探索 Dell PowerEdge™ M1000e 网络结构架构

作者：JOHN LOFFINK

模块化组件是赶上 IT 高速发展的关键。 新 Dell™ PowerEdge™ M1000e 模块化刀片服务器机箱的架构及其第 10 代戴尔服务器技术的设计用途是同时支持现在和将来的网络技术，保护企业的模块化服务器投资。

相关类别：

刀片服务器

Dell PowerEdge 刀片服务器

Dell PowerEdge 服务器

有关完整目录索引，请访问 DELL.COM/PowerSolutions。

PowerEdge M1000e 模块化刀片服务器机箱的设计完全是为了支持现在和将来的服务器、存储、网络和管理技术，其中为多种连接留下了余地。作为模块化架构的一部分，关键的 PowerEdge M1000e 组件一个标准的千兆以太网（GbE）结构和两个额外的可自定义的结构，提供充分的带宽，支持 10 千兆以太网（10GbE）、结构通道和 InfiniBand 的互连。 机箱全部使用冗余的热插拔组件，以便最大程度地实现系统可用性。

PowerEdge M1000e 机箱支持多达 16 个半高服务器模块，每个占用机箱前面的一个插槽。本文介绍使用这些半高刀片服务器的 PowerEdge M1000e 结构架构。 较大机箱中的刀片服务器可以相应地支持更高级别的结构连接[1]。

模块化服务器 I/O

为了了解 PowerEdge M1000e 的架构，就必须首先定义四个关键术语： 结构、通路、链接和端口。

结构是在多个设备之间编码、传输和同步数据的一种方法。例如，GbE、结构通道或者 InfiniBand 等。结构在 PowerEdge M1000e 系统内的服务器模块和 IOM 之间通过中平面进行传输。它们也可以通过物理铜线或者 IOM 上的光接口传输到外部空间。

通路的定义是 I/O 终端设备之间的结构数据传送路径。 在现代的高速串行接口中，每个通路包括一个发送差异对和一个接收差异对。 实际上，一个通路是印刷电路板上电源线或者铜线轨迹中的四条线组成的： 包括发送正信号、发送负信号、接收正信号和接收负信号。 发送差异对信号位这些高速通路提供了更好的噪声容限。 在提到通路时，结构标准中使用了不同的术语。 PCI Express（PCIe）用的是通路，InfiniBand 用的是物理通路，而结构通道与以太网用的是链接。

在这里，链接的定义是在 I/O 终端设备之间组成一个通信输送路径的多个结构通路的集合。例如，四通路（x4）、八通路（x8）、十六通路 (x16)PCIe 以及四通路 10GBase-KX4。PCIe、InfiniBand 和以太网使用的术语是链接。使用通路

[1] 有关 PowerEdge M1000e 的更多信息，请参阅 Chad Fenner 发表在 Dell Power Solutions（2008 年 2 月期）上的文章 “ 新一代的 Dell PowerEdge M1000e 模块化刀片服务器机箱 ”，下载网址是：DELL.COM/download/Global/Power/ps1q08 - 20080206 - Fenner.pdf。

> “根据设计，关键 PowerEdge M1000e 组件可在每个刀片服务器上支持三个高速光纤：即一个标准千兆以太网光纤和两个额外的可自定义光纤。”

与链接两个不同术语的原因是为了防止出现混乱，因为在以太网中，链接同时指单通路结构传输与多通路结构传输。一些结构（如结构通道）没有定义链接，因为它们的单个传输运行的是多通路，旨在提高带宽。 这里定义的链接可在多通路之间提供同步，从而可以有效地协作提供单个传输。

