The Dell Force10 ExaScale E-Series

A Dell Technical White Paper

September 2011

Contents

Figures

Overview

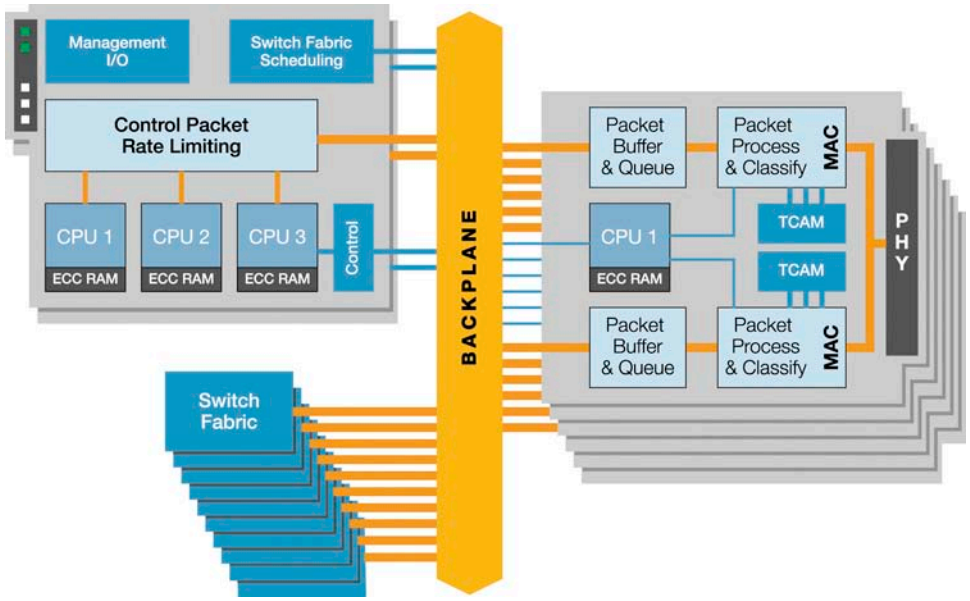
The Dell Force10 ExaScale E-Series virtualized core switch/router provides the unmatched scalability, line-rate performance, and full L2 switching, L3 routing and MPLS functionality that is essential for today's most demanding network applications. To provide this unprecedented combination of power and economy, at never-before-realized price/performance ratios, the ExaScale E-Series architecture leverages innovations in switch-fabric, backplane, Application-Specific Integrated Circuit (ASIC), and system-control-plane design.

Scalable, High-performance Ethernet

The ExaScale E-Series provides unmatched scalability to 1,260 Gigabit Ethernet (GbE) or 140 10 Gigabit Ethernet (10 GbE) ports per chassis, line-rate forwarding with Access Control Lists (ACLs) on all ports, and comprehensive L2 switching and L3 routing. The ExaScale E-Series architecture delivers these capabilities via a number of technology breakthroughs:

- The ExaScale E-Series switch fabric, featuring N:1 redundancy, advanced queuing, multicast, and jumbo frame support, delivers unmatched scalability with up to 3.5 Tbps of non-blocking connectivity.
- An industry-first 100 Gbps of data capacity per slot (125 Gbps/slot raw capacity) equips the switch fabric to support 40 GbE and 100 GbE in the future.
- An industry first, a passive copper backplane reliably cost-efficiently provides 125Gbps between line card slot — without the cost and complexity optical or active backplanes.
- Distributed ExaScale ASICs — purpose built for switching Ethernet frames and routing IP— provide full hardware support for line-rate processing, classification, buffer and traffic, scheduling, and switching.
- A high-performance multiprocessor control plane line-rate L2 switching, routing, and MPLS full L2/L3/MPLS protocol and feature support.Encryption

Figure 1. Dell Force10 ExaScale Architecture



Line-rate performance, cost-effective scalability, and robust L2 switching and L3 routing:

- Scalable, non-blocking switch fabric enables the low latency and jitter critical for streaming media applications
- Cost-effective, reliable passive copper backplane maximizes system reliability and minimizes cost
- High-performance Dell Force10 ASICs distribute packet forwarding, ACL processing, Quality of Service (QoS), and buffering to every line card
- Robust L2/L3/MPLS multiprocessor control plane with innovative control traffic filtering and rate limiting capabilities

The Dell Force10 ExaScale E-Series combines these elements of the architecture with high-availability features such as hot-swappability of all key components and system-wide environmental monitoring. This combination enables the platform to deliver maximum system uptime and serviceability along with unprecedented performance and scalability. As a result, the ExaScale E-Series can bring the simplicity and economics of Ethernet to some of today's most demanding network applications, including virtualized data center, Cluster/Grid Computing, next-generation Internet Exchanges (IXs), cloud computing networks, and Enterprise Backbones.

