Powering next-generation BI systems

Madan Sheina
Principal Analyst, Ovum

October 2009

Powering next-generation BI systems

Executive summary

In a nutshell

Many server functions, such as application and web serving, can be handled easily by "commodity" Windows platforms. But requirements for core business intelligence (BI) database serving are significantly different. This paper explains the requirements of next-generation BI and details the technological components and features that companies should look for in their choice of server hardware, arguing that it is best served by a scalable, high-performance and reliable server platform. It looks at how one supplier, IBM, is bringing together the smarts of its Cognos BI platform and processing brawn of its Power Systems to support customers adopting BI or expanding their BI ambitions across the enterprise. It examines the key value proposition that joint IBM Cognos-Power customers can gain from implementing BI software optimized to run on a scalable, high-performance and reliable platform.

The Ovum view

Organisations want to implement a scalable, high-performance server platform with as low total cost of ownership (TCO) as possible on which to run their most important mission-critical business applications. In short, server platforms need to deliver processing muscle and resources without breaking the IT budget.

At a time when IT departments are pressured to do more with less, many are looking to make their IT infrastructure leaner in terms of implementation, running and management costs. At the same time, corporate data volumes continue to grow exponentially and processing windows shrink. Reconciling the need for performance and cost has proven difficult. The old ways of throwing more expensive server hardware at a performance problem is no longer an option. It often raises overall TCO of IT systems as data center managers grapple with sprawling server farms and the attendant costs of staffing, floor space, energy consumption and systems management.

One such application that's now in the crossroads of these two conflicting pressures is business intelligence (BI). A wave of next-generation BI applications is emerging. These applications extend beyond the offline, batch-driven load-query-analyze architectures, which are ridden with latency and serve only a handful of elite analysts to provide near-realtime response and analysis for a much broader set of operational business users. In short, BI has become a mission-critical application. That raises the stakes considerably for performance, scale, reliability and TCO. To run this mixed BI workload across the enterprise, companies need to make sure it is implemented on an equally capable mission-critical platform.

Arguably, no physical component of an IT strategy is more important than the choice of the underlying server infrastructure. Yet surprisingly, this is one of the most overlooked aspects of BI implementations. In most cases the choice is dictated by whatever is already in place. Many server functions, such as application and web serving, can be handled comparatively easily by "commodity" Windows x86 platforms. However, the demands of BI processing are different and need to be supported by a server platform that can deliver the scalability, performance and reliability a mission-critical application demands, while providing a low enough TCO to open BI up to the entire enterprise.

One company stepping up to the challenges of next-generation BI implementations is IBM. The company is smartly pairing the scalable processing brawn of its Power Systems servers with the brains of its Cognos 8 enterprise BI system. IBM has already proven the tight integration of these two technologies in its Smart Analytics System. Cognos BI for Power Systems focuses on providing a fit-for-purpose system aimed at running enterprise BI applications. The combination promises high performance and low TCO through an optimized tuning of Power System servers. IBM thoroughly analyzed Power Systems and came up with a proven set of configurations to achieve optimal performance, scale and resilience when running Cognos 8 BI workloads. Effectively, IBM removes the laborious, complex graft that IT departments undergo to make sure their servers and BI software are working in best concert.

Meanwhile, the TCO part of the equation is addressed by Power Systems' inherent server consolidation and virtualization technologies, which, if implemented correctly, can dramatically slash infrastructure costs. The ability to virtualize processor, memory and I/O resources ensures efficient use of server capacity resource and, therefore, less energy costs (after all, running applications on servers systems that run at 100% power consumption but 20% usage is wasted money and energy) and simplified growth management without having to manage sprawling farms of Windows and Unix servers.

IBM might be doing customers a favor. But we see it differently. If a company is sourcing its BI software and server hardware from the same vendor, it should expect – or, rather, demand – that its business applications run much better than on a competitor's server platform.

Reshaping BI requirements

Emergence of next-generation BI

The demands on business responsiveness and operational speed and flexibility for enterprises competing in today's economically challenging environment makes BI a necessity, rather than a luxury. At a time when survival depends on visibility into operations and making the right decisions, it's clear why BI initiatives continue to top CIO agendas. Companies want to use BI to minimize their risk, avoid surprises and anticipate opportunities before their competitors.

However, BI requirements have evolved along with the business climate. The traditional batch-driven extract, transform and load (ETL) data warehousing model, which worked well for simple, predictable and non-time-sensitive queries and analyses, can no longer keep pace with a new wave of BI environments that mandate greater analytic speed and currency. In short, analyzing past performance is no longer a guarantee of success. The problem is that BI has often fallen short of the ideal, delivering insight into the past but not into up-to-the-moment performance or future prospects.