端口的定义是所链接设备的物理 I/O 端接口。 一个端口可以连接一个或者多个通路的结构 I/O。

刀片服务器 I/O 架构

每个 PowerEdge M1000e 半高服务器模块有三种受到支持的高速结构，其中通过是使用服务器上可选插件夹层卡的、灵活的结构。 根据设计，PowerEdge M1000e 机箱中使用的刀片服务器可支持多个网络布局，并为未来的升级提供足够的带宽。 I/O 结构整合包含(LAN)、存储区域网络（SAN）和进程通信（Interprocess Communication，IPC）网络。 刀片服务器端口通过机箱中平面与机箱后部的相关 IOM 连接，然后，该机箱再连接到 LAN、SAN 或者 IPC 网络。

如图 1 所示，第一个嵌入的高速结构是结构 A，它由两个 GbE LOM 以及相关的机箱 IOM 组成。 LOM 基于 Broadcom BCM5708 NetXtreme II 以太网控制器，并支持 TCP/IP 卸载引擎(TCP Offload Engine, TOE）和 Internet SCSI (iSCSI) 启动。 尽管刀片服务器目前已经拥有一个双 LOM 配置，机箱中平面设计将来仍然可在每个半高刀片服务器上支持最多四个 GbE LOM。 刀片服务器 LOM 和以太网结构 B 和 C 夹层卡也支持局域网唤醒(Wake-on-LAN，WOL)。 基于 iSCSI 和结构通道的卡也支持 SAN 启动。

直通模块对信号进行电子缓冲处理，并执行一些低级别的链接层转换；它不会接触数据，而且，所有的数据流都与夹层卡中的数据保持一对一的通信。

除了结构 A 之外，刀片服务器在每个半高刀片服务器上安装两个可选的双端口 I/O 夹层卡，为额外的结构 B 和 C 提供支持。 这些卡目前支持各种以太网（包括 iSCSI）、结构通道和 InfiniBand 技术，并使用常规机箱提供结构配置灵活性。 结构 B 和 C 的 I/O 夹层卡使用的是相同的机械、电气和管理技术规范，具有很高的灵活性和模块度。

可选的夹层卡通过八通路（x8）PCIe 接口与刀片服务器芯片组连接，并通过 Gen1 PCIe 提供多达 16 千兆位/每秒/每夹层卡的带宽。 PCIe 结构和外部结构都是通过高速 10 千兆位/每秒空气介质连接器插脚、平面和中平面进行路由的。 为了提高信号完整性，信号会隔

