IBM

Web Wombat search engine accents the power of DB2 in B2B applications.

Web Wombat's busy spiders crawl the Web for new sites to add to its already vast cache of 100 million Web sites, making it Australia's leading search engine.

In 1994, when the Internet was still in its infancy, one enterprising engineer and a visionary journalist from the land down under hooked up their home PCs in the garage to cobble together a Web server and a search engine. At that time there were no more than 100,000 Australian Web sites, and a 28Kbps modem connection provided all the bandwidth needed to search the Web. Today, that same search engine has indexed more than 100 million Web sites in its data warehouse, attracts more than 100,000 unique users every month and, on average, performs 30,000 searches a day.

"Suddenly we found that we needed more—an enterprise data management solution in place of our existing Paradox database, hardware with much more processing power and a good deal more bandwidth."

–Phillip Bertolus, Technical Director, Wombat Technology

Key Components

Software

- IBM DB2® Universal Database™ for AIX®
- Linux®
- Tivoli® Storage Manager

Servers

- IBM RS/6000®
- IBM Netfinity® 5500
- IBM AS/400®
- IBM Enterprise Storage Server™
- IBM Magstar® 3590 Tape Subsystem

"DB2 is especially suited for integrated B2B applications. With other solutions, you run into all sorts of integration problems because you can't incorporate data from disparate data sources."

— Phillip Bertolus

Web Wombat, one of Australia's largest search engines, may have begun small, but its growth trajectory has been positively explosive. "We grew so fast that we quickly began to tax our infrastructure," recalls Phillip Bertolus, co-founder and technical director of Wombat Technology, the company behind the search engine. "Suddenly we found that we needed more—an enterprise data management solution in place of our existing Paradox database, hardware with much more processing power and a good deal more bandwidth."

That was six years ago. The original database has since been replaced by a one-terabyte data warehouse built on IBM DB2 Universal Database for AIX, Version 5.2. One IBM RS/6000 S80 server and one RS/6000 Workstation 43P (260) server give it all the processing muscle it needs, and a 10Mbps fiber-optic link allows the search engine spiders to trawl the Web endlessly for more sites to index.

"IBM technologies have allowed us to build an immensely scalable and reliable e-business infrastructure," remarks Bertolus. "How we use it is limited only by our imagination and the market's needs. Today, we've adapted it to develop a new portfolio of integrated business-to-business (B2B) knowledge management applications—targeting a new, and lucrative, market that will help fuel our future growth."

DB2 makes small work of surging data volumes
During this evolution, DB2 Universal Database has been the cornerstone of Wombat Technology's IT infrastructure. "We began with DB2 Developer's Edition, Version 2.1, running on an IBM OS/2 Warp® Server," says Bertolus. "Since then, we've scaled it by an order of magnitude, but in all these years DB2 has never once crashed or lost data. We've moved it from IBM OS/2® to Microsoft® Windows NT® to Linux, IBM AIX and even Sun's Solaris™ Operating Environment—that kind of multiplatform scalability is extraordinary."

Notes Michael Tancredi, managing director, Wombat Technology, "The only reason we haven't produced a larger database is that commercially there isn't a need for one at present. From a technology perspective, there is virtually no limit to how big our database can be or how much of the Internet we can actually index in it. With DB2 and the RS/6000, our system is designed for near-linear scalability."

To efficiently and reliably manage data back-up and overflow in its high-volume data storage environment, Wombat has deployed the IBM Enterprise Storage Server in conjunction with Tivoli Storage Manager and the IBM Magstar 3590 Tape Subsystem.

The search engine, developed in C++, has a mass-distributed spider architecture that gives it great flexibility. Out on the Web, the spiders trawl the data highways, looking for new Web sites that fit predetermined criteria. These remote spiders compress the content from these Web sites and send it back to the core spider residing on an IBM Netfinity 5500 server. The data is then parsed, indexed and stored in the DB2 data warehouse. "We can create and retrieve very large indexes from DB2 in seconds," says Bertolus. The search engine can perform 60 searches per second on the Web Wombat proprietary indexes, using only six processors. With the potential to go up to 24 processors or beyond, performance will never be an issue for growing user volumes.

B2B solutions enable backend integration, improve business performance
With its search engine firmly entrenched in the business-to-consumer (B2C) marketplace, the company is employing the technology to develop innovative B2B applications that enable corporate knowledge management. Integrating directly with companies' online transaction processing (OLTP) and backend enterprise resource planning (ERP) systems, these applications support advanced business intelligence needs with data mining, catalog and database search capabilities. All new application development at Wombat is done on DB2 Universal Database for AIX, Version 7.1.

"From a technology perspective, there is virtually no limit to how big our database can be or how much of the Internet we can actually index in it. With DB2 and the RS/6000, our system is designed for near-linear scalability."

–Michael Tancredi, Managing Director, Wombat Technology

With DB2 scaling to store a growing volume of Web site information, Web Wombat can provide users a rich Web searching experience.

Offered both as licensed applications and ASP services, the B2B solutions are quickly gaining market traction. The Wombat Catalog Search, for example, integrates with the ERP systems of companies that sell online and retrieves product information directly from their backend databases, improving transaction performance and reducing throughput loads on these companies' backend servers.

An early adopter of this technology was Buy.com, which had 60,000 products in its catalog but needed the ability to go up to half a million. "Basic search operations were taking up to 90 percent of their CPU capacity and impacting the performance of high-value sales transactions. Customer service was suffering," says Bertolus. To address this, Web Wombat integrated its search engine with Buy.com's ERP system, enabling direct online access to the catalogs. "We reduced processor load during the peak holiday season from 90 to 5 percent," says Bertolus. Buy.com exports XML forms containing product information from its backend IBM AS/400 server and puts them on an FTP site from which the search engine retrieves and indexes them.

"DB2 is especially suited for integrated B2B applications," Bertolus adds. "With other solutions, you run into all sorts of integration problems because you can't incorporate data from disparate data sources." News Interactive, the online division of Melbourne-based News Corporation, used this capability in deploying

the search engine for its online job placement service. Thousands of job descriptions that employment agencies submit to News Interactive in XML format are extracted by the search engine, indexed and made available for quick, relevant searches by job applicants.

Spinning predictive solutions

Wombat Mining, the company's data mining solution, is used to uncover data generated by OLTP systems and index it to perform pattern searches. Thus, a mobile phone network can define search parameters to determine which users are most likely to change their service providers. More than 20,000 searches can be executed overnight, across multiple transactional databases, and the results can be fed back into the customer relationship management system, setting off proactive customer-care processes.

"We've evolved from an Australian search engine to a global search engine without experiencing any of the technological pains associated with such rapid expansion," notes Bertolus. "DB2 Universal Database and RS/6000 provide us with the high-performance, reliable and scalable e-business infrastructure we need to sustain future growth."

For more information
Please contact your
IBM marketing representative or
IBM Business Partner.

Visit us at:
ibm.com/e-business

For more information about
Wombat Technology, visit:
www.webwombat.com.au

IBM

G325-1871-00