**VCF纵向虚拟化技术架构**

文/刘刀桂 祁正林

VM及其迁移驱动着数据中心大规模二层网络的发展，随着网络规模的扩大，网络设备数量随之增大，网络管理成为数据中心基础设施管理中的一个棘手问题。同时，现代大数据中心对网络提供给服务器的端口密度也提出了更高的要求，例如万台服务器的规模已是互联网数据中心现实中的普遍需求。端口扩展技术作为提高接入设备端口密度的一种有效手段逐渐成熟并获得了业界的认可。VCF纵向虚拟化技术（Vertical Converged Framework，纵向融合框架，以下简称VCF）即是该技术的一种实现方式，满足数据中心虚拟化高密接入并可以简化管理。Cisco公司相类似的技术是FEX。

VCF在纵向维度上支持对系统进行异构扩展，即在形成一台逻辑虚拟设备的基础上，把一台盒式设备作为一块远程接口板加入主设备系统，以达到扩展I/O端口能力和进行集中控制管理的目的。为叙述方便，后文会把纵向VCF的建立和管理过程等与IRF传统的横向相关功能进行对比。

IRF（横向）堆叠拓扑主要有链型和环形两种。设备按角色可分为Master和Slave。Slave在一定条件下可转变为Master，两者业务处理能力是同一水平的，只不过Slave处于“非不能也，实不为也”的状态。

对于VCF（即纵向）来说，设备按角色分为CB（Controlling Bridge）和PE（Port Extender）两种。CB表示控制设备，PE表示纵向扩展设备，即端口扩展器（或称远程接口板）。通常来说，PE设备的能力不足以充当CB，管理拓扑上难以越级，处于“非不为也，实不能也”的状态。

如图1所示，左边是框式设备或者是盒式设备各自形成IRF堆叠横向虚拟化系统，有环形堆叠和链型堆叠（虚线存在的情况）两种拓扑形式；右边是框式设备与盒式设备形成VCF纵向虚拟化系统（简称 VCF Fabric），为便于对比CB由IRF堆叠组成。

1. IRF横向虚拟化和VCF纵向虚拟化对比

一般来说，对于IRF（横向）堆叠，控制平面由Master管理，转发能力和端口密度随着Slave增加而增加。对于VCF（纵向） Fabric，控制平面由CB（或IRF中的Master）管理，端口密度随着PE增加而增加，但总体上转发能力仍取决于CB设备。

VCF可与IRF技术组合使用， 所形成的系统具有单一管理点、跨设备聚合以及即插即用等优点，同时加强了纵向端口扩展能力。

# VCF技术机制

对VCF来说，CB角色可以由处理能力较强的盒式设备和框式设备承担，也可以是基于IRF技术建立的横向堆叠。PE一般来说是低成本的盒式设备。实际应用中，CB角色多为横向堆叠，这样有益于PE上行冗余。以下技术说明以此为主。

1. **拓扑管理**

图2中CB角色是一个典型的IRF堆叠。PE角色为盒式设备。CB与PE互联口称为纵向 Fabric口，纵向 Fabric口是一个逻辑概念，可以是一个物理端口或者多个物理端口组成的聚合口。CB与PE之间可以使用专用线缆或光纤连接。



1. VCF典型拓扑

PE根据组网需要可以连到一台或多台CB设备上，PE与PE之间不能再有其他连线。从模型上说，PE相当于CB的一块远程接口板。从功能上看，CB与PE间的纵向 Fabric连接相当于框式设备的“背板”。从管理上看，所有CB和PE设备组成一个堆叠，对外是一台设备，一个管理点。

整个拓扑建立包括两个方面：一方面是多台CB设备依据IRF相关规则和拓扑计算建立横向堆叠；另一方面是CB通过纵向 Fabric口向外发送HELLO报文，根据PE反馈信息建立纵向 Fabric。

如图3所示，纵向 Fabric建立过程主要分为四步：

* 第一步，完成扩展板编号（Slot-ID）的分配和获取。CB上VCF Fabric口使能后会周期性地发送探测报文，一旦Slot-ID分配完成则停止。
* 第二步，完成软件的加载。包括PE发送加载请求，CB提供版本文件描述信息，以及确认加载和加载完成等几个子过程。这其中，Bootware（类似于个人电脑上的BIOS）和App（即主机软件）的加载实现过程类似。
* 第三步，PE以下载后的版本重启并完成在CB的注册。
* 最后，CB向PE下发配置信息。