Scalable Switching Capacity

Non-Blocking 3.5 Tbps Switch Fabric

The switch fabric provides 3.5 Tbps non-blocking switching capacity along with support for advanced queuing, multicast, and jumbo frames. The Switch Fabric Modules (SFMs) that comprise the switch fabric are N:1 redundant and hot swappable, maximizing the ExaScale E-Series' system availability. The platform provides 125 Gbps of full-duplex bandwidth to each line card slot. This ensures line-rate performance on up to 1,260 GbE ports or 140 10 GbE ports in a single ExaScale E-Series chassis and leaves customers with enough headroom to double the system's port density in the future without the hassle of a forklift upgrade. The switch fabric's N:1 redundant design enables the platform to provide redundancy without replicating the entire switch fabric complex — a costly strategy necessitated by other architectures.

The switch fabric is designed to forward full Ethernet and IP frames in hardware, switching them without any need for segmentation and reassembly. The Virtual Output Queue (VOQ)-based crossbar switching architecture overcomes the cost and complexity of multi-level or clustered switch fabrics.

In contrast to previous switch/router architectures, the scalable VOQ fabric model enables high-performance switching of complete packets across the backplane. This capability enables the ExaScale E-Series switch/ routers to scale well beyond software-based forwarding architectures, eliminate the throughput limitations of bus-based architectures, and avoid the non-deterministic characteristics of cell-based architectures (such as sequencing and reassembly issues).

Passive Copper Backplane

The reliable and cost-efficient backplane is the industry's first high-speed non-optical backplane to achieve 3.5 Tbps in a half-rack switch/router chassis. Unlike optical backplane interconnect systems or active copper backplanes, the patent-pending backplane has no single points of failure and eliminates costly electrical-optical-electrical conversions. The resulting system simplicity means bulletproof reliability, system simplicity, and minimum cost.

True Line-rate L2/L3 Forwarding

Distributed Forwarding with Deterministic Low Latency

ExaScale ASICs are present on every line card, providing predictable, hardware-based distributed packet processing up to 2,000 Mpps. Unlike centralized architectures, the distributed processing architecture enables high throughput and deterministic performance even as more ports are added to the system and as more traffic load is added to the network. In contrast with low-performance processor or route-cache-based forwarding architectures, the ExaScale E-Series has no "slow-path" or software-based forwarding.

The classification ASICs, along with advanced Ternary Content Addressable Memories (TCAM), allow line-rate lookups within L2, L3, and MPLS forwarding tables. Similarly, as packets enter the system, the classification ASICs enable on-the-fly line-rate lookups of ACL entries for destination, policy and Quality of Service (QoS) mappings. This simultaneous packet processing and classification allows the ExaScale E-Series to provide line-rate L2/L3/MPLS forwarding performance independent of table lengths, IP address prefix lengths, or packet size — even when all ACLs and QoS features are enabled. The hardware forwarding afforded by the ExaScale ASICs gives the E-Series the ability to combine the true line-rate performance and low, deterministic latency and jitter required to build server aggregation networks in next-generation streaming media applications.

Line-Rate with All L2/L3/MPLS Features

Perhaps the biggest problem with today's switch/routers is the performance degradation they suffer when L2/L3/MPLS features are turned on. As opposed to these traditional architectures, the ExaScale E-Series does not compromise on features to provide line-rate forwarding performance. The ExaScale ASICs have line-rate support for filtering, statistics collection, QoS, rate policing, and limiting. ExaScale ASICs also deliver protocol-specific hardware support at line-rate for L2 switching, L3 routing, and MPLS.

The ExaScale E-Series' line-rate features include:

- Filtering with standard and extended ACLs
- QoS (DiffServ, IEEE 802.1p)
- Rate policing and limiting
- L2 switching features
    - Source address learning and limiting
    - Link Aggregation
    - VLAN stacking
- L3 routing features
    - ECMP
    - Inter-VLAN routing
    - IP multicast
- MPLS features
    - Transit LSP
    - Push, Pop, Swap, Classify Labels
    - FRR
- Statistics collection

Service-Aware QoS Architecture

The ExaScale E-Series provides extensive QoS and traffic management capabilities designed according to DiffServ and IEEE 802.1p requirements.

The ExaScale ASICs provide robust packet marking capabilities, including the ability to re-map packet priorities between IP and Ethernet schemes. Traffic conditioning is based upon two-rate, three-color token bucket-based metering and marking. Eight queues per destination port map directly to class-based DiffServ and IEEE 802.1p queuing models. Congestion avoidance is enabled by configurable drop precedence probability curves of Weighted Random Early Discard (WRED). The combination of these features enables Committed Access Rate (CAR)-based service offerings with rate policing and limiting.

