THE CLIPPER GROUP

Navigator™

Navigating Information Technology Horizons℠

Published Since 1993 Report #TCG2005006 February 13, 2005

Azul Systems Offloads Enterprise Java Processing – And Changes How Enterprises Do Computing

Analyst: Anne MacFarland

Management Summary

Cost saving has become second nature in enterprise IT. Pervasive use of less than premium, *good-enough* computing components has been a natural result of limited or decreasing budgets and increasing business processes to be supported. But **there is a point when good enough really isn't good enough – when buying cheap is a waste of money.** As when you buy cheap clothing that must be dry cleaned, false economies cost dearly, over time. Azul Systems, located in Mountain View, CA, sees the typical application tier of commodity servers running J2EE or other virtual-machine based applications in demand-driven environments as a glaring case of the false economy brought on by the *good-enough* habit.

Azul feels that the multi-threaded nature of J2EE and other virtual-machine-based application processes is inadequately served by general-purpose processors - processors that are optimized to minimize single thread latency. Even those that purport to multi-thread do so at a more coarse level by multiplexing processes, not at a sub-process level. So, there is a lot of duplication of effort between the runtime and operating system instruction sets – akin to the unnecessary spinning of tires. More generally, Azul feels that application computing should not be constrained by the size of the application hosts an enterprise chooses to buy. **Box-based capacity planning for an application with unpredictable demand is inexact, and provisioning of commodity application servers errs on the side of generosity, leading to underutilization of assets. Failover and load balancing of application servers add to the complexity. Deft management of these environments, whether by humans of by software, is not cheap.** Overall, the use of *good-enough* commodity elements can be a frightful extravagance.

Azul proposes a comprehensive solution – a purpose-built compute pool architecture that Azul calls *Network Attached Processing*. Azul uses virtual machine proxies on application servers to redirect suitable application workloads to its *Compute Appliance*. The appliance processes these workloads more efficiently *and at a lower cost* than general-purpose processors. The application tier, relieved of bursty workloads, can be configured lean, and becomes easier to manage. This is an architectural-scale shift, not a product choice. It is very congruent with enterprise IT needs and solves some significant IT pains. For more details, read on.

IN THIS ISSUE

The Inherent Unpredictability of Demand-Driven IT Environments

Optimization, in the classic sense, only works with well-defined parameters, i.e., in stabile environments. The *wherever, whenever, immediate, self-service access* demanded by users of large-scale computer systems has led to an inevitably uneven demand on computing processes. And the way applications are being developed adds another dimension of unpredictability. The need to build new competitive advantages into their business processes has led to an enterprise need for faster (but safe) implementation, and more interoperability between components. Critical business applications, including applications from SAP, Microsoft, and Oracle, and those written on application platforms like IBM *WebSphere* and BEA *WebLogic*, are now written using virtual machines as runtimes to accommodate the plug-and play portability and degree of interoperability that modern business practices demand. The limited scalability of traditional IT optimization has led some enterprises to take a more architectural approach.

Enterprises have adopted two classic server architectures: *scale-out clusters* of commodity (low processor count) servers, and *fail-over clusters* of partitionable, high-processor count servers. Both architectures require significant tuning and management costs. Neither solves the problem of unpredictable demand, and the ensuing bottleneck that crashes the system. Bigger bottles and wider mouths can make the problem happen *less frequently,* but **it takes an offload architecture, like a bathtub's overflow aperture or a heating system's expansion tank, to prevent an over-capacity problem from happening.**

A History of IT Offloads

- In early IT systems, servers became a bottleneck to system expansion, so routers and switches were developed to offload networking procedures.
- Applications were limited in the amount of data they could directly support, so separate database applications were developed as offset application accessories.
- For similar reasons of scale, data storage was weaned from servers onto SANs.
- And, for a slightly different purpose, files to be widely shared were separated from the applications server that generated them and stored on NAS storage arrays, or behind NAS gateways, for easier, pervasive access.
- Virtual machines allowed applications to be segregated from the platforms they ran on, and to be run more densely according to their needs.
- Cloggy stuff, like TCP/IP and graphics processing, were offloaded onto TOE cards and graphics processors.
- In 2004, the IBM *zSeries* offered *a zAAP processor* to enhance mainframe use for new workloads.[1]

These and other trends in IT represent a long legacy of separating the function of compute from the computer in order to unsnarl the knots that block the flow of computing.

Azul's Compute Pool Architecture

Azul has taken the offload concept and expanded it from an accessory function to a core capability. Azul bifurcates the virtual environments in which J2EE applications run into a proxy stub, run on the application server[2], and the runtime, offloaded to their purpose-built Compute Appliance.

The Front End of the Azul System

With the computing workloads offloaded, enterprise application servers become a bunch of front-end gateway-style mount points at which I/O, security and

[1] See **The Clipper Group Navigator** dated April 7, 2004, entitled "*zSeries Zips Through Java with zAAP*," available at http://www.clipper.com/research/TCG2004030.pdf.