The next generation of BI requirements has arrived, and is being driven by four factors.

- **Mission-critical use** – BI has morphed from a niche, back-office function to an operationally focused application woven into the fabric of daily business processes and activities. Analyzing the right data at the right time and delivering it in the right presentation style is now a mission-critical element of supporting daily operational tasks around sales, marketing, customer management, manufacturing and other core business functions. Many of these operationalized BI applications need to run continuously – often 24×7.
- **Pervasive deployment** – BI is extending beyond the traditional handful of power users or analysts to end users at all levels of the enterprise, as well as customers, partners and suppliers. Advances in GUI designs, self-serve capabilities and dashboarding tools have allowed BI to permeate up and down the enterprise ranks. However, it's not only user scale that BI systems need to grapple with. They also have to deal with increasingly mixed and unpredictable processing workloads.
- **Data volume and complexity** – Corporate data structures are growing more complex, demands for cross-organizational access to information are becoming pervasive, and external users increasingly want to interface to core databases through customer and supplier self-service systems.

- **Realtime performance** – The application of BI to operational processes has raised the bar for analytic speed to insight. The more operationalized BI gets, the greater the need for realtime processing. Many BI users need detailed information about the state of the business today, not last month, so they can visualize, analyze and respond to problems immediately. The expectation of immediate data access and query response times demands high-performance data processing.

Simply put, BI is getting bigger, faster and more sophisticated, all at the same time. These requirements are reshaping how BI systems are built, run and managed. They also have serious architectural ramifications; not least is the choice of an operating platform on which they can run efficiently and cost-effectively.

The right server platform is key

If BI has become a business-critical application, it stands to reason that it needs to be built and run on a suitably mission-critical architecture. Having the right BI software is a key consideration. However, the choice of server platform is also critical. Organizations have to be confident that their servers can meet the high service levels demanded by processing-hungry applications such as BI, and easily adapt to escalating and changing requirements.

Investing in a server platform is not a trivial investment or decision. Of all of the strategic infrastructure decisions IT organizations make, the choice of processing architecture may be one of the most important, and one that can't be left to chance. It will be a key factor in defining the current, and planned, scope of an enterprise BI system, including its agility to scale, performance, cost model and vendor relationships. Companies need to ensure that their server infrastructure is open, modern and purpose-built to accelerate architectural IT business decisions and future-proof existing IT investments.

Factoring in cost of ownership

It's easy for organisations implementing BI to be narrowly focused on the linear performance and processing capacity of a server. It's true that a server platform should have the muscle to easily adjust to expanding workload characteristics. However, it should do so without adding overheads (i.e. more servers) and complexity (management of those servers). Like any other IT investment, another important dimension needs to be factored into a BI investment – namely, cost of ownership.

Economic concerns are at the forefront of IT decisions. Application and database workload demands might be exploding, but corporate budgets are tightening. IT departments are therefore being asked to do more with less. The risk for an organization is that its underlying hardware infrastructure might not be up to the task of running the latest applications.

Balance is the key – facilitating the growth and overall performance of the BI system, while controlling costs of the operation and acquisition of the BI environment. It requires performance to be reconciled with price to provide total cost of ownership (TCO) efficiencies. Additionally, a myriad of other metrics also need to be distilled into the TCO assessment, apart from the upfront-cost purchase costs of the BI software and server platform. These relate to generic operational factors in all data centers, such as real estate (i.e. smaller server footprints and floor space requirements), energy efficiency and cooling costs, and minimizing IT staffing resources. The overriding goal is to create a BI software environment that provides optimal response time for its users and scales well to allow a growing number of concurrent users – all at the lowest cost of deployment and ownership as possible.

Important technical attributes for BI servers

So what should organizations look for in their server platform to support next-generation BI requirements and their attendant demands in scale, performance and TCO? In some cases, that decision might already be made, since there is always a tendency to deploy BI applications on whatever server platform is in place. Such an approach might be in harmony with an organization's attempts to standardize their IT infrastructure. However, it will not necessarily be the most ideal choice for BI.

Organizations should carefully consider the following capabilities of their server platforms before embarking on a BI implementation.

Smooth scalability

For organizations experiencing 30–60% annual data growth and increasing data complexity, the ability of a server to "scale with simplicity" and in a non-disruptive manner is extremely important to maximize returns on BI investments. That is because BI growth rates follow Moore's Law – i.e. doubling in size every 18 months. Companies looking to grow their BI systems out of a niche departmental use should, therefore, guard against getting caught short by implementing phase 1 and forgetting about how the system will scale to phase 2.