QoS is also built into the switch fabric. Both ingress and egress buffering are provided, including backpressure mechanisms that ward off the possibility of head-of-line blocking. Separate unicast and multicast queues with 82 milliseconds of buffering enable minimal packet loss even in oversubscribed network conditions. The ExaScale E-Series switch fabric uses Interleaved Weighted Fair Queuing (IWFQ) to schedule traffic out of the ingress and egress queues, and programmable queue sizes allow seamless handling of both real-time and bursty traffic patterns.

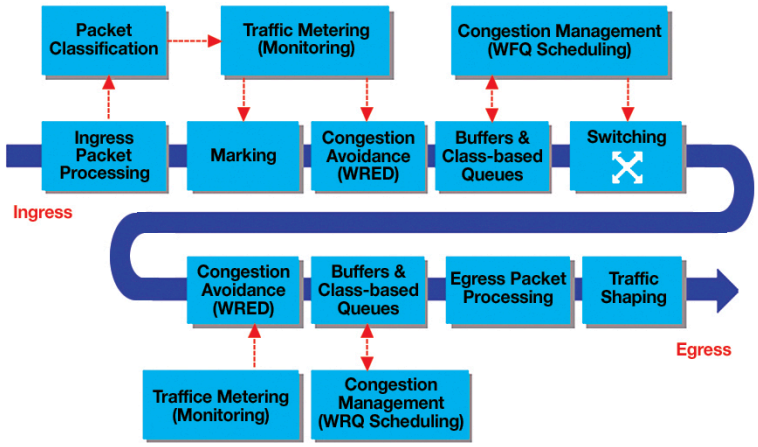
With the service-aware QoS capabilities of the ExaScale E-Series architecture, service providers can honor customer-defined traffic priorities, or assign their own class of service policies to enforce simple and manageable Service Level Agreements (SLAs).

Robust System Control Plane

Distributed Multiprocessor Control Plane

With three processors on each Route Processor Module (RPM) and one processor on every line card, the ExaScale E-Series system control plane is purpose-built to deliver high-performance and fault-tolerance to the full Dell Force10 Operating System (FTOS™) suite of L2 switching features and L3 routing protocols. Designed to meet the needs of Internet-scale networks, the system control plane supports millions of routing table entries, up to 768K forwarding table entries, and tens of thousands of ACLs on every line card.

Figure 2. QoS Within the Dell Force10 ASICs



Upon receipt, the packet classifier ASIC may mark or re-mark the packet based on configured policy including Multi-Field filtering (an ACL matching on source, destination, protocol, and port), Virtual Local Area Network (VLAN), or physical port mapping. After classification, a two-rate, three-color marker mechanism built into the ASICs performs ingress traffic shaping. The packet is then placed in the appropriate output queue. Queues are serviced by the IWFQ algorithm and sent across the fabric to the egress queues. A second, identical, two-rate, three-color marker mechanism performs traffic shaping on egress.

Software processes are distributed among the processors, allowing true real-time multiprocessing. This enables process isolation with memory protection. These features are absolutely necessary for fault tolerance and rapid convergence in large-scale enterprise and service provider networks.

Resilience and Security

The RPMs provide innovative traffic control, rate limiting, and filtering. These capabilities empower network administrators to suppress harmful Denial of Service (DoS) attacks and prevent flooding of unwanted traffic onto the network — an event that places an unnecessary burden on control processors. Dedicated 100 Mbps switched paths from the RPMs to every line card eliminate sluggish forwarding table updates that could otherwise jeopardize network stability.

Fault Tolerance and High Availability

To maximize network uptime, the E-Series architecture supports redundancy, availability, and serviceability features. All key components are redundant, including the RPMs, SFMs, power, and cooling components. All memory systems are ECC/parity protected. System-wide environmental monitoring and persistent configuration synchronization enable the FTOS to detect, report, and correct faults with minimum system interruption. In addition, serviceability features such as hot-swapability of all key components, cable management, and front-side access to all cabling and cards minimize mean time to repair.

Full L2/L3/MPLS Functionality

Reliable Real-Time Operating System

Dell Force10 FTOS software is purpose-built for scalable, high-performance Ethernet applications that span the LAN, MAN, and WAN. FTOS harnesses the massive performance of the ExaScale switch/router platform and provides end users with the functionality they need to utilize the power of the E-Series architecture.

A real-time operating system, FTOS is customized for an extensive range of high-performance L2 switching, L3 routing, and MPLS features. Process modularity and distribution in a protected multiprocessor environment ensures stability and fault tolerance. Critical FTOS features include robust IP routing control plane, hardware and software fault-tolerance, highly granular traffic management and accounting, industry standard Command Line Interface (CLI), and system diagnostics.

Extensive L2/L3/MPLS Control Plane Features

FTOS works in harmony with the hardware-assisted per-packet features provided by ExaScale ASICs. As a result, FTOS provides the system's L2/L3/MPLS control plane features and functionality, making it an integral part of the E-Series architecture.

