

QLogic's 3GCNA: SAN Leadership Meets LAN Convergence

By Bob Wheeler
Senior Analyst

October 2010



The Linley Group

www.linleygroup.com

This paper examines QLogic's third-generation converged network adapter (CNA) technology, which combines elements from the company's Fibre Channel HBA, iSCSI HBA, and intelligent NIC product lines.

Introduction

On October 7, 2010, QLogic announced a portfolio of third-generation converged-networking products under the code name 3GCNA. The new 8200-series 10GbE CNAs represent the next generation of QLogic's 8100-series CNAs. Thanks to its first-to-market availability and strong software support, the 8100 was the CNA market-share leader for 1H10. QLogic has CNA design wins at the top four server vendors. The 8200-series adds full iSCSI offload, advanced TCP/IP offloads, and new virtualization features while improving networking performance.

The new 3200-series 10GbE Intelligent Ethernet Adapters are the logical successors to QLogic's 3100-series NICs. Compared with the 8200 CNAs, the 3200 NICs omit full Fibre Channel over Ethernet (FCoE) and iSCSI offload but offer the same advanced TCP/IP features.

In addition to CNAs and NICs, QLogic is offering its 3GCNA technology in the form of a controller chip for what it calls converged-LOM (cLOM) designs. The 8200 cLOM chip comes in several versions optimized for specific LOM configurations. The cLOM chips support the same protocols as the 8200 CNAs.

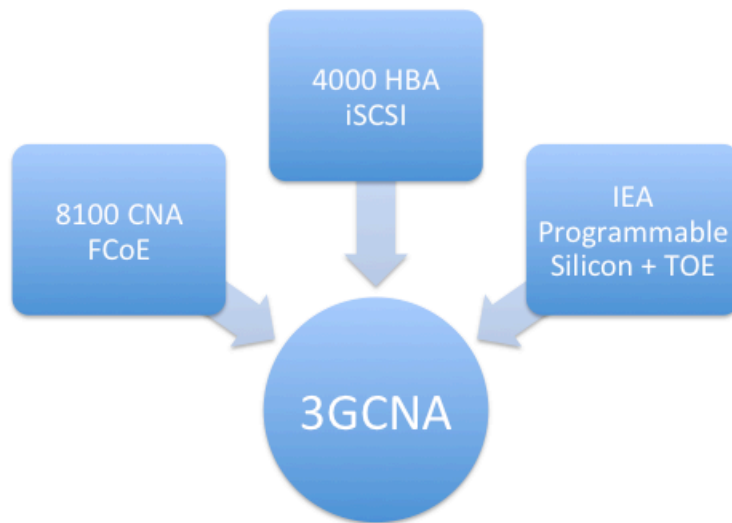


Figure 1. Key technology elements of QLogic's 3GCNA design.

QLogic's CNA and NIC/LOM Technologies

Before examining the details of the new 3GCNA products, we review QLogic's existing CNA and network-adapter technologies and how these feed into the new unified design. Specifically, the 3GCNA builds on QLogic's Intelligent Ethernet Adapter (IEA) silicon architecture, the same FC HBA drivers used by the 8100-series CNAs, and an iSCSI stack from the 4000-series HBAs. QLogic's industry-leading iSCSI stack has shipped with multiple HBA generations since 2003.

QLogic's 8100-series second-generation CNAs have been in production since 2Q09. The 8100 cards were the industry's first single-chip CNA design, using a single QLogic ISP8112 controller chip to perform both FCoE and NIC functions. Based on QLogic's proven FC technology, the 8100 CNAs offer the industry's broadest HBA driver support, and they come with QLogic's SANsurfer HBA Manager. The 8100 implements a relatively simple NIC function, which supports RSS and NetQueue as well as basic stateless offloads including LSO/GSO and TCP/UDP/IP checksums. To support FCoE, the 8100 implements PFC, ETS, and DCBX.

QLogic's IEA technology comes from the April 2009 acquisition of 10GbE startup NetXen, which had already shipped multiple generations of chips or adapters to leading OEMs including HP, IBM, and Dell. The shipping 3100-series IEAs are based on the NX3031 controller chip, which supports dual 10GbE ports, four GbE ports, and PCI Express v2.0. The chip uses external DRAM to buffer packets and store packet context for a large number of connections. HP is also shipping the NX3031 in ProLiant DL/ML370 G6 server LAN-on-motherboard (LOM) designs as well as in its NC522SFP NIC.