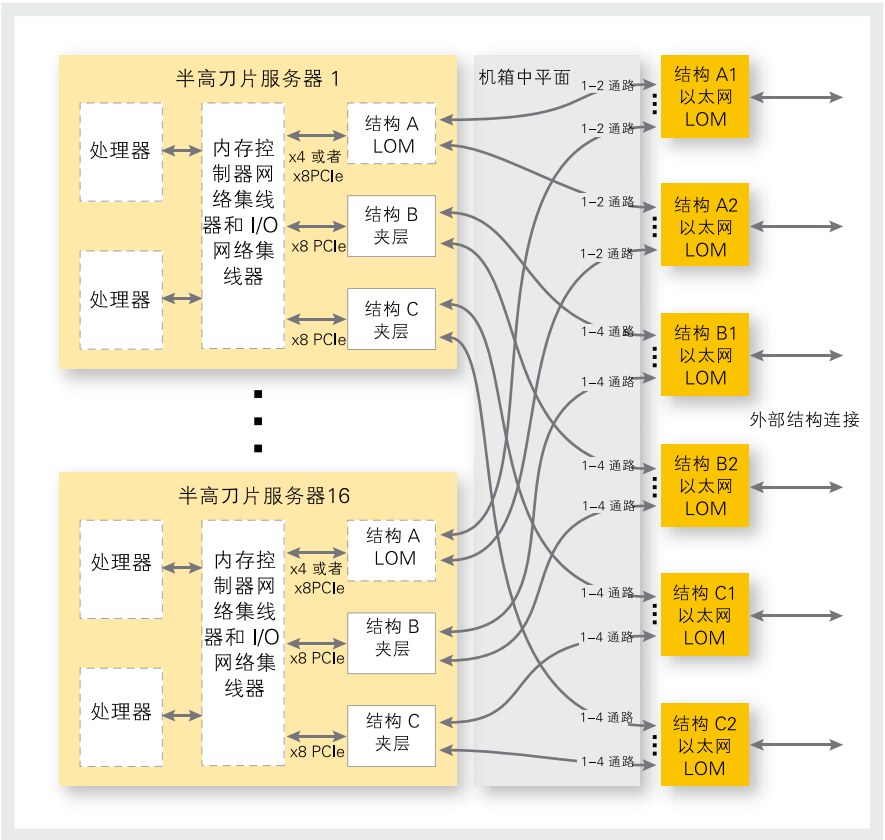

图 1. Dell PowerEdge M1000e 高速结构架构

离发送与接收信号，将串话干扰降到最低程度。 差异对是通过地面插脚隔离的，而信号连接器列是交错排列——有助于将信号耦合降到最低程度。

因为采用了模块化架构，且可以灵活地配置系统中的夹层和 IOM，所以，管理员会无意地使用错误的夹层卡将刀片服务器热插拔到系统中。 为了防止意外激活刀片服务器上错误配置的结构，PowerEdge M1000e 系统管理硬件和软件中包括了结构一致性检验。 例如，如果管理员在结构 C 插槽中配置了结构通道 IOM，那么，所有的刀片服务器都必须有结构通道夹层卡或者，结构插槽中不使用夹层卡。 在这种情况下，如果将刀片服务器热插拔到机箱中，而机箱的结构 C 插槽中有 GbE 夹层卡，那么，系统会自动地检测这一错误配置，并向管理员发送错误警报，以便适当地重新配置刀片服务器。

机箱 I/O 架构

PowerEdge M1000e 中推出的一个主要功能是，在使用以太网直通模块时，全面支持每秒 10/100/1,000 兆比特以太网速度。以前，直通连接仅限于每秒 1,000 兆比特或者 GbE 速度。 PowerEdge M1000e 可让机构使用以太网直通或交换机技术连接到旧式 10/100 兆比特每秒基础结构。 这一特性使用的是 1000Base-KX 传输上的带内信号，不需要管理员的交互。

机箱中平面可以对结构 A 提供多达四个 GbE 链接/刀片服务器的支持，每个半高刀片服务器提供多达 4 千兆位/每秒的带宽。 结构 B 和 C 以刀片服务器夹层卡中的两组四个通路路由到机箱后部的 IOM。 支持带宽范围从每通路 1 千兆位/每秒到 10 千兆位/每秒，具体情况取决于结构，或者每夹层卡多达 80 千兆位/每秒。

因为每个夹层卡都通过八通路 PCIe 链接连接到刀片服务器，所以，系统没有限制 I/O 带宽的节流点。 根据 Dell 的预计，当多通路 10GBase-KR 技术可用时，刀片服务器将迁移到 PCIe 2.0 或者更好的设备上，在刀片服务器和机箱 IOM 之间提供完全的端到端 I/O 带宽。

中平面

中平面——提供电源分配、结构连接和系统管理基础结构的大型印刷电路板——是 PowerEdge M1000e 机箱内所有连接的重点；它旨在为现在和将来的服务器和基础结构提供可扩展的带宽。 根据容错系统的要求，PowerEdge M1000e 中平面是被动式的，没有隐藏的堆栈中平面或者带有活动组件的内插器。 I/O 结构和系统管理基础结构的设计用途是为每个热插拔组件提供全部冗余。

I/O 结构通过支持 10 千兆位/每秒的高速连接器和绝缘材料进行路由。 I/O 渠道模拟的是 10GBase-KR 渠道模型。该结构符合各种行业标准,支持 10-12 或者更高的位错误率。Dell 进行大量的投资，确保位现在和将来的服务器与基础结构提供可扩展的带宽。

