that provides comprehensive facilities for analysis and planning, reporting, dashboards, and scorecards with a broad reach and open access to a variety of structured and unstructured data sources. Cognos BI

[18] Within multi-year data warehouses, historical data can often account for up to 95% of the volume.

[19] For PureSystems fans, if this seems familiar it is because the hardware technology is shared with the *IBM PureData System for Analytics N1001*, announced in October, which is a standalone system for business analytic applications. The Accelerator V3 is a much different "personality", in that it is a fully integrated extension of DB2 for z/OS and thus shares its mainframe attributes.

is enabled for z/OS, AIX, Linux, and Windows platforms. Key features include:

- ***Dynamic Query*** – Interactive reporting and ad-hoc analysis, maximizes use of in-memory cache, dynamically selects SQL or MDX (Multidimensional Expressions) for best results.
- ***Dynamic Cubes*** – High-performance analyses through a combination of caching, optimized aggregates, and optimized SQL.
- ***Dynamic Query Analyzer*** – A graphical tool illustrating how a query was generated, thus assisting administrators in tuning, maintenance, and troubleshooting.
- ***Framework Manager*** – Metadata modeling tool that combines data from multiple sources into a unified view; enables merging of historical and current data.
- ***Data Manager*** – An ETL tool to build multi-dimensional data warehouses and data marts; it moves data into an optimal state for analytics.

IBM SPSS is another essential analytical tool. Its product portfolio consists of four families: statistics, modeling, data collection, and deployment. For example, the statistics family of tools is the most widely used suite of statistical software in the world and includes facilities for linear and non-linear models, simulations, customizing tables, data preparation, and validity checking. The modeling family of tools includes accessing operational data from a variety of sources, such as Cognos BI, DB2, IBM Netezza, Microsoft SQL Server, and Oracle MySQL. Modeling algorithms include CHAID, K-Means, neural networks, linear regression, and time-series forecasts, plus facilities for textual analysis, scoring, and to export in numerous file formats and data types. Scoring functions have been ported from SPSS Modeler 15 to DB2 for z/OS, thereby now being able to leverage the zEC12's very-high speed. Scoring may be invoked by SQL statements and run in user-defined functions (UDFs). Rather than extracting data from DB2 and sending it to SPSS on another platform, the data can be scored against the primary data from within the DB2 transaction stream, resulting in faster and more accurate results for subsequent analytics.

Conclusion

Many IT server, storage, and software tools vendors are competing for mind-share and installations in the business analytics arena. The broadest and most comprehensive set of solutions belongs to IBM, which includes offerings within each of its server families, System z, Power Systems, System x, and PureSystems, and combinations thereof, with complementary storage technologies to support them. Moreover, IBM offers several suites of software products including:

- **Information management** (InfoSphere Warehouse),
- **Integrated development** (Rational), and
- **Integrated management** (Tivoli).

The great advantage is that these suites are common to and support all of IBM's server environments. IBM also has a range of robust *SmartCloud* business analytics solution offerings, both private and public, that also address a wide range of business needs.

Business analytics applications certainly are not a new phenomenon, but new focus now is being driven by the rapidly emerging application space commonly referred to as "Big Data". I don't believe that IBM will attempt to attack this space head on with System z per se, because most solutions now being developed are focused on massively parallel, distributed commodity processors and storage media. However, I would not discount the possibility of a hybrid solution sometime in the future that includes zEnterprise as a control and centralization point.

Nonetheless, for those clients who are currently invested in System z or expect to be as they grow their enterprises, the System z architecture and particularly the zEnterprise family of hybrid "sidekick" servers continues to be most capable of delivering an exceptional range of very flexible, high-performance BA solutions. These have been designed to satisfy the most demanding requirements effectively and economically while maintaining the qualities of service that IBM's mainframe customers demand. When you really need a vehicle that does many things, particularly business analytics, very well and offers few compromises, System z probably is your answer. It should be at the top of your selection short list.

About The Clipper Group, Inc.

The Clipper Group, Inc., now in its twentieth year, is an independent publishing and consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

***Stephen D. (Steve) Bartlett* is a Senior Contributing Analyst for The Clipper Group.** Mr. Bartlett's interests include enterprise solutions including servers, system software, middleware, their underlying technologies, and their optimal deployment for responsiveness to emerging business requirements. In 2010, he joined the Clipper team after over 42 years with the IBM Corporation as an account and program manager in large system sales, product development, strategy, marketing, market research, and finance. During that time, he received several awards for his contributions in innovative market research and contributions to the business. He has a B.S. from Rensselaer Polytechnic Institute, and an M.S. from Union College.

- ***Reach Steve Bartlett via e-mail at steve.bartlett@clipper.com or at 845-452-4111.***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager**, *Clipper* **Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.