IDC
Analyze the Future

INSIGHT

Azul Systems Targets Java-Enabled Workloads for Improved Throughput

Matthew Eastwood Jean S. Bozman

IDC OPINION

Azul Systems is a Silicon Valley company that aims to improve Java-enabled workloads by processing these applications on a specialized, highly optimized compute appliance. The company is focused on improving the performance of customers' Java-enabled workloads as follows:

- By taking Java-enabled workloads from a variety of networked server hosts and processing the compute load on behalf of these host systems. (The Azul Compute Appliance aims to be a Java application workhorse that alleviates application performance bottlenecks and accelerates business results from these mission-critical applications.)
- By allowing the originating servers to remain in place as their workloads are directed to Azul's multicore systems for faster processing, then returning the results to the originating servers — without disrupting current IT infrastructure
- By coupling this appliance-based approach with an understanding that customers are increasingly virtualizing their software environment — and optimizing Azul appliances to support this trend through their shared-resource design

IN THIS INSIGHT

This IDC Insight examines Azul Systems, a Silicon Valley firm that is focused on optimizing the performance of Java-enabled workloads. The company, now four years old, is working to improve business results from Java-based applications that are running across a multitier enterprise computing environment.

SITUATION OVERVIEW

Introduction

Azul Systems has a mission: It is working to change the way Java applications run across a multitiered heterogeneous network of servers. The company does that by redirecting those application workloads to an Azul Compute Appliance, which houses two to 16 powerful multicore chips in an array providing up to 384 coherent processor cores and 256GB of memory within each appliance. Additionally, the company is now shipping its second-generation appliances based on the new Vega 2 (48-core) processor.

Filing Information: February 2007, IDC #205368, Volume: 1, Tab: Vendors

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

The company founders believe that the phenomena of underutilized server resources, information silos, and nonoptimized application code all persist in the enterprise today, causing application performance to suffer. Azul Systems addresses this problem by redirecting Java workloads to its compute appliances, which have been optimized to efficiently process Java code. Each appliance provides compute resources for the application hosted on the target application server and can be viewed as a form of proxy system, which takes on Java workloads on behalf of these host application servers.

The business benefits of this approach are driving the solution, with the key business drivers including the following:

- Improved performance for the Java-enabled applications, resulting in improved throughput and response times for data associated with these applications
- Improved ability of IT to meet service-level agreements (SLAs) and quality of service (QoS) for the enterprise' business units
- A dynamic infrastructure that can adapt quickly to changing business requirements
- Accelerated time to market for IT services by reducing tuning, provisioning, and capacity planning efforts

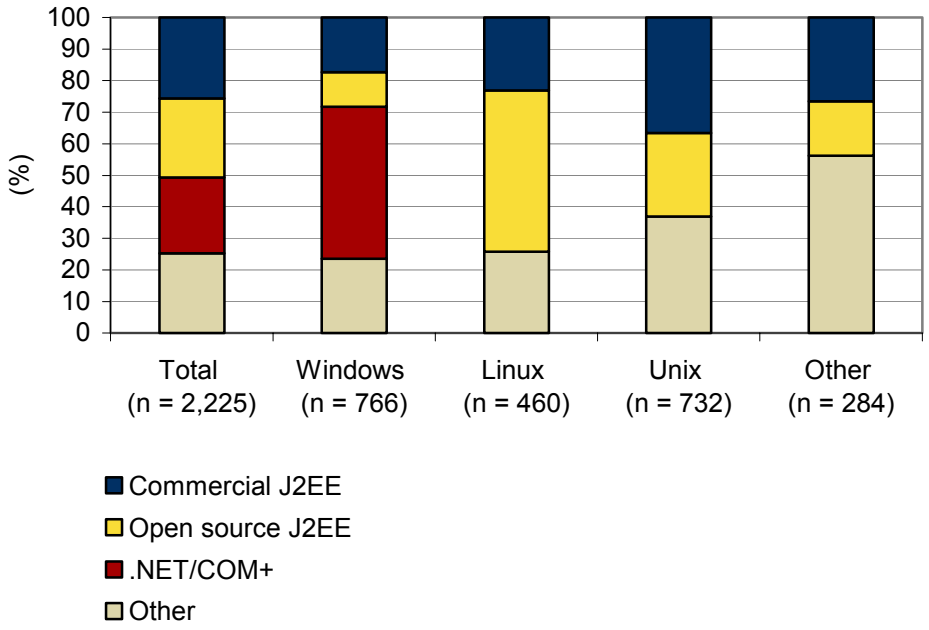
The Java infrastructure opportunity is large, as ongoing IDC server research shows. As Figure 1 illustrates, IDC estimates that more than $11 billion — or 20% of all server spend worldwide — was invested in server infrastructure in support of Java application in 2005. This data is based on IDC's demand-side 2005 server workloads research published as part of a series of studies that look at server workloads annually. Based on this research, it appears that the opportunity for Azul Systems is largest in the Unix server market — given the company's extensive use of enterprise Java workloads within the datacenter — but the Windows, Linux, and mainframe markets also present strong opportunities for the firm's products.