I/O 模块

PowerEdge M1000e 机箱背后配备了系统管理、冷却、电源和 I/O 组件。IOM 都是成对使用的，其中两个全冗余的模块供每个刀片服务器结构使用，可以是直通或交换机模块。直通模块在每个刀片服务器上的夹层卡端口和外部网络之间提供一对一的直接连接。交换机以有效地汇总刀片服务器上的 LOM 或者夹层卡中的链接，并将其连接到网络上行链路中。 根据设计，插槽之间的 IOM 是完全兼容的。 图 2 介绍了这些不同的组件，它们是 PowerEdge M1000e 架构的组成部分。

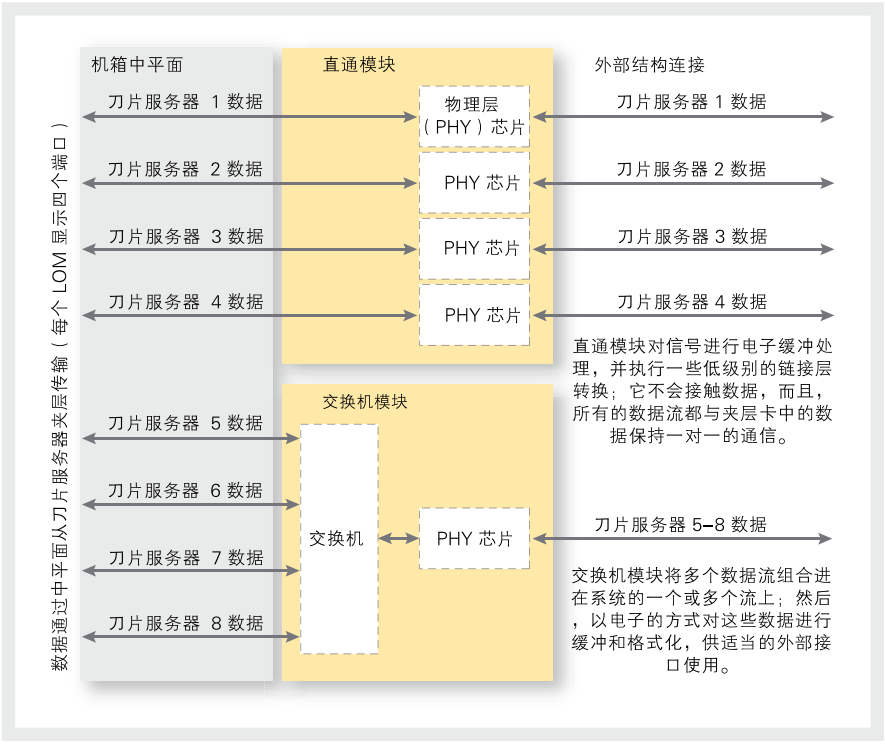

图2. Dell PowerEdge M1000e 架构的直通模块与交换机模块

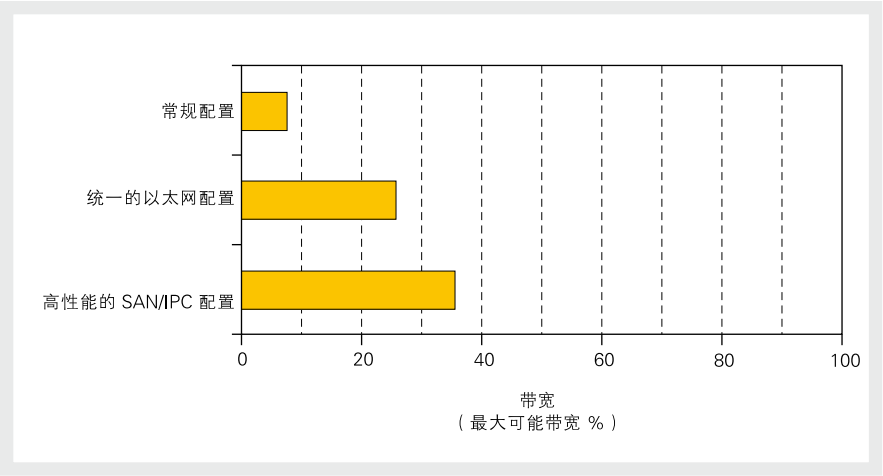

图 3. 三种 Dell PowerEdge M1000e 配置的带宽要求（占总系统带宽的百分比）

对于 PowerEdge M1000e 机箱而言，Dell PowerConnect™ M6220 以太网交换机与 Cisco Catalyst 刀片服务器交换机（3032、3130S-G 和 3130S-X）表明了 IOM 的先进模化程度。 这些以太网交换机模块中推出子模块 I/O 扩展，有助于最大程度地实现灵活地与扩展。 它们提供统一的硬件设计，支持从低成本的 GbE 配置扩展到利用交换机堆栈将机箱内或者之间的多个交换机互连的配置，或者扩展到支持使用铜质接口与光接口将 10GbE 上行链路连接到核心网络的配置。

可扩展的系统带宽

如果所有三种结构中配备全部的 GbE 与 10GbE 通路，PowerEdge M1000e 架构总共可以支持 5.44 Tbps 的带宽。 如图 3 所示，每个刀片服务器中采用四通路 GbE 与两通路结构通道的常规配置使用的带宽不到总带宽的 10%。 一个使用两通路 GbE 与四通路 10GbE/刀片服务器的、汇总了所有网络、存储和 IPC-over-Ethernet 链接的以太网配置使用的带宽大约只有总带宽的 25%。 即使是使用两通路 GbE、两通路 8 千兆位/每秒结构通道的高性能 SAN/IPC 配置，两通路的 InfiniBand 使用的带宽仍然只有总带宽的大约 35%。