[2] The Azul proxy on these application servers does not use local resources.

application patching are done.[3] With the compute workloads offloaded, the enterprise may need fewer instances of an application. With a reduced compute function, the application servers don't have to exist physically as separate nodes. A whole raft of application servers can be co-resident virtual machines on a single server. This server consolidation gives obvious CAPEX and OPEX savings.

The Azul Compute Appliance

The Azul Compute Appliance executes the application, interacting with the outside world through the mount point of the proxy on the application server. Because the compute appliance has no state, it does not add a security exposure to be managed. Because its binaries are not exposed, its processors can work with any operating system, as is needed.

The Compute Appliance is a self-sufficient, self-reliant computing brick, available in increments granular enough to satisfy the smaller enterprise and scaling to sate the highest demand. The backplane is multi-terabyte. The connection, in the first generation product, is Ethernet. The 96-processor entry-level model of the appliance is a 5U unit that consumes only 600 watts of power. **The Compute Appliance can scale from 96 to 384 processors, at which size it has 3½ times the threaded compute capacity of any existing server. It is easy to imagine how such an environmentally-polite but enormously-potent appliance could change what your run where, and, indeed, the very nature (and size) of the environment in which computing is done.** Energy use savings alone would be huge.

The Azul NAP has policies to enforce absolute guarantees of capacity, priorities, and the need of each application for headroom. The compute pool provides a single, large, coherent image for each server. As Azul CEO Steven DeWitt put it, "Azul is effectively a Layer 7 compute-coherent switch."

As a system element, the NAP compute pool has the instrumentation to be managed as a pool, satisfying the need for performance tuning. It has the SNMP and MIB-2 hooks to be managed in a larger environment. When another brick of compute capacity is added, it advertises itself and joins its colleagues. If a brick fails, its cohorts pick up the slack.

Azul's Compute Appliance has been tested with IBM *WebSphere,* BEA *WebLogic,* and *JBOSS*. Azul hopes to support *.Net* in the future.

Benefits of the Azul System

What Azul has created is more than a product. It's a system. Few people build systems these days, because there are many components involved that all must be designed right. But, if you have the experience, expertise, and patience to do it, getting all those things right can bring benefits that go beyond the first order benefits of reduced costs that are outlined above.

- **Azul's purpose-built, 24-core chip processes virtual machine workloads more efficiently than can be done by traditional instruction sets**. For instance, the processors do not pause for garbage collection. It happens concurrently, as do other utility functions.
- **Azul processors are designed to meet the multi-threading needs of J2EE and .Net applications.** Processors classically have been designed to minimize single-threaded latency. The current multi-threaded architectures are designed to use those single-threaded architectures more efficiently by interleaving (multiplexing) processes, which does little to optimize threads within a process, as are found in J2EE and .Net environments.
- **An application's share of this pool of processing can expand, in real time, to meet the needs of the application.** Azul

[3] This is an immediate source of both CAPEX and OPEX savings, particularly if the alternative is a server farm with a high rate of swap-in, swap-out.

calls this capability *Unbound Compute*. Because the unit of expansion is a self-managed appliance of focused capability, it does not incur the overhead of multiple instances of device-specific management overhead that general-purpose nodes of a server farm, or the partition management needed by large UNIX and Intel servers. The elasticity comes just where it is functionally required.

- **Azul's compute pool allows an off-loaded point of control**. The unified command and control structure over a reconfigurable processing environment for multiple applications that results would otherwise be very difficult to achieve in distributed situations. It allows instrumentation exactly where it is needed.

Deploying the system

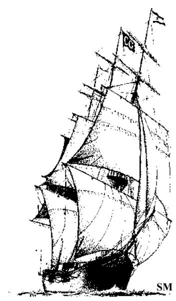
Installing Azul means no change to the existing system (beyond a download of proxies on the servers involved), and no porting. The Compute Appliance plugs in like any node on the LAN. You do not open the appliance. It is pre-configured to process J2EE workloads. Open standards do have great benefits. This is one of them. The appliance may be extremely proprietary in its inner configuration, but, in the environment, it just works.

Go to Market

Azul's go-to-market plan is, initially, direct-to-buyer in the US and through partners in Japan, the two geographies where Azul is being customer-tested. Azul will expand its efforts to Europe in the spring.

Conclusion

Enterprises have recently seen the traditional partitions of business units become obstructions to getting the common functionalities of the enterprise efficiently done. Extracting common functionalities from fiefdoms has been hard, not because business units crave doing mundane things, but because it requires a different way of thinking. Similarly, Azul, though it is a product, requires a rethink at the system level. It is revolutionary on a larger-than-product scale – but it also is a revolution not just of design, but of the enterprise mindset, as well. If your enterprise has significant J2EE workloads (and most large enterprises do), it is time to revise your mindset and figure out all that the Azul Compute Appliance could do for your environment.



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent 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

***Anne MacFarland* is Director of Enterprise Architectures and Infrastructure Solutions for The Clipper Group.** Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

- ***Reach Anne MacFarland via e-mail at Anne.MacFarland@clipper.com or at 781-235-0085 Ext. 28. (Please dial "1-28" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group *Captain's Log***, 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.

Disclosure

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.

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.