Azul Systems believes its solution addresses some continuing problems with Java workloads that are affected by scalability and response time consistency. The company is providing resources in its appliances that are designed to provide significant scalability and consistent response times for Java applications. These resources are intended to improve scalability, computing response times and, thus, SLAs. Examples of these resources include:

- Large coherent compute capacity, provided by multichip, multithreaded system design
- Large amounts of available memory (Single instances of a Java application can access up to 200GB of memory.)
- Hardware-assisted garbage collection that minimizes application pause times required for reclamation of metadata associated with object-oriented Java processing (aka garbage collection)

FIGURE 1

Java Server Customer Revenue Share by Operating Environment and Deployment Platform, 2005



Notes:

$11.2 billion was spent on more than 1.3 million Java servers worldwide in 2005.

Commercial Java does best on Solaris and HP-UX.

Commercial J2EE does best in the financial, manufacturing, and communications segments.

Open source J2EE Linux penetration is twice that of the overall market.

Open source J2EE holds steady with company size.

Communications and healthcare favor open source J2EE.

Source: IDC, 2006

Company Description

Started in 2002 with private funding and several rounds of venture capital investment, Azul Systems is headquartered in Mountain View, California, and maintains offices in Tokyo and London. These cities, not coincidentally, maintain the greatest concentration of financial services companies on the planet. The company also has a presence in Bangalore, India, where development and testing are supported.

Azul Systems was founded by Shyam Pillalamarri, Scott Sellers, and Gil Tene — and is led by chief executive officer (CEO) and president Stephen DeWitt, who was CEO of Cobalt Systems until that company was acquired by Sun Microsystems in CY00. Pillalamarri serves as vice president of software engineering, Sellers as chief operating officer (COO), and Tene as vice president of technology and chief technology officer (CTO). Other executives have corporate experience at Sun, HP, Sequent Systems, Checkpoint, and other companies with histories in the SMP server, appliance server, networking, and software infrastructure markets.

To address the business opportunity it targeted, Azul Systems has designed its own multicore processors, which are manufactured by TSMC in Taiwan per the company's design specifications. Additionally, Azul Systems recently launched its second-generation processor technology, called Vega 2 (with 48 cores), which will allow the company to run more workloads with the support of additional threads per core — and more cores per Vega processor than the Vega 1 processor, which had 24 cores.

At the same time, Azul Systems is building up an ecosystem of software partners that want to see their Java-enabled applications optimized to run faster than with traditional general-purpose servers alone. To meet this goal, Azul Systems has relationships with more than 20 technology partner companies, including three large app server providers — Oracle, BEA Systems, and Red Hat/JBoss — which provide licensing agreements to Azul customers that co-deploy Azul appliance technology along with the app server software. Additionally, the company added support for Microsoft .NET–enabled workloads via an alliance with Mainsoft in 2006. Azul is also building a channel for its technology designed to drive market penetration across multiple geographies worldwide. Reseller partners include ITOCHU Techno-Science Corp., Nissho Electronics, and Sumisho Electronics in Japan. Azul technology is also distributed through Morse Ltd. in the United Kingdom and World IT Systems in Poland.