按照设计，PowerEdge M1000e 可以灵活地、低成本地全面支持短期的、中期的、长期的 I/O 基础结构要求。 例如，可以灵活地路由结构 A 的两个双通路路径以及结构 B 与 C 的两个四通路路径。短期来看，10GBase-KX4 路由支持 10GbE 连接。 10GBase-KX4 路由使用的是所有 4 个通路，每个通路运行的数据速率是 2.5 千兆位/每秒，总链接数据带宽为 10 千兆位/每秒。 在今天和不远的将来，10GBase-KX4 是成本最低的、无所不在的中平面 10GbE 结构解决方案。 全面配置的 PowerEdge M1000e 系统支持两个双 GbE 结构和两个配有 10GBase-KX4 路由的双 10GbE 结构，为尖端的统一网络布局提供支持。 例如，此类配置可为传统的网络通信量的每刀片服务器提供冗余的 10GbE 链接，为 iSCSI 或者以太网上光纤通道（Fibre Channel Over Ethernet）连接的存储提供另一系列冗余的 10GbE 链接，同时，继续维护两个双通路的 GbE 链接，以便满足系统管理或者其他窄带要求。

模块化刀片服务器架构

根据设计，Dell PowerEdge M1000e 模块化刀片服务器机箱的架构及其第 10 代戴尔服务器技术专门着眼于模块度，用于提供可自定义的多通路结构，为现在和将来的网络技术提供支持，同时，保护企业的刀片服务器投资。 因为 Dell 的部分侧重点是简化 IT，因此，对于 PowerEdge M1000e 夹层或者 IOM、多种夹层和 LOM 机箱以及设计标准而言，它们的产品支持列表不会出现混乱，从而有助于避免潜在的兼容性和配置困难。 PowerEdge M1000e 支持夹层设计标准和设计标准，可在系统级别和子系统级别支持真正的模化程度，从而可以简化现在和将来的扩展与增强。

John Loffink 是戴尔服务器高级工程事业部的工程师/战略家。 在服务器、企业存储、硬件设计和高可用性计算方面拥有超过 20 年的背景。 John 荣获了佛罗里达州工学院的电子工程学士学位。

MORE ONLINE
www.dell.com.cn/PowerSolutions

快速链接

Dell PowerEdge M1000e:
www.dell.com.cn/Servers