The NX3031 is built around five custom processors, which NetXen called MetaCores. The chip also includes an encryption engine designed for 10Gbps throughput, but QLogic has not enabled this feature through software/firmware. The NX3031 implements dynamic power-management features that reduce typical power dissipation by about 50% when operating in a GbE-only configuration. QLogic uses the internal programmability of its design to deliver advanced protocol offloads such as Windows TCP Chimney and Large Receive Offload (LRO). QLogic supports the NX3031 with drivers for Linux, Solaris, VMware ESX, Windows, and XenServer.

3GCNA: Behind the Adaptive Convergence Curtain

In launching its 3GCNA, QLogic stressed flexibility as being critical during an industry technology transition. Under the umbrella of Adaptive Convergence, the company branded key features and functions tied to flexibility. Here, we examine the technology behind the brands.

VMflex: Virtualizing the NIC/HBA

Anybody working around data centers and servers is well aware of the server virtualization trend. But the effects of virtualization on the network stack and on I/O performance are less well understood. As virtualization increases server utilization, it also increases the importance of I/O processing overhead, as fewer “free” CPU cycles are available.

In virtualized environments running Citrix XenServer, Microsoft Hyper-V, or VMware ESX, the physical NIC becomes isolated from the guest operating system. The hypervisor translates requests from the guest OSs into requests for a physical NIC. In addition to allowing multiple guest OSs to share a single physical NIC, the hypervisor can emulate an Ethernet (Layer 2) switch connecting virtual machines to each other and to physical NIC ports. Virtual-NIC and virtual-switch emulation is performed in software, thereby adding a great deal of processing overhead in the network path.

Virtualization-software vendors have been working with NIC and HBA vendors to develop hardware-offload techniques, which offset I/O-virtualization overhead. For example, VMware added NetQueue hypervisor offloads to ESX Server 3.5, boosting 10GbE performance nearly to wire speed but still consuming many CPU cycles. The next step in I/O virtualization is to bypass the hypervisor altogether, allowing the guest OS to directly access a virtual NIC. The single-root I/O virtualization (SR-IOV) standard for PCI Express enables this next-generation offload technique.

An SR-IOV device has one physical function (PF) and a number of virtual functions (VFs). The hypervisor uses the PF to manage the VFs, each of which maps to a virtual machine. Each VF replicates the resources required by the VM. Once the SR-IOV device is configured, the VMs can independently access their associated VFs. This enables a hypervisor-bypass or direct-access model, which removes the device-emulation overhead normally associated with sharing a device through the hypervisor. Because hypervisors normally handle VM-to-VM traffic, hypervisor-bypass architectures must push VM-to-VM switching out to the NIC or an adjacent physical switch.

QLogic's 3GCNA implements state-of-the-art I/O virtualization, including support for hypervisor offloads and hypervisor bypass using SR-IOV. The 3GCNA design supports 64 VFs, matching the maximum number of threads in shipping four-socket servers. Each VM can be allocated the amount of bandwidth and the traffic priority required by its specific workload. The 3GCNA's programmability is crucial to supporting the virtual L2 switch function, which is designed to work with any vendor's 10GbE switch. Whereas SR-IOV is already standardized, competing VM-to-VM switching approaches are being developed under IEEE 802.1Qbg and 802.1Qbh. QLogic can support these emerging standards using firmware upgrades, easing customer concerns about future switch interoperability.

Although hypervisor bypass promises greater performance, SR-IOV requires as-yet unavailable support from OS vendors. In the meantime, QLogic's VMflex technology

enables the 3200 and 8200 adapters to support up to eight virtual ports without OS changes. Each physical 10GbE port appears to the OS as four virtual ports. Users can then divide the 10Gbps of available bandwidth per port by assigning guaranteed bandwidth amounts to each virtual port in 100Mbps increments. As with SR-IOV, the 3GCNA's virtual L2 switch forwards traffic between the virtual ports.

This virtual-port feature provides users with a straightforward upgrade path from GbE to 10GbE. In ESX environments, common practice is to use multiport GbE NICs and dedicate GbE ports to specific functions such as VM traffic, VM migration (VMotion), and management. With VMflex, a single QLogic adapter can replace multiple GbE NICs by dedicating a virtual port to each function. Furthermore, bandwidth can be distributed flexibly to optimize system performance rather than being fixed at 1Gbps per function.

ConvergeFlex: Concurrent FCoE, iSCSI, and TCP/IP

Because FC compatibility was paramount, most early CNA designs were based on existing FC HBA architectures combined with a simple L2 NIC function. QLogic's 8100-series CNA is an example of such a dual-function design; it was compatible with existing FC drivers and integrated a basic Ethernet controller. The only option for customers wishing to support iSCSI in addition to FCoE was to use an iSCSI software initiator.