Technology Strategy

Azul Systems identified the explosive growth of Java workloads in the enterprise at the time of its founding and set out to build a system designed to optimize the execution of Java code. It does so by shifting Java workloads from a variety of host servers — all running Java Virtual Machines (JVMs) and some form of app server software (e.g., IBM WebSphere, BEA WebLogic, or Oracle Application Server) — to Azul Compute Appliances.

The Azul appliances are based on a custom-designed multicore processor that is intended to efficiently run multithreaded workloads. There are two to 16 Vega chips per Azul Compute Appliance, providing support for up to 384 active threads per clock cycle based on the use of the 24-core Vega processors. When the new Vega 2 processors are used, the number of simultaneous executable threads is effectively doubled.

This shift is transparent to the application itself and is accomplished by replacing the JVM on the host server with the Azul Virtual Machine. This Java-compatible VM moves all Java executables off the host server and over a Gigabit Ethernet network to the compute appliance, where they are instantiated. To all elements of the infrastructure that interact with the Java code, the application appears to be resident on the host application server the entire time.

Multiple applications can tap into the same shared resources within the Azul Compute Appliance — and the appliance can manage a "pool" of compute resources via its Compute Pool Manager software. In effect, the appliance serves as a large shared compute and memory pool for the targeted workloads.

Underneath the application server software layer, a range of server operating systems are supported on the customer's installed servers — which remain in place as Java-enabled workloads are redirected to the Azul Compute Appliances. On first shipment, the operating systems supported by the Azul Compute Appliances included Sun Solaris 8 and 9 for Sun UltraSPARC-based servers, Red Hat Enterprise Linux (RHEL 2 and RHEL 3), and Novell SUSE Linux Enterprise Server (SLES 9). Since then, Azul Systems has added support for IBM AIX 5L Unix and HP-UX 11i Unix. Additionally, support for Microsoft Windows servers is expected to be made available in 2007. By addressing many of the top operating systems in the enterprise computing world, Azul Systems is acknowledging the inherent heterogeneity of host application server systems found in most enterprises — and underscoring the promise of Java's "write once, run anywhere" technology, which was launched by Sun Microsystems in May 1995.

The Systems

The company launched a family of three compute appliance products in April 2005. The Azul Compute Appliance 960, housing four Vega 1 chips (96 processor cores in total) and 32–64GB of memory; the 1920, with eight Vega 1 chips (192 processor cores) and 64–128GB of memory; and the 3840, with 16 Vega 1 chips (384 processor cores) and 128–256GB of memory. All appliances provide up to 4x Gigabit Ethernet ports, two 10/100 Ethernet management ports, redundant network processors, and N+1 hot-plug power and fan trays. The 960 model comes in a 5U enclosure, while the 1920 and 3840 models are delivered within an 11U, 24in.-wide enclosure designed for a standard datacenter rack environment.

With the launch of the Vega 2–based compute appliances, Azul Systems has added two additional models in the same 5U form factor as the 960. The 3210 model has two Vega 2 chips (96 processor cores) and up to 48GB of memory, and the 3220 model has four Vega 2 chips (192 processor cores) and up to 192GB of memory. Pricing for Azul Compute Appliances starts at $49,995 for the 3210 model.

Business Objective

Azul Systems' stated business objective is to become the preferred platform for business-critical Java workloads. The company wanted to accomplish this goal by shifting Java workloads transparently, via the network, to its appliances. The aim is to combat the phenomenon of server sprawl within the application tier that has developed in recent years. Azul Systems contends that datacenter capacity and cost can be dramatically reduced through the use of its appliances because they offer server density along with improved power efficiency and application performance compared with a collection of host application–tier servers delivering the same overall throughput.

IT users are voicing frustrations within the IT infrastructure, particularly within the critical application-serving tier. Azul is providing both compute and memory capacity in the form of a shared resource pool capable of running these Java applications efficiently. In the most demanding environments, such as those found in financial services markets, in telecommunications, and in ecommerce systems, absolute performance is key to business success. A lack of performance can reduce an

organization's ability to serve end-user requests (such as for trades or stock quotes), reducing QoS and negatively impacting SLAs made with business units within an enterprise. Further, the lack of scalability found in most x86 servers today can sometimes cause Java workloads to run out of "headroom" and resources — often requiring a move of that workload to a larger, more expensive SMP server for effective processing, which may be determined by the strict SLA requirements in many enterprises.

