Cisco Data Center Ethernet

Q. What is Data Center Ethernet? Is it a product, a protocol, or a solution?

A. The Cisco® Data Center Ethernet architecture is a collection of Ethernet extensions providing enhancements to classical Ethernet targeted for specific use in the data center. The enhancements, when grouped together, form solutions for scenarios encountered in the data center. The Cisco Data Center Ethernet architecture has three components: lossless fabric, Layer 2 multipathing (L2MP), and extensions to Ethernet.

Q. What is required to be a Data Center Ethernet switch or endpoint?

A. A Data Center Ethernet switch must have L2MP, lossless transport capability and the following Data Center Ethernet extensions: Priority-based Flow Control (PFC), Enhanced Transmission Selection (ETS), and Data Center Bridging Exchange (DCBX) Protocol. A Data Center Ethernet endpoint must have the PFC, ETS, and DCBX Data Center Ethernet extensions.

Support for congestion notification and Fibre Channel over Ethernet (FCoE) is optional for Data Center Ethernet switches and endpoints. There will be Data Center Ethernet switches and endpoints that do not deploy the congestion notification and FCoE options but will still benefit from the other Data Center Ethernet requirements.

Q. What are the Data Center Ethernet protocol extensions to Ethernet?

A. The extensions are listed here. The first three (PFC, ETS, and DCBX) are required; congestion notification is optional.

- **Priority-based Flow Control (required).** PFC is an enhancement to the existing pause mechanism in Ethernet. The current Ethernet pause option stops all traffic on a link; essentially, it is a link pause for the entire link. PFC creates eight separate virtual links on the physical link and allows any of these links to be paused and restarted independently, enabling the network to create a no-drop class of service for an individual virtual link. It also allows differentiated quality-of-service (QoS) policies for the eight unique virtual links. (IEEE 802.1Qbb; see http://www.ieee802.org/1/files/public/docs2007/new-cm-barrass-pause-proposal.pdf.)
- **Enhanced Transmission Selection (required).** ETS is also called priority grouping. Eight distinct virtual link types can be created by implementing PFC, and it can be advantageous to have different traffic classes defined within the different PFC traffic types. ETS enables these differentiated treatments within the same priority classes of PFC. ETS provides prioritized processing based on bandwidth allocation, low latency, or best effort, resulting in per-group traffic class allocation. For example, an Ethernet class of traffic may have a high-priority designation and a best effort within that same class. ETS allows differentiation between traffic of the same priority class, thus creating priority groups. (IEEE 802.1Qaz; see http://www.ieee802.org/1/pages/802.1az.html.)
- **Data Center Bridging Exchange Protocol (required).** DCBX is a discovery and capability exchange protocol that is used by devices enabled for Data Center Ethernet to exchange configuration information. The following parameters of the Data Center Ethernet features can be exchanged:

- Priority groups in ETS
- PFC
- Congestion notification
- Applications
- Logical link down
- Network interface virtualization
  (See http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overview-rev0.2.pdf.)

- **Congestion Notification (optional).** Congestion notification is a form of traffic management that shapes traffic as close to the edge of the network as possible to minimize the negative effects of congestion and enforce rate limiting at the source. A congestion notification management mechanism uses notification and probing to send messages back to the originating source to slow down transmission when congestion is encountered along the network route. The architecture defines a congestion point and a reaction point. In this model, congestion is measured at the congestion point, and rate-limiting, or back pressure, is imposed on the reaction point to shape traffic and reduce the effects of congestion. (IEEE 802.1Qau; see http://www.ieee802.org/1/pages/802.1au.html.)

Q. What is the status of the Data Center Ethernet extensions with the standards bodies?

A. Since they are all different specifications, they are all at different stages of the IEEE standardization process. They have all been proposed as standards and will have different lifecycles and timelines in the standards working groups.

Q. Why are there different names for the collection of these Ethernet enhancements (DCE™, CEE, and DCB) and how do the terms differ?

A. Different organizations created different names to identify the collection of the specifications, all based on the same core specifications. IEEE has used the term “Data Center Bridging (DCB).” IEEE typically calls a standard specification by a number: for example, IEEE 802.1az. IEEE did not have a way to identify the group of specifications with a standard number, so the organization grouped the specifications into DCB.

The term “Converged Enhanced Ethernet (CEE)” was coined by IBM, again to reflect the core group of specifications, to gain consensus among industry vendors (including Cisco) as to what a Version 0 list of the specifications would be before they all become standards.

Cisco uses the term “Data Center Ethernet” to refer to its architecture for a next generation of Ethernet for the data center. Cisco Data Center Ethernet is a superset of the CEE and DCB proposals, using the same three specifications: PFC, ETS, and DCBX. In addition, Cisco includes L2MP, lossless fabric, and congestion notification. The DCE™ designation will be used when a Cisco product or Cisco partner product includes those specifications and meets the requirements as defined by the Cisco Data Center Ethernet architecture.