BI scalability refers to the ability of an application to process increasing workloads. That depends on the server's ability to support linear (predictable) and proportional (constant). Typically, BI systems need to support a scale-up or scale-out architecture. Each presents its own challenges. In a scale-up approach, a single device is used for the BI server and, as more throughput is required, a larger server (16 to 32 to 64 CPUs) is necessary. Using a single server often reduces the I/O and network delay of the system, particularly if the data source for the BI server is housed on the same device. A scale-out approach adds commodity servers (typically rack-mounted servers). As more throughput is required, more servers are added, one at a time. Because multiple BI servers are involved,

technologies such as a load balancer, messaging system or enterprise service bus (ESB) must be used, potentially introducing a performance bottleneck or overhead.

Scalability is not an exclusive issue for large, enterprise-wide BI deployments. It's also an issue for mid-sized organizations, many of which run their BI systems on Windows x86 platforms. Arguably, these companies suffer the most from massive Windows server farm management, as they need to continually add new servers to support growth. Additionally, scalability is not always linear – more so in a BI environment, in which its ability to scale (and deliver consistent performance) across a range of unpredictable processing workloads is, perhaps, more important. BI workloads are notoriously hard to predict, and often have peak times that are cyclical or outgrow the system resource originally planned. Hence, they are dependent on gaining easy, flexible and cost-effective access to processor and memory resources through the efficient sharing of pooled resources and flexible "capacity-on-demand" options that activate dormant processors.

Stable processing performance

Query performance is used to describe the speed of an application's response time to a system or end-user request or query. Query response time is critical for most BI applications.

A BI system can return good performance at low loads but perform poorly at higher loads. Why is this so? The limiting factors relate partly to the design of the BI application software used, while others relate to the choice of server. The main factors are influenced by the latter concern, memory, CPU count and speed and I/O bandwidth. Other factors that impact performance include the type of workload, how the software components and middleware products are configured, how the operating system is tuned, and the server capacity. It's also true that the transactional data sources accessed to generate BI content have a large impact on the performance of BI systems. That is because their schemas are not structured or indexed for BI reporting and analysis purposes.

Efficient use of capacity and resources

The introduction of new BI applications and unanticipated growth usually add complexity to the server infrastructure, leading to inefficiencies. Adding a sprawl of servers to keep up with these demands adds costs in terms of additional licenses, floor or rack space, power and cooling, data storage and administrative resources.

One answer is physical server consolidation through virtualization – i.e. provisioning multiple instances of operating systems across a single server and managing them through a single console. The technology has already had a profound change on the way IT managers consider cost-cutting in their data centers. Many turn to virtualization to efficiently scale their computing environments and deliver the high-performance experience and power that BI users demand, controlling server sprawl in their data centers at the same.

Understanding server usage is necessary for efficient virtualization and consolidation of servers. It's one thing to provide processing capacity, but another thing to fully exploit it. A recent study by 1E showed that idle servers cost businesses nearly $24.7 billion per annum in licenses, maintenance and wasted energy. Efficient server capacity usage, therefore, relies on mechanisms that intelligently and efficiently allocate and consolidate workloads, or both, on the most efficient servers and monitor and control workload execution processes across partitions. That becomes more pronounced in BI environments, in which BI workloads are constantly changing and smooth performance of the system relies on shifting them to available processing resources. Hence, the quality of virtualization and workload management tools supported by the server can be a key differentiator for enabling efficient virtualized BI server environments.

System resilience and availability

Resilience and the ability to meet and exceed demanding service levels of critical transactional business applications are top priorities for many companies. There's now an increased expectation that the same critical resilience will be mirrored in BI and analytic processing environments so that they are able to run uninterrupted, 24×7.

Outages affecting business applications or web servers can be dealt with relatively easily by switching to alternate platforms. However, an unplanned hardware or software failure on a BI database server, which may cause data to be lost or corrupted, means dealing with issues of data replication, synchronization and recovery that are generally more complex than for other server functions. Even planned outages – for such functions as hardware, operating system, database and applications software upgrades and scheduled hardware and software maintenance – can have adverse business consequences to operations and decision-making.

The underlying server system must, therefore, ensure continuous availability with minimal down time. This is particularly important as BI applications grow in scope and sophistication (as BI becomes more operationally ingrained into business processes).