1. PE加入以及VCF建立过程
2. **VCF Fabric连接方式**

如前所述，PE到CB间纵向 Fabric连接类似于框式设备的“背板”，为了增加带宽并使上下行流量保持合适的收敛比，两者间链路通常由多个物理线路组成，逻辑上可采用HASH方式来实现。这样一条链路Down，不会引起挂服务器的下行端口Down，但带宽变小，相关流量也会重新进行HASH计算并分配到剩余链路上（如图4所示）。



1. VCF Fabric连接方式
2. **PE管理**

横向配置、Master选举以及整个堆叠建立和维护与IRF没有纵向功能前完全一样。纵向VCF加上PE后，建立过程相对复杂一些，但本质上所有CB和PE形成一个单一的逻辑实体，可以通过任何一台CB上的用户管理接口，如Console口、Telnet或者网管口来进行配置和管理。

IRF（横向）系统使用成员编号（Member-ID）来标识和管理成员设备，在一个IRF中所有设备的成员编号是唯一的。成员编号被引入到端口编号中，便于用户配置和识别成员设备上的接口。类似地，在VCF（纵向）中，系统使用扩展板编号（Slot -ID）来标识和管理纵向扩展设备，在整个 系统中扩展板编号也是唯一的且同样被引入到端口编号中。如果CB是框式设备，这个编号也绝对不能与框式设备上已有接口板（LPU）的编号重复。在使用上两者机制稍有区别，成员编号（Member-ID）需要设备重启才能生效；而扩展板编号（Slot -ID）在CB上配置后可立即生效。

* **PE加入**。当VCF系统有新的成员设备加入时，会根据系统所处状态或者PE设备的状态采取不同的处理过程。假设横向IRF已配置，且在CB上已为PE分配了Slot -ID。（1）此时PE以缺省出厂配置可即插即用。正常运行的纵向 VCF系统，当因某些外在因素引起断电或重启，系统不需要干预的情况下将自动恢复。（2）运行过程中，PE可以通过纵向 Fabric口随时接入系统，CB会自动计算拓扑以防止新的PE接入时产生环路。从虚拟化的角度来看，这个过程相当于框式设备的接口板插入。当然，由于此时的“背板”链接是动态端口，需要进行拓扑计算以阻断环路；而实际的框式设备在初始化时已经完成了这一动作。
* **PE离开**。PE离开相对来说简单一些，当CB与PE链接电缆拔出或者对应端口Down掉，系统即产生远程接口板离开事件。这一过程与框式设备的接口板拔出基本一致。

1. **盒式设备作为CB**

盒式设备充当CB并下挂PE时，横向CB通过IRF互联形成的虚拟设备相当于一台框式分布式设备主控板；纵向PE通过VCF互联形成虚拟框式设备的分布式设备接口板（或称线卡）。横向IRF互联电缆模拟了交换背板中主控板互联，IRF中的Master相当于虚拟设备的主用主控板，Slave设备相当于备用主控板。同样地，纵向VCF的CB与PE间互联电缆模拟交换背板中接口板到背板的链接，PE设备相当于虚拟设备的I/O接口板。如图5所示，右边为虚拟化设备的逻辑视图。



1. 盒式设备作为CB时形成的VCF虚拟化设备
2. **框式设备作为CB**

框式设备充当CB并下挂PE时，对于横向，框式设备通过IRF互联形成的虚拟设备也相当于一台框式分布式设备，此时该虚拟框式设备拥有更多的主控板和接口板；对于纵向，PE通过VCF互联形成虚拟框式设备的分布式设备接口板。横向IRF中的Master的主用主控板相当于虚拟设备的主用主控板，Master的备用主控板以及Slave的主用、备用主控板均相当于虚拟设备的备用主控板（同时可担任接口板的角色）；Master和Slave中的接口板继续担当接口板的角色，其中接口板的部分或者全部端口与PE相连。同样地，对于纵向，盒式PE通过VCF与CB相连（一般来说是框式CB的接口板），PE设备相当于虚拟设备的I/O接口板（如图6所示，右边为虚拟化设备的逻辑视图）。