Q. Why are Data Center Ethernet and lossless Ethernet important?

A. To transport any application that requires lossless service, such as Fibre Channel storage traffic, over an Ethernet network and achieve a unified fabric, a method for providing a lossless service class is required. Fibre Channel storage traffic requires a no-drop capability. A no-drop traffic type can be created using Data Center Ethernet and a lossless Ethernet switch fabric.

Q. What is Layer 2 Multipathing (L2MP) and why is it important?

A. L2MP is an enhancement to increase the bisectional bandwidth between nodes in the data center by enabling multiple parallel paths between nodes. To do this, Ethernet must adopt alternatives to overcome the limitations of the Spanning Tree Protocol, which blocks all but one path to avoid loops. Although numerous advancements have improved Spanning Tree Protocol, one area where Spanning Tree Protocol is still lagging is the capability to load balance traffic among alternative paths. Data Center Ethernet with L2MP will enable use of all available connections between nodes and eliminate the single-path requirement and slow convergence of Spanning Tree Protocol. Multiple options are being proposed in the standards to achieve multipathing at Layer 2.

Q. Is lossless Ethernet synonymous with Data Center Ethernet, or is something else required?

A. A lossless Ethernet environment has two fundamental requirements: a method to pause the link, such as PFC, found in Data Center Ethernet; and a method to tie the pause commands from the ingress to the egress port across the internal switch fabric. The pause option in Ethernet and PFC in Data Center Ethernet take care of providing lossless Ethernet on each link. Additionally, a lossless intraswitch fabric architecture is required to provide truly lossless Ethernet. A lossless intraswitch fabric is achieved by logically connecting the PFC pause of virtual links at the ingress port, with the virtual output queues at the egress port of the switch, so that packets are not switched from the ingress to the egress port when the port is full. This resource arbitration of the switch fabric eliminates packet loss across the switch fabric when output port resources are not available. When Data Center Ethernet and intraswitch resource arbitration are combined, they fulfill the two requirements for a lossless Ethernet fabric.

Q. What is FCoE?

A. Fibre Channel over Ethernet (FCoE) is a protocol specification that maps native Fibre Channel over Ethernet, independent of the native Ethernet forwarding scheme. It allows an evolutionary approach to I/O consolidation by preserving all Fibre Channel constructs. It is not a requirement for deploying Data Center Ethernet, but rather an option that takes advantage of a fabric that is enabled for Data Center Ethernet.

Q. Are Data Center Ethernet and lossless Ethernet required to support FCoE?

A. INCITS T11 is writing the standard for FCoE that will mandate that a lossless Ethernet network is required to support FCoE. Standard pause (as well as PFC) makes a link lossless, but pause or PFC alone are not enough to make a network lossless. To make the network lossless, each switch needs to correlate the buffers of the incoming links to the buffers of the egress links and tie them to the pause implementation. This is a capability of the architecture and has nothing to do with protocols. The Cisco Nexus 7000 Series with DCE™ modules will provide this lossless network capability.

Q. Why do we need Data Center Ethernet? Isn't Ethernet good enough?

A. Data Center Ethernet is adding new capabilities specifically for Ethernet in the data center to the existing strong base of Ethernet. With these new data center extensions, Ethernet will provide solutions for consolidating I/O and carrying multiple protocols, such as IP and Fibre Channel, over Ethernet on the same network fabric, as opposed to separate networks. Also, the continuing spread of 10 Gigabit Ethernet provides the opportunity for greater mixtures of traffic types. With the enhancements in Data Center Ethernet, (PFC, ETS, DCBX, and congestion notification) enabled, a 10 Gigabit Ethernet connection will be able to support multiple traffic types simultaneously, while preserving the respective traffic treatments. With

these extensions, the same 10 Gigabit Ethernet link will also be able to support Fibre Channel storage traffic by offering a no-drop capability for FCoE traffic.

Q. Besides the obvious benefits of reduced number of cables, potential power reduction, reduced number of interfaces, and increased storage utilization enabled by storage consolidation, what other advantages does consolidated I/O provide?

A. A consolidated I/O link, or a unified I/O cable, can present multi-protocol traffic to a unified fabric on a single cable. A unified fabric is a single, multipurpose data center transport that can transmit IP traffic and Fibre Channel storage data simultaneously across the same interface and the same switch fabric, preserving differentiated classes of service. Use cases include multi-protocol transport, FCoE, and Remote Direct Memory Access (RDMA) over low-latency Ethernet. Other benefits include the capability of all hosts to access storage resources over the common unified fabric; any host with FCoE support will be able to mount any storage area network (SAN) resource.