In an industry first, QLogic's 3GCNA supports three functions simultaneously: FCoE using an FC HBA driver, iSCSI using an HBA driver, and an advanced NIC function. The FC function uses the same proven and qualified FC drivers as the 8100 series, but these drivers have been ported to the 3GCNA's new architecture. Although the 8100 and 3GCNA use different firmware, this difference is transparent to operating systems and management software. Similarly, QLogic developed iSCSI firmware for the 3GCNA design that provides compatibility with field-proven 4000-series iSCSI HBA drivers. To provide simultaneous operation, the FCoE and iSCSI firmware are, in fact, combined. Replacing the SANsurfer HBA manager is the new QConvergeConsole utility, which enables single-pane-of-glass management for FCoE, iSCSI, and NIC functions on 3GCNAs as well as FC HBAs.

In a single-application server, few customers require FC and iSCSI SAN support at the same time. With the adoption of virtualized servers that support VM migration, however, application workloads are no longer tied to physical servers. As an application migrates from one physical server to another, it requires access to the same storage resources regardless of its physical location. The 3GCNA allows VMs to access FC or iSCSI storage without rebooting the server to change protocols. In a private data center, this flexibility eliminates downtime when moving workloads across servers. In a public cloud-computing environment, simultaneous multiprotocol support enables a unified server design regardless of physical storage resources or application workloads.

FlexOffload: Delivering Application Performance

QLogic uses FlexOffload to describe the 3GCNA's protocol-offload features. These features include full offload of FCoE and iSCSI protocol processing as well as multiple offload techniques for TCP/IP. Using an HBA driver for FCoE, the 8100-series CNA fully offloads FC protocol processing. For iSCSI, however, the 8100 requires a software initiator, which performs iSCSI protocol processing on the host processor. Although iSCSI software initiators are popular for GbE speeds, this protocol processing consumes significant CPU resources at 10GbE data rates. With the 3GCNA, an iSCSI HBA driver fully offloads this processing from the host, leaving processor cycles available for applications.

For file-based storage protocols, such as NFS over TCP, the 8100 provides little in the way of hardware offloads. Conversely, the 3000-series Intelligent Ethernet Adapters offloaded TCP processing from the host but did not support FCoE or iSCSI HBA drivers. As described above, virtualized environments need the flexibility to access the storage resources required by a given workload regardless of what physical server is running the application. By concurrently supporting TCP offload in parallel with full FCoE and iSCSI offload, the 3GCNA minimizes server CPU utilization for both file- and block-storage protocols.

Avoiding proprietary TCP stacks, the 3GCNA supports TCP offload using standard operating-system TCP/IP stacks. Specifically, the 3GCNA supports LRO under all operating systems and Microsoft TCP Chimney under Windows. LRO partially offloads TCP-receive processing by performing segment reassembly for a small number of connections. Whereas the Linux kernel supports a software LRO implementation, other operating systems currently require the NIC hardware to perform LRO such that segment reassembly is transparent to the TCP/IP stack. By implementing hardware LRO, the 3GCNA can offload TCP reassembly across Linux, VMware ESX, and Windows environments.

Like the common LSO/TSO and TCP/IP-checksum offloads, LRO remains only a partial offload. Fully offloading TCP processing to the NIC requires a TCP-offload engine (TOE) capable of maintaining TCP state for a large number of connections. Although TOE implementations have been available for years, Microsoft's TCP Chimney remains the only TOE-enabled networking stack in a mainstream OS. For this reason, the 3GCNA supports full TCP offload in Windows environments but only partial TCP offloads in other environments. As network stacks evolve, the 3GCNA's programmable architecture can implement new and improved offload techniques through firmware and driver upgrades.

For years, TOE-NIC vendors argued that TCP processing in hardware provided lower server-CPU utilization and reduced system power consumption. But in the old world of single-application servers, CPU utilization was rarely a limiting factor for real-world performance. Today, virtualized servers are achieving utilization of about 90%. In this context, protocol offloads can directly affect delivered performance. Instead of

performing simple benchmarks showing maximum throughput or IOPS, vendors and customers need to begin benchmarking application performance for real-world workloads. This type of benchmarking is already prevalent in the world of high-performance computing and has proven valuable in exposing the true performance of I/O technologies and protocols.

