

# DB2 EMBRACES LINUX

BY RAYME JERNIGAN

Which database provides the best value? Which works best for deploying applications on a Linux\* cluster? Which is most scalable? And is high availability really achievable on Linux? To get IBM's take on the continued growth of Linux databases, IBM's *Linux Executive Report* talked to Robert Picciano, IBM's vice president of Database Technology, and Gary Schneider, DB2\* director of Linux Business Development. Here's what they had to say.

**Q:** What's driving customers to move to Linux for their IT infrastructure?

**Gary Schneider:** There are a number of factors, although total cost of ownership (TCO) remains a key driver. One way to do this is to deploy your IT infrastructure on Linux using commodity-based servers.

Regardless of the platform, customers require scalability, performance and high availability. With the onslaught of data facing customers (Radio Frequency Identification Device, sensors, pervasive devices), databases are growing quickly. If customers don't think about how their database applications can expand from the beginning, they can be boxed in later. It's important to think about these issues up front.

**Q:** Are you seeing customers deploy mission-critical applications, like a database, on Linux?

**GS:** Yes. We've come through the early adopter stage where Linux was used primarily for file and print serving. As Linux matures, customers are putting more mission-critical applications on Linux. Database software is fundamental to just about every

major business application out there, whether home-grown or from commercial Independent Software Vendors (ISVs).

There's research that shows that customers see database deployments as the main reason they'll purchase their next Linux server.

**Q:** Which industries in particular seem to be deploying database on Linux?

**GS:** It's happening across the board, but we're seeing the strongest growth in financial markets, public sector and distribution. These sectors are typically very cost-conscious, so that isn't surprising.

**Q:** Does DB2 on Linux have any specific advantages over the other major database brands?

**GS:** There are several: First, DB2 has the lowest total cost of ownership in the industry, so it's a natural choice when considering Linux. Why would you make the move to a low-cost operating system (OS) on a commodity hardware platform, and then potentially overspend on middleware?



Gary Schneider, DB2 director of Linux Business Development, IBM

**IBM**  
sets the  
record  
straight  
on Linux  
databases



Bob Picciano, vice president of Database Technology, IBM

PHOTOS COURTESY OF DB2

To reduce costs further, DB2 has included advanced automatic capabilities that IBM's competitors don't have. This also contributes to IBM's TCO advantage, because it means you need fewer people to manage your database environment. It's all about reducing the cost of computing while delivering the enterprise capabilities customers need.

**GS:** And let's talk scalability. Data volumes are exploding, so scalability should be one of the primary considerations when looking for a database that can support an information infrastructure. Some database products just don't scale linearly. That means that as the database becomes larger, it will begin to bog down. Adding new servers will only provide so much help, so your cost per transaction begins to rise and manageability often becomes a problem. This can eat up the value that initially brought the customer to Linux.

DB2 is unique in that it delivers unprecedented scalability characteristics that have been designed into the product and perfected over many years. We've also worked hand-in-hand

with our Linux distribution partners to ensure that DB2 leverages the extended capabilities of the 2.6 Linux kernel. The result is that DB2 really shines in Linux environments and helps us deliver the promise of scalability for Linux today. In fact, DB2 offers scalability to 1,000 nodes. We know our competitors' products don't scale linearly and are only "certified" to a small number of nodes, like 12 or 16. We back up our claims visibly in the market place, and DB2 has leading performance benchmarks for TPC-C and TPC-H workloads as well as ISV applications.

In the end, all that matters is how our clients' applications perform in delivering critical business value. Head-to-head, DB2 offers better price/performance and lower TCO than Oracle\* in key industry standard benchmarks.

**Q:** What about availability? Isn't it more difficult to achieve high availability in a clustered environment?

**GS:** You can achieve the same availability on Linux clusters that you can on other platforms. In fact, the lower hardware

# IBM DB2 for Linux

## Infrastructure on demand— to satisfy tennis fans' demands

By Barry Mittelman and IBM Sponsorship Marketing

Watched on TV in more than 161 countries, the US Open is tennis' premier championship. And for parent organization the United States Tennis Association (USTA), capturing that interest and turning it into revenue was key. But to do that, they needed to set their Web site apart from the scores of competing sports-related Web sites. The USTA can take advantage of access to elements of information from match results and statistics, as well as text, audio and video features and player bios—but so can a lot of other sites. Where the USTA has the most significant advantage is in its exclusive access to the scores of US Open matches in progress. Any sports site can post game-by-game scores, but the ability to publish official results as they occur straight from the umpire's chair, point-by-point, gives the USTA a natural lead. All they needed was the IT infrastructure to make it happen.