Following an early focus on financial services companies, Azul Systems is now branching out to a variety of IT organizations, which vary by their size, industry, IT maturity, and the number and types of Java-based applications being supported. This broadening of focus is to be expected as the company matures and moves to widen its installed base beyond early adopters. That kind of demand for mission-critical workloads extends beyond financial services to include other vertical markets requiring timely results that closely affect the bottom line of the overall enterprise's business results.

Technology Profile — How It Works

Java applications have historically provided wide-ranging interoperability between hardware platforms but have faced performance problems related to the just-in-time (JIT) Java bytecode compilation that takes place when the code is executed on the JVM aboard the server. Java-enabled applications also rely on object-oriented technology that often requires garbage collection of unused objects to optimize performance. These intermediate results must be continuously removed to reduce latency that slows throughput.

Azul Systems claims that its technology provides roughly three times the performance and throughput for typical Java applications, according to company testing and measurements, and it enhances garbage collection for unused Java objects, which is an important aspect of this performance improvement. However, additional testing results from Azul Systems show that the throughput results can go much higher — exceeding 10 times the original throughput, or more. IDC notes that these test results are posted on the company's Web site at **www.azulsystems.com**.

The Azul Compute Appliance is installed at customer sites, taking on the Java-enabled application workloads from a variety of host application servers on the network and processing them on behalf of those originating systems. Its multithreaded compute technology, coupled with large memory resources, and its hardware-assisted approach to garbage collection of used Java objects are combined to improve data processing when compared with throughput on the combined collection of host servers where the workload originated. This means that current systems can stay in place — but they can get a workload optimization assist for their Java applications when placed in the same network as the Azul Compute Appliance.

Azul Compute Appliances were designed using a new type of processor that was custom designed by the company and fabricated in a specialty semiconductor fabrication facility in Taiwan by TSMC. Multiple applications can be supported on each compute appliance — while the computer systems that originally hosted the workloads remain in place at the customer site, thus minimizing disruption to the core

business needs of each Azul customer. When the processing on the Azul Compute Appliance is completed, computing results are seamlessly delivered back to these host systems.

Each appliance utilizes a symmetric, flat SMP design to aggregate all processing and memory in each system into a single logical set of resources that multiple Java applications can leverage. This flat design allows applications of all sizes to be hosted on the systems with no concerns about where an application must be placed to achieve maximum efficiency. Memory access is uniform throughout the system, so large applications have fully contiguous access to RAM anywhere in the system. The Vega processors used in the appliances do not run the x86 or SPARC instruction set. They are uniquely designed to execute highly multithreaded managed code, such as Java. Unique instruction sets are provided in each processor to assist with several routine tasks in these types of applications, such as garbage collection and thread management.

To tap into an Azul Compute Appliance, users install the Azul Virtual Machine on the originating servers, where the Java applications are installed. The IT site changes the JAVA_HOME command to leverage the Azul VM instead of the virtual machine being used on the originating app server. Users then restart their application on the Azul Systems machine.

Upon restart, the Azul VM redirects the Java application to the compute appliance, where it is instantiated. After redirection, the host server plays the role of proxy for the compute appliance, handling all communications between the Java application and the other elements in the infrastructure. Application-level clustering, monitoring, profiling, and all other programming continue, without modification, although the processing has been relocated to the Azul Compute Appliance. Importantly, the Java-enabled application appears to be resident on the traditional server.

Technical Benefits of the Azul Compute Appliance

- **Scalability for growing workloads.** Azul Compute Appliances allow single instances of an application to expand their footprint from the typical limit of 2–4GB and four to eight cores on originating servers and to access up to 200GB and hundreds of simultaneous processor cores to execute hundreds of simultaneous threads.