SecureFlex: Securing the SAN

The key to SecureFlex, according to QLogic, is what it is not: it is not proprietary. SecureFlex is, quite simply, support for industry-standard security protocols for protecting data in flight. For FCoE, the 3GCNA supports the FC Security Protocol (FC-SP). By authenticating messages, FC-SP prevents man-in-the-middle attacks. Because iSCSI runs on top of TCP/IP, IPsec is the underlying security protocol. The 3GCNA architecture is designed to support IPsec, but QLogic has not yet announced software/firmware support for this feature. QLogic's standards-based approach enables interoperability with storage arrays from various vendors, which can then add data-at-rest encryption as a part of their system's value.

FlexLOM Meets cLOM

QLogic's new cLOM chip brings all of the features and capabilities discussed above to LOM designs. For server OEMs, the cLOM chip provides a single hardware design capable of supporting FCoE, iSCSI, and advanced offloads for TCP/IP and virtualization. Furthermore, one firmware load supports all features, avoiding the need to change firmware personalities when customizing features.

Another key cLOM feature, branded FlexLOM, carries over from the NX3031. The FlexLOM design enables OEMs to ship a 4xGbE LOM design that is upgradeable to 10GbE using a daughtercard. HP adopted this approach in its ProLiant DL/ML370 G6 servers, which ship with 4xGbE ports and use the low-cost NC524SFP option to add a pair of SFP+ 10GbE ports. The FlexLOM design starts by placing the cLOM device on the server motherboard along with a third-party GbE-over-copper quad PHY and four RJ45 ports. In addition, the cLOM chip's 10GbE ports and a few control signals are routed to a daughtercard connector. The optional daughtercards include media-specific PHYs and connectors for optical, UTP, or twinax cabling. This approach decouples the 10GbE PHY from the motherboard design, thereby reducing entry cost and improving media flexibility.

Compared with the NX3031, the new cLOM chip improves integration by reducing external-memory requirements and integrating optional PHYs. QLogic is offering several cLOM variants optimized for specific LOM implementations. For FlexLOM designs, the base cLOM chip provides the smallest footprint while omitting 10GbE PHYs. For blade-server designs, QLogic offers a cLOM chip with integrated 10GBase-KR PHYs for direct connection to a backplane. Finally, for emerging rack/tower LOM designs, the company offers a cLOM device with two 10GBase-T ports that support RJ45 connectors.

Conclusions

QLogic's 3GCNA design is significant both to the company and to the broader industry. For QLogic, the 3GCNA unifies its CNA and NIC product lines around one chip architecture. In doing so, the company has added strong TCP-offload features to its already well-proven FCoE capabilities. The IEA architecture's programmability also allowed QLogic to port its existing iSCSI stack to the 3GCNA, giving the company its first product with 10GbE iSCSI HBA capability. Whereas the 8100 CNA principally competed with other FC-centric CNAs, the 3GCNA enables QLogic to compete equally well for iSCSI HBA and Ethernet-centric 10GbE designs.

From an industry perspective, the 3GCNA stands out as the first product to support concurrent FCoE, iSCSI, and TCP offload. Currently, only one competing product supports full iSCSI and FCoE offload; that product, however, supports only one block-mode protocol (FCoE or iSCSI) per adapter. Although mixing FCoE and iSCSI on a physical server is rare today, VM migration is breaking the traditional linkage between physical servers and application workloads. Cloud computing, in its ultimate instantiation, promises I/O resources that dynamically follow workloads as they move around the pool of compute resources.

By making the 3GCNA available as a chip-level product, QLogic is giving its OEM customers added flexibility in how they configure their systems. Server vendors are using various methods to deliver 10GbE in current-generation servers, including standard form-factor NICs, proprietary NICs and mezzanine cards, and blade-server LOM designs. QLogic's FlexLOM design provides the added option of a 10GbE-capable field-upgradeable rack/tower LOM design. By requiring only a media-specific daughtercard to enable 10GbE, QLogic's design provides the lowest-cost 10GbE upgrade while also preserving PCIe slots.

The dual trends of server virtualization and network convergence are driving major changes in server I/O requirements. In this dynamic market environment, programmable designs are best able to adapt to changing requirements. With the 3GCNA, QLogic has developed a powerful design that protects customers' investments through its flexible architecture.

About the Author

Bob Wheeler is a senior analyst at The Linley Group and co-author of A Guide to 10G Ethernet Controllers and Adapters. The Linley Group offers the most comprehensive analysis of the networking-silicon industry. We analyze not only the business strategy but also the technology inside all the announced products. Our in-depth reports cover topics including Ethernet chips, network processors, high-speed embedded processors, and handset processors. For more information, see our web site at www.linleygroup.com.

Trademark names are used throughout this paper in an editorial fashion and are not denoted with a trademark symbol. These trademarks are the property of their respective owners.

This paper is sponsored by QLogic, but all opinions and analysis are those of the author.