To fulfill this goal, the US Open site needed to be dynamically responsive to highly unpredictable geographic spikes in traffic triggered by, among other things, the nationality of a match participant. The USTA wanted a solution that could provide on demand access when it was most needed. Economically, it was important for the USTA to rally the resources of a fully implemented large solution, but without the investment in the fixed infrastructure and skills that go along with it. Luckily, IBM Global Services had a Linux\*-enabled solution that met the challenge.

### An Infrastructure That Serves All

With IBM\* @server xSeries\* server running Linux and IBM @server pSeries\* server running AIX\*, plus software such as IBM DB2\* for Linux, IBM WebSphere\* Application Server and IBM Tivoli\* Intelligent Orchestrator, the tournament—and its Web coverage—went off without a hitch.

Data, statistics and player biographical information are first stored in IBM DB2 for Linux, then published to WebSphere Business Integration Event Broker, which, in turn, through publish-and-subscribe technology, "pushes" real-time score information to fans via formats tailored to their preferences. This means that the scores can be

costs that Linux introduces into your environment means you can do things that would be cost-prohibitive in a more expensive server environment.

**Robert Picciano:** That's exactly right. For example, you could choose to mirror your compute nodes one-to-one where there is a backup server for every active server in the cluster. That may sound like an expensive architecture, but we're sometimes talking about only a \$2,000 compute node. Compare this to larger UNIX\* servers that can cost hundreds of thousands of dollars.

So it's not more difficult, it's just different. The strategy you use to create a highly available architecture will vary from one operating system platform to the other.

**Q:** Oracle tells customers that Real Application Clusters (RAC) makes high availability really easy for Linux clusters. Is DB2 as easy as Oracle RAC?

**RP:** Aside from having scalability limitations, the architecture of RAC potentially introduces several single points of failures in the system—customers need to be aware of that. And for smaller configurations, customers may be paying too much for Oracle RAC. DB2 comes with system failover software from Tivoli\* for the 2-node configuration at no extra charge, for example.

**GS:** Also, consider that another aspect of availability is scheduled downtime for maintenance. DB2 has several exclusive utilities built into it to allow you to perform preventive maintenance while keeping the database online. And there are many more high-availability features coming in DB2 Stinger, the next update of the DB2 UDB and DB2 Connect products, which will be released later this year.

**Q:** What do you see as the primary application of DB2 on a Linux cluster?

**GS:** Business intelligence is a key area where people are using Linux clusters. Business intelligence is the practice of examining, or "mining," data you already have to run your business better or uncover new opportunities.

Mining data like this often requires new infrastructure, and clustered or partitioned databases are very well-suited for business intelligence workloads. They let you break database queries up into smaller pieces and execute them against subsets of the data for faster results.

Look at the Herzberg Institute of Astrophysics under the National Research Council in Canada. They're using DB2 on a Linux cluster of 10 IBM\* xSeries\* servers because of the price/performance advantages and scalability of DB2. They're analyzing astronomical data, so you can imagine the volumes.

Using DB2 on a Linux cluster, you start small and can scale out with new compute nodes as your data grows, so you don't have to buy all the hardware up front. DB2's scalability



updated automatically as play happens, without the need for users to manually refresh their browsers to receive updated information. Furthermore, because the infrastructure only serves information that changes (instead of refreshing an entire page), bandwidth requirements are reduced substantially, reducing infrastructure load and improving performance.

"For such large sporting events, IBM needs to deliver on the promise of a highly available and highly scalable Web site. The system for the US Open has exceeded all expectations," says John J. Kent, program manager of IBM's Worldwide Sponsorship Marketing.

## From Courtside Seat to "Virtual Seat"

The Web site isn't the only place the IBM Linux infrastructure is put to use. Here's how it works: During the game, each point, as well as other information such as whether the serve was an ace or a double fault, is recorded by a staff member at courtside using a handheld device or laptop. This real-time data is automatically routed to the IBM scoring system where it's stored in an IBM DB2 for Linux database. Once in the system, the data is automatically tabulated and published to broadcasters who use the real-time data to display the score on television or create graphics related to the match. Information is also fed to the US Open's large video scoreboard, the source of match information for fans at the tournament, and to [www.usopen.org](http://www.usopen.org), the source of match information for Internet fans. Viewed this way, it becomes obvious that the IBM Linux scoring system has a profound effect on every fan following the matches.