- **Provision of RAS characteristics.** Azul Compute Appliances exhibit high reliability, with good mean time between failure (MTBF) characteristics. There are fewer moving parts than in many other server designs because there no hard disks, switches, fabrics, or mechanicals within the appliance systems. In addition, there are redundant network connections and hot-swappable power supplies and fans, reducing failure from those types of components.

- **Large global shared memory.** Each appliance provides between 32 and 256GB of memory that can be shared among multiple Java applications. This large, contiguous memory set can be allocated in any way between applications, from 1GB heaps for small applications up to a single 200GB heap for a single large application.

- **Hardware-assisted garbage collection.** Memory allocations larger than 2GB are impractical on traditional servers due to the need to clean up and reallocate that memory during the run of an application. This process of garbage collection — involving cleanup and reallocation of memory — can stop an application for a period of a few hundred milliseconds up to minutes, based on how much garbage has to be collected. Azul Systems appliances employ hardware-assisted garbage collection, allowing for parallel, concurrent cleanup of used objects while the application is running, reducing garbage-collection pauses.

- **Compute pool manager.** Administration software called the Compute Pool Manager governs resource allocation within the system. This Web-based tool allows administrators to ensure resource allocation and assign priorities to each application sharing the appliance and memory and processor reservations.

- **Appliance operating system.** Each Azul Compute Appliance is powered by a Linux-based embedded operating system, which does not expose its binaries. This operating system, working in conjunction with the Azul Systems VM and Compute Pool Manager administration tool, controls the dynamic allocation of resources, the resource availability and reliability features of the system, and the communications between multiple Azul Compute Appliances.

Inclusion of Microsoft and Microsoft .NET Technologies

Although Azul Systems is focused on improving Java application performance, and has been since its launch in 2004, it can also execute Microsoft .NET workloads via an alliance with Mainsoft and its Visual MainWin product. This is an important alliance because most companies have substantial investments in Microsoft Windows and .NET technology, as well as in Java applications and J2EE technology. Both environments coexist, and both have been widely deployed as customers have been involved in the process of deploying Web services across the networked enterprise.

Here, there is a need to deal with both Java objects and .NET objects — as has been the case since Microsoft accelerated its .NET strategy in the late 1990s and early 2000s. Because objects "inherit" the characteristics of parent objects, the object types are very different from one another, even if the business goals of the Java and .NET environments are similar. IDC believes that Azul Systems' alliance with Mainsoft is a good step in the direction of improved interoperability between these two widely deployed app server and processing environments, particularly for end-to-end solutions that tap both Java and .NET functionality along the way.

A Changing Technology Environment

In the four years that Azul Systems has been growing as a company, a number of technologies have appeared in the marketplace, and these have affected the shape of customer deployments. Among these technologies are the following:

- **Multicore processors.** The shipping of dual-core and quad-core processors has made multicore capabilities widespread within the marketplace. Although the number of cores is far less than that of Azul Systems' Vega processors, it has already affected the way that customers deploy workloads via the increasing server density of rack-optimized and blade server form factors.

- **Virtualization.** This technology has long been leveraged within the IT datacenter but was typically within the realm of mainframes and Unix/RISC servers, Now, the x86 server platform, the largest server market segment, can exercise virtualization in a number of ways, as supported by the VMware, Xen, and Microsoft software–based hypervisors — the VM managers built into Intel and AMD processors (Intel VT and AMD Virtualization) and the Sun Solaris x86 support for logically partitioned "containers" within a Solaris instance.

- **Improved workload-balancing techniques.** The proliferation of volume servers (servers priced less than $25,000) in recent years has brought with it a number of approaches to group these small servers together via clustering software, grid software, or workload-balancing software — all of which have the effect of redirecting new requests to less heavily burdened servers for faster processing on available compute resources.

Taken together, these elements of multicore, virtualization, and workload balancing are setting the stage for IT transformation in many IT shops — so that more flexibility is being gained through their deployment. Although not adequate for the most pressing data processing demands, where milliseconds can be a competitive differentiator, they are forming a foundation for "good enough" computing, in which demanding tasks are directed to the most capable servers in the datacenter, and many more workloads are not optimized for application speed-ups due to less business sensitivity across a range of workloads to high-throughput performance.