Q. What about classical or traditional Ethernet? Will Data Center Ethernet ports be able to connect to existing Gigabit Ethernet and 10 Gigabit Ethernet ports?

A. Yes, ports enabled for Data Center Ethernet will act as regular Ethernet ports when connecting to other regular Ethernet ports. By definition, backward compatibility is supported.

Q. How will hosts support Data Center Ethernet?

A. Cisco is working with a broad group of partners who will deliver 10 Gigabit Ethernet network interface controllers (NICs) and next-generation converged network adapters (CNAs) that support the Data Center Ethernet features.

Q. What makes a product compatible with Data Center Ethernet?

A. When a product has implemented the required Data Center Ethernet specifications, it should be compatible with other products that have implemented the same set of specifications. The minimal requirements for DCE™ endpoints (hosts, targets, servers, etc.) are PFC, ETS, and DCBX. A DCE™ switch must also include a lossless intra-switch fabric architecture capable of providing a no-drop service and L2MP.

A DCE™ switch must provide a lossless fabric for a lossless transmission service class that will not drop a frame. To support FCoE, a lossless fabric is mandatory to essentially guarantee storage traffic a no-drop service. To create a lossless Ethernet fabric with multi-protocol support, two elements are required: a priority-based pause mechanism (such as PFC), and an intelligent switch fabric arbitration that ties ingress port traffic to egress port resources to honor any pause requirements.

Today's standard pause in Ethernet halts all traffic types on the link. PFC creates up to eight separate logical links over the same physical link, thus allowing any one of eight traffic types to be paused independently while allowing the other traffic types to flow without interruption. PFC uses a priority-based pause mechanism to select the IEEE 802.1p traffic type to be paused on a physical link. The capability to invoke pause on differentiated traffic types enables the traffic to be consolidated over a single interface, resulting in a single Ethernet connection for unified I/O links. Single unified I/O connections aggregate multiple traffic types for delivery to a unified fabric.

PFC provides a no-drop option for each logical link with the capability to halt independent logical traffic types. PFC (as well as standard pause) makes a link lossless, but that is not enough to make a network a lossless fabric. In addition to no-drop capability on the link, a method to tie ingress port pause behavior to egress port resources is required across the intra-switch fabric. To make the network lossless, each switch also needs to relate the buffers of the incoming links to the buffers of the egress links. By logically tying the availability of egress port resources to the ingress port traffic, arbitration can occur to help ensure that no packets are dropped—this is the definition of a lossless, switch fabric architecture. This lossless Ethernet intra-switch fabric behavior provides the required no-drop service that emulates the buffer credit management system seen in Fibre Channel switches today.

Q. Will all Cisco products support Data Center Ethernet?

A. It is unlikely that all Cisco products will implement Data Center Ethernet, since many Cisco products are not built just for the data center. Data Center Ethernet is designed primarily to enhance data center networks, although there will be benefits in other environments. Most of the enhancements will be offered in Cisco products that are found in the data center. Since there are numerous components to Data Center Ethernet, it is not unlikely that a portion of the enhancements will be added to certain platforms. For example, a Cisco switch might support I/O consolidation but not implement congestion notification. In this case, the switch would be considered compatible with Data Center Ethernet without the optional congestion notification.

Q. Does the Cisco Nexus 7000 Series Switch hardware support Data Center Ethernet today? Is a software or hardware upgrade required?

A. The Cisco Nexus 7000 Series will support Data Center Ethernet with new line card modules and an upgrade to the Cisco NX-OS Software. The supervisors, fabric modules, and common equipment are ready for Data Center Ethernet. Most current Ethernet switches will require new hardware and upgrades to software since several of the extensions are enabled in hardware.

Q. The Cisco Nexus 7000 Series has a lossless fabric. Does that mean that it supports Data Center Ethernet already?

A. The fabric of the Cisco Nexus 7000 Series is lossless, and it can support Data Center Ethernet today.

Q. If the Cisco Nexus 7000 Series has a lossless fabric, can the first-generation line cards support Data Center Ethernet and support lossless transmission?

A. The first-generation line cards are classical Ethernet and do not support the Data Center Ethernet extensions, and they can incur some packet loss today.

Q. Do the Cisco Nexus 5000 Series Switches support Data Center Ethernet?

A. Yes. The Cisco Nexus 5000 Series Switches are Cisco DCE™ switches.

Q. Where can I find more information on Data Center Ethernet?

A. Visit the following URLs:

- Cisco Data Center Ethernet: http://www.cisco.com/go/dce
- DCBX explanation: http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overview-rev0.2.pdf
- Data Center Bridging Tutorial: http://ieee802.org/802_tutorials/nov07/Data-Center-Bridging-Tutorial-Nov-2007-v2.pdf

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C67-461717-00 03/08