In the case of L3, FTOS delivers robust routing protocols — including Border Gateway Protocol 4 (BGP4), IS-IS, Open Shortest Path First (OSPF), and Routing Information Protocol 2 (RIP) v2 — that scale to

millions of routes and 512K forwarding entries. Virtual Router Redundancy Protocol (VRRP) and Equal Cost Multi-Path routing (ECMP) in FTOS enable reliable network design while support for IP Multicast — ranging from Internet Group Management Protocol (IGMP) and Protocol Independent Multicast-Sparse Mode (PIM-SM) to Multiprotocol Border Gateway Protocol (MBGP) and Multicast Source Discovery Protocol (MSDP) — make the ExaScale E-Series the platform of choice in data center, enterprise backbone and service provider network core applications.

In the case of L2, FTOS scales to thousands of MAC entries. It also supports 4096 VLANs, VLAN tagging and stacking, load sharing and fail-over with Link Aggregation (LAG), and Rapid Spanning Tree Protocol (RSTP).

In the case of MPLS, FTOS delivers LSR (Label Switch Router) functionality with Aggregation, Fast Reroute, Node/Link Protection, RSVP-TE, LDP, OSPF-TE, IS-IS-TE, MPLS over LAG, LAG Hashing on Labels, and 8 queues.

Flexible Services and Management

Customers demand flexibility, and the Dell Force10 ExaScale E-Series gives them the tools they need. These features include congestion control with WRED and WFQ, QoS interworking between L2, L3, and MPLS, CAR (policing), and detailed statistics for accounting and billing.

Integral to the architecture of the ExaScale platform are its flexible and secure management capabilities. User access can be authenticated via RADIUS and TACACS+, while secure access methods include Secure Shell (SSH) and Secure Copy (SCP). Built in support for SNMP MIBs and a dedicated 100BaseT management port allows seamless integration in any in-band and out-of-band Network Management environment.

Leader in Virtual Services

The ExaScale E-Series was engineered with virtualization services in mind. Current and future services in Force10’s virtualization framework include VLAN support, VRF (Virtual Routing and Forwarding) support to deliver multiple routing domains on one chassis as part of VirtualControl, Virtual Chassis that supports multiple chassis behaving as one while being controlled by a single master as part of VirtualScale, and sFlow support as part of VirtualView. With the ExaScale E-Series, these services are tools that can be utilized by network architects and administrators to reduce complexity, and streamline data flows through the data canter, while increasing reliability and performance. Refer to the white paper " Dell Force10's FTOS Delivers End-to-End Reliability While Lowering Network TCO" for more information on Dell Force10's virtualization framework.

Leader in Automation Services

The ExaScale E-Series has also been engineered with open, standards-based automation as a primary design goal. As such, it strives for a complete and robust solution as it interfaces and cooperates with emerging server, network and desktop virtualization technologies. Collaborating with server hypervisor or virtual switch resource provisioning functionalities, Dell Force10 is striving to ease the load and reduce the risk of heretofore required administrative intervention. The network should detect and reflect the needs of cloud based applications and services. As part of FTOS, Dell Force10 has developed an open automation framework to deliver the necessary API and hooks consisting of two main elements. One element of this framework bridges the gap between the network and virtualized server technologies, enabling these tools to push new configurations to the network, thereby providing a closed-loop provisioning system. The second element empowers middleware by providing access to critical switch/router resources to report and respond as resource needs change. Dell Force10's

standards-based automation technology is designed to transform data center infrastructure into a self-optimizing, highly available and accountable cloud computing fabric.

Conclusion

The Dell Force10 ExaScale E-Series architecture enables enterprises and service providers to build scalable, resilient, and high-performance Ethernet networks. Technological innovation across a range of areas enables the ExaScale platform to deliver a scalable 3.5 Tbps switch fabric, high-performance ASIC-based distributed forwarding, a robust control plane, and the feature-rich FTOS operating system.

The ExaScale E-Series has a per slot capacity of 125 Gbps across 14 slots. This allows massive GbE and 10 GbE port density today plus enough scalability to double that density tomorrow — without a forklift upgrade. Dell Force10's ASIC-based distributed forwarding provides true line-rate performance for all packet sizes, even with all L2/L3/MPLS features enabled. The multiprocessor control plane, fortified with DoS protection, supplies scalability and resilience to large networks. In addition, FTOS's process modularity and isolation allow control plane stability with extensive features.

From virtualized data center and Cluster/Grid Computing to next-generation Internet Exchanges and cloud computing networks, the breakthrough ExaScale E-Series architecture supports the most demanding network applications and enables tomorrow's ultra high-speed network applications. The Dell Force10 ExaScale E-Series provide an unprecedented combination of scalability, performance, and full-featured L2 switching, L3 routing and MPLS.