Importantly, Azul Systems must make the case that it can work well with these new technologies that are already in place in customer environments — and it must embrace the reasons customers have been deploying virtualization and multicore processors and build upon them as part of the value proposition for Azul Systems deployment. Specifically, the emergence of VMs does not mean that Azul Systems cannot work with them; just the opposite. It can manage applications that are built on top of VM technology, facilitating the pooling of server resources and improving management of applications built on Java, .NET, and SAP NetWeaver technologies. Finally, some customers are choosing to optimize their app server software as a way to improve Java workload performance. Again, Azul Systems must show prospective customers how these approaches compare with its solution.

Customer Examples

Early adopters of Azul Compute Appliances included financial services companies, with several located on Wall Street. But customer use has moved into other types of businesses — spreading into other vertical markets outside financial services, including travel reservations centers with high online transaction rates, credit agencies, and online, Web-enabled sports and entertainment companies:

- Credit Suisse is using the Azul Compute Appliances to achieve substantial datacenter consolidation benefits and to improve application performance. The company saw an initial 45% reduction in TCO and a 75% reduction in server count for the Java-based services that are now leveraging the Azul Systems appliances.

- TrueCredit, a division of credit reporting leader TransUnion, leverages the Azul Compute Appliances to support its customer-facing credit reporting services. TrueCredit used the Azul Systems appliances to consolidate its server resources and has found that it is able to support at least 50% more customers than before and that it has improved asset productivity fourfold.

- British Telecom (BT) provides local-loop gateways to its worldwide network backbone for a variety of telecommunications companies. BT utilizes Azul Compute Appliances as the infrastructure supporting these gateways, providing SLA assurance and increased computing capacity.

Challenges and Opportunities

Like most new companies, Azul Systems faces the ongoing challenge of competing with some of the world's largest technology providers, many of which provide their own Java technology solutions, comprising hardware and software components. Further, large systems companies and systems integrators provide consulting services that are focused on optimizing the performance of Java-enabled applications and workloads through the use of software alone.

Azul Systems' solution is its own — based on an Azul Systems processor design and an approach to throughout optimization and memory management and garbage collection that provides a differentiated solution for Java-workload processing. However, the resources of larger competitors cannot be underestimated, given their worldwide advertising, marketing activities, and broad go-to-market capabilities. Azul can address these challenges by continuing to drive its own message consistently in the marketplace. Furthermore, it can enhance its position in the market by leveraging customer references, developing viral marketing campaigns, and enhancing its Partner Program to amplify this message in the worldwide IT marketplace.

It is critical for Azul Systems to explain the problem that it is seeking to solve because many of today's customers have taken other approaches to improving Java performance, including modifying the Java code itself; optimizing the performance of app servers that direct, and redirect, transactional workloads between servers; and orchestrating the workloads so that they are provisioned across available compute resources. All are valid approaches to improve Java performance. Rather, it is the degree of improvement, and the time-to-solution advantage, on which Azul's technology is focused, and the company points to a series of processing benchmarks to strengthen its case.

FUTURE OUTLOOK

The ability of Java-based applications to run across multiple types of server architectures is what has led to their extensive use worldwide, and to their importance in updating mission-critical databases and data warehouses. Java-based workloads are widely used across the enterprise — and they are accessed via ISV-based host application servers within the middle tier of multitier networks. The efficiency with which those Java-based applications can be run, and their functionality for end users, is of time-critical importance to many organizations.

Azul Systems is focused on enhancing the efficiency with which Java workloads are processed. Its compute appliance architecture provides a hardware-based platform that supports high performance, high availability, and high levels of security to the data being processed by Java-based workloads. In addition, its ability to interoperate with many types of systems across multiple operating systems in virtualized host server environments allows the Azul Compute Appliances to act as a leverage point within the enterprise, improving throughput and enhancing business results for Java-enabled workloads. Furthermore, this is done without disrupting the compute infrastructure that is already in place in the customer's datacenter.

Published Under Services: Enterprise Servers: Technology Markets; Enterprise Servers: Technology Markets With Vertical Views