Green footprint

Energy consumption management has become an increasingly important cost factor for data center expansion. Corporations are forced to be concerned by environmental regulations and scrutiny, or by the potential cost savings gained by reducing the physical footprint of their data centers (i.e. less floor space) and improving the operating efficiencies of the servers that run in those data centers through lowering the energy costs of spinning disks and cooling processor units. These costs can mount up for large data centers that have hundreds or thousands of servers installed.

Resource-hungry applications such as BI often test the capacity of current facilities to supply and manage server power and cooling costs. They tend to create

unpredictable troughs and valleys of processing demands. Building intelligence into the server layer that automatically shuts down power usage during the valleys is important in reducing the carbon footprint of the system and saving data center costs.

Optimizing your BI environment

Tuning the server

Simply installing and running BI software using the default configuration settings of a server's operating system or relying on the application design does not always guarantee that the system will perform at optimal speeds. It's true that server resources can be added to accommodate increasing load and manage service-level agreements for application performance. However, simply adding more servers might not always improve BI performance if the system is not tuned correctly. However, it will raise the overall TCO of running the system.

That requires companies to fine-tune the multiple configuration settings of the server to determine the best performance scenario according to various BI workloads. The primary optimization approach focuses on the operating system kernel. This typically involves the tuning of thread handling, memory allocation and lock performance management because most critical BI processing is implemented in a multithreaded, distributed architecture.

Challenges of manual fine-tuning

Tuning a server for optimal BI performance in-house is a complex task, given the mind-boggling numbers and combinations (literally hundreds) of operating system kernel parameters, ranges and values (often derived from others) that may need to be changed or tweaked to deliver an optimal configuration.

The task often becomes a substantial academic exercise and increasingly complex for constantly shifting BI workload environments. On top of server-resource configuration, it also requires careful management of database objects and structures. In addition, performance and scalability load testing is difficult to simulate, though a number of BI industry-recognized usage profiles and test case uses are emerging, which blend multiple BI workloads of various sizes and complexities running simultaneously over extremely large numbers of concurrent simulated-user sessions to fully stress all application and environmental server components.

While manual tuning that optimizes processor, memory and I/O performance can be a complex and tedious task, it also assumes the availability of skills, expertise, experience and resources. These need to be dedicated, since constantly changing BI requirements make configuration tuning a continuous process of refinement.

Naturally, the cost and complexity of configuration can mount up, and are often beyond the technical (and financial) capabilities of companies. Many companies are

looking for information and guidance on how to select the appropriate hardware model and configuration to run it on. Most look to their IT suppliers to provide this optimization out of the box (or server). Organizations tend to expect their IT vendors to provide more than just technology.

Many rely on their application vendors to provide integrated solutions that fit their needs and infrastructure. At a software level, they expect their IT providers to supply applications that understand their business environment – either horizontally (in terms of business processes) or vertically (in terms of industry data models). At a hardware level, they expect their platforms to be pre-tuned (or at least easily tuned) to optimally run the applications they deploy. This expectation becomes a demand when the BI software and server platforms are provided by the same IT vendor.

Running Cognos 8 on Power Systems

Two platforms – one optimized system

Recognizing the need to balance performance and TCO, IBM has paired two technologies in its portfolio – IBM Cognos 8 BI software and IBM Power Systems platform – to offer an optimized, “fit-for-purpose” software-server platform for running enterprise BI applications.

The pairing provides a scalable, high-performance foundation for supporting large BI processing workloads, while lowering the TCO of the system.

IBM Cognos 8

IBM Cognos 8 consolidates a broad array of BI capabilities onto a single platform, including query, reporting, multidimensional analysis, dashboards, business event management and scorecarding. Because these applications are designed to scale up and across enterprise needs and serve large and diverse communities of BI end users, scale and performance are key. IBM Cognos 8 meets the requirement with a multi-tiered SOA architecture that ensures that, when a request is submitted to IBM Cognos 8 BI installation, the right processing is distributed to the right tier (web, application and data) optimally. The BI system also employs an efficient performance layer built on the principle of peer-to-peer request distribution – the same principle that powers some of the most scalable file-sharing networks in the world. Of course, realizing performance potential relies on having a server optimally tuned to provide the necessary level of enterprise scalability, performance and availability.

IBM Power Systems

IBM POWER – an acronym for “Performance Optimization With Enhanced RISC” – is both a RISC instruction set architecture developed by IBM and a series of microprocessors that consolidates IBM System p and System i (AS 400) onto one

server platform capable of running operating systems, including AIX, Linux on Power and IBM i (the rebranded i5/OS). All of these operating systems, as well as the applications on top of them, run simultaneously on a virtualized machine.