The incorporation of Linux servers into the USTA solution began in mid-2001, so that by the beginning of the US Open in late summer of that year, several of the solution's xSeries and pSeries servers were already running. While originally deployed for testing purposes, the servers performed as site volume reached a record-setting two million unique visitors. By the 2002 US Open, the IBM solution was fully implemented on xSeries running Linux and pSeries servers. IBM WebSphere Business Integration Event Broker was also deployed to power the delivery of real-time scores to the site's IBM On Demand Scoreboard application. Thrilling the USTA, in 2003 the site was able to maintain 100-percent availability despite more than 15.2 million visits, an increase in site visits of 18 percent from the previous year.

means that you don't run into diminishing returns when it comes time to expand.

**Q:** What about transactional applications?

**RP:** For online transaction processing applications, achieving the best performance on anyone's clustered database can be a bit more complicated. You may need to pay attention to how the data's partitioned in order to achieve optimal application scalability in a clustered configuration.

In DB2 Stinger, we'll be helping customers do this with the Partition Advisor, one of DB2's autonomic capabilities. The Partition Advisor allows customers to describe their workload, and then DB2 will make recommendations about how to partition the data on disk. This helps optimize performance of the application for DB2 ICE, or Integrated Cluster Environment. (We'll continue to enhance DB2's autonomic capabilities in this area over time.)

It's also a common misconception that you have to rewrite the entire application to achieve scalability. Some applications partition better than others, but the same considerations apply whether it's DB2 or Oracle RAC. For example, we recently achieved over 1.3 million transactions per hour with a major Wall Street customer using DB2 on a 5-node Linux cluster with IBM xSeries servers, and the application did not have to be rewritten to achieve this result.

**Q:** What about commercial ISV applications and DB2 on Linux? Are ISVs adopting Linux?

**GS:** Yes. We're seeing a significant number of ISVs, both large and small, porting their applications over to Linux. PeopleSoft\*, for example, has made a significant commitment to Linux, porting 170 applications to the platform this year, and SAP\* has been enabled on Linux for some time now. Overall, we have almost 2,000 ISVs today who are enabled on IBM middleware on Linux.


**Q:** Last year, you introduced DB2 Integrated Cluster Environment. How's that going?

**GS:** The IBM DB2 Integrated Cluster Environment is a blueprint for deploying DB2 Linux clusters. It's based on the IBM Linux Cluster 1350 offering, which provides a complete Linux cluster/hardware platform with all the pieces and parts fully integrated. It leverages DB2's ability to scale out in that environment.

We have major customers, like Herzberg, using DB2 ICE configurations, but that's just of the tip of the iceberg. Our Linux story goes well beyond this one offering.

**Q:** So, what else is new for DB2 on Linux?

**RP:** DB2 Stinger will have enhancements for faster transaction performance and better availability. For example, we'll have High Availability Disaster Recovery, or HADR—



a data replication feature that provides high-availability solutions for both partial and complete site failures. DB2 Stinger also adds additional autonomic features to increase the ease of managing the database, with tools like the Partition Advisor I mentioned earlier and the Design Advisor. It will have better integration with Rational XDE to provide better tools to data architects. DB2 Stinger has first-class integration into popular Integrated Development Environments (IDEs) to ensure that developer productivity is as high as that of the database administrators (DBAs) that rely on DB2.

And we've been working for months with the key Linux distributors to exploit the new 2.6 Linux kernel, so that DB2 is ready when those distributions include that kernel.

**Q:** Oracle seems to have shifted their business focus to Linux. Is IBM doing the same with DB2?

**GS:** I think most people know IBM has been an early supporter of Linux and continues to demonstrate leadership on this platform across all of our middleware products. DB2's been available on Linux for more than five years now.

As customers move to Linux, we're finding that they want to re-evaluate their middleware infrastructure decisions, even re-evaluate past vendor relationships. This provides us

the opportunity to work with them on a complete middleware strategy to help them become an on demand business.

**RP:** It's really important to understand that a seamless flow of information across the enterprise is a critical element of an on demand operating environment. Businesses are striving to increase their productivity and get more leverage and value out of their IT investments. In today's world this often means focusing horizontally across business processes and applications and even outside of the firewall. We live in a heterogenous world and we can't expect customers to rip-and-replace their data management infrastructure to achieve this type of integration. A great example of integration includes the work we did as a vital part of a successful solution for the USTA (see the "IBM DB2 for Linux" sidebar on page 48).

Linux is a great unifier because it runs across all of IBM's hardware platforms. DB2 was designed to exist in a heterogenous world. Linux is an opportunity IBM fully intends to capitalize on for the benefit of our clients and partners.

**Rayme Jernigan** is a technology analyst living in the Research Triangle Park area of North Carolina. He was the founder and first-elected president of the Triangle Java\* Users Group, and has published various articles, interviews and essays about Java and Linux-related technology.