IBM's POWER architecture is no longer just about server consolidation. One of its primary benefits is that it allows server systems built on POWER processor technology to scale in terms of silicon footprint and high-performance computing throughput. Power Systems also incorporates features such as system and workload management, availability and recovery optimization, security and autonomic computing – many of which are derived from IBM's mainframe systems. That makes for a server platform well suited for enterprise BI.

With Cognos BI for Power Systems, IBM has not just thrown these two platforms together. The company also made sure they can leverage the strongest attributes of each other. IBM's research and development division conducted an analysis of how IBM Cognos 8 BI runs on the AIX-based Power Systems platforms. The significance is not necessarily in the secrecy of the sauce but, rather, in the stirring. IBM has dedicated considerable effort, time and resources to research, test and analyze a variety of specialized configurations and tuning variables against live BI workloads. It has documented its findings as set of best practices that companies can use to ensure efficient transaction throughput and response times for their BI systems.

Customer benefits

The net result is a solution that can accelerate the deployment of a data warehousing system or BI applications and can grow in a cost-effective manner while keeping TCO in check. Presented below are some of the top reasons and benefits for running IBM Cognos BI applications on Power Systems.

Faster time to analysis and predictable scalability

Power Systems was built from the ground up to perform and scale across demanding 64-bit, high-end BI requirements, against both predictable and unpredictable mixed-enterprise workloads. The underlying POWER processer technology includes a highly effective compiler/operating system-level performance acceleration, low levels of symmetric multiprocessing (SMP) overheads and system-level integration.

Leveraging Power Systems' performance optimization features, IBM has devised several configurations around AIX memory, thread and lock management settings that are proven to run IBM Cognos 8 BI applications faster. IBM claims a 60% improvement in throughput BI processing performance (i.e. response times) with some of the best optimized configurations.

Just as importantly, Power Systems can safeguard BI and query and analysis performance by automatically adjusting server capacity to meet changing BI workloads. For example, tests showed twofold improvements in throughput when used with IBM Cognos BI software, and benchmark scalability up to 10,000 users.

It's important to note that these scalability and performance benefits are not restricted to large-enterprise deployments. However, IBM Cognos-Power optimization benefits can also scale down for mid-sized organizations through IBM's low-entry Express 520 Power System server range. Arguably, mid-sized firms stand to gain the most from the pre-optimization because they often lack the in-house skills to configure the system. Moreover, they often experience server sprawl on their x86 server environment as they grow their systems outwards.

Improved server usage

Server consolidation saves money, and Power Systems has a compelling story to tell. The PowerVM virtualization architecture provides a range of flexible (hardware, software and workload) partitioning schemes that lets companies optimally configure their mixed AIX, Linux and IBM i processing workloads onto a single, virtualized server. PowerVM comes with a host of virtualization management tools to track usage of shared system resources (useful for implementing chargeback procedures).

The main benefit is a much lower TCO enabled by vastly improved server capacity usage. PowerVM's integrated partitioning technologies allow the system to dynamically reallocate resources to portions of the BI application as needed to better match changing business cycles or handle unexpected surges in demand. For example, during peak BI processing cycles and times that outgrow system resources, PowerVM provides quick access to dormant processor and memory resource capacity. Power Systems also comes with a standby "capacity-on-demand" scalability to support a "pay-as-you-grow" strategy. That allows for substantial savings in BI software priced on a per-processor basis.

Smarter virtualization and use of server resources also accrues other TCO benefits through reduced management complexity and smaller server footprints.

The simplicity of PowerVM's unified hardware platform translates to lower personnel costs for systems administrative overheads when consolidating systems. This is achieved by reducing the number different operating system instances managed across multiple servers.

Power Systems' virtualization and consolidation capabilities also offer a space-saving and green design that conserves and manages energy usage in the data center. The system comes with its own energy management tool that monitors and optimizes power consumption.

Reliable and available BI analysis

Power Systems is equipped with a wide range of embedded hardware and software-based reliability, availability and serviceability (RAS) features. Capabilities such as component redundancy, pervasive monitoring, diagnostic, and fault isolation and resolution facilities built into processors, main memory, cache and packaging modules all help to minimize the potential for unplanned hardware and software outages.

Other capabilities in Power Systems that guard against outages include: failure masking (i.e. standby processors may be automatically activated if one begins to malfunction, without interrupting operations); concurrent maintenance ("hot plugging") functions and dynamic sparing, which also reduce requirements for planned outages; and live partition/application mobility, which enables users to move partitions between systems with no application downtime.