1. 框式设备作为CB时形成的VCF虚拟化设备

# VCF系统管理

上文已提到整个 Fabric系统可作为一个逻辑实体，通过一个IP进行管理。但是系统层面如何进行软件版本管理，如何进行配置和如何通过即插即用来建立VCF系统的呢？

* **软件版本管理。**IRF在建立横向堆叠的时候会比较版本，最终所有成员都会统一于Master的版本。对VCF来说， PE在加入堆叠时，从CB下载版本；当CB是IRF堆叠时，无论PE是否直接与Master相连，都会从Master获取版本。因此，从结果看，整个堆叠系统版本都会与Master统一。横向堆叠Slave获取的版本与Master自身运行的版本是同步的；纵向 Fabric各PE获取的是适合PE运行的部分。一般来说，CB和PE各自由不同的CPU和交换芯片等构成，因此实际上在CB（或Master）上有两个不同功能用途的软件包，系统启动或运行过程中会自适应各取所需。
* **配置管理。**整个 Fabric系统作为一个逻辑实体进行管理时，可通过IRF成员如Master或Slave的Console等进行配置；一般来说，PE不提供Console等配置口。对于VCF，当在CB上指定与物理端口（或逻辑上的聚合端口）相应的Slot-ID后，且PE已正常加入系统，此时便可通过CB对PE进行配置，例如PE上端口所属VLAN，QoS规则等。系统保存配置后，PE对应配置信息保存在CB上。当系统重启或者更换PE时，PE对应配置信息也从CB上下发，即PE配置可以“继承”。
* **PE即插即用。**PE相当于VCF虚拟化框式设备的一块接口板，实际框式设备通过热插拔来实现即插即用，为了实现类似功能和简化管理，PE通过纵向 Fabric口以及纵向 Fabric连接Up/Down事件感知支持即插即用。PE“插入”虚拟框的过程同图3纵向 Fabric建立过程，此处不再赘述。

# VCF上层控制协议

VCF侧重对CB设备进行I/O端口扩展，除了和端口密切相关的功能外，其他上层协议基本上都在CB上实现。这样做的好处显而易见，PE仅作为接口板插入虚拟框式设备，提高了端口密度而减少了管理网元，且系统控制管理平面上移，有利于对大二层多服务器环境下的集中控制和网络策略管理。其次，对PE的性能规格要求不高，有利于成本控制。

CB在横向IRF组成的Master和多个Salve是1:N备份模型，作为纵向诸多PE的管理控制单元起着冗余备份作用。纵向 Fabric PE作为接口板加入，协议控制平面继承了横向堆叠的实现和优点。例如对三层报文TTL跳数仍只加1；支持跨PE聚合等。

# VCF转发平面实现

一般来说，VCF中CB设备相对PE性能更好，承担VCF的业务数据转发决策；而PE主要承担CB端口扩展器角色。在VCF CB设备中，不管业务流量来自PE设备，还是来自CB设备自身的非纵向 Fabric口，都根据业务报文的目的进行查表转发。

* **VCF单播转发**

VCF上行方向（即从PE到CB），来自PE的UNI口的流量，在扩展设备上不做查表转发处理，而是将原始的业务报文直接重定向到CB设备。CB设备收到业务报文后，从其中提取扩展端口等信息，并基于该信息完成地址学习及业务控制。

下行方向。如果业务报文需要单播到PE的某UNI端口，CB设备在完成业务报文转发决策和必要的报文修改后，通过纵向 Fabric互联口发送到PE设备。PE收到业务报文后，从其中直接提取出端口等完成业务报文的发送。

* **VCF组播转发**

上行方向，即从PE向CB的报文流程同前文单播一致。下行方向，对于需要组播（广播处理过程同组播）的业务报文，CB设备会为每个PE设备拷贝一份业务报文，通过纵向 Fabric互联口发送到PE设备。PE收到这类业务报文后，如果是广播报文，则在对应的VLAN内广播业务报文；如果是组播报文，则按照组播索引查找对应的UNI端口列表复制并发送业务报文。

* **VCF多互联链路选路机制**

一般来说，CB设备和PE之间会配置多个互联链路，而且IRF横向堆叠作为CB设备时，多个互联链路可以分布在不同的IRF成员设备上。

PE到CB设备的上行方向，单播和组播实现方式一致。采用HASH方式使得流量在互联链路上分布更为均匀。

CB设备到PE设备的下行选路，对于单播，采用最短路径原则，即如果CB设备是IRF，且到某PE的互联链路分布在多个IRF成员上，则IRF选择到PE的最短路径。如果单个IRF成员与某PE有多条互联链路，则在该IRF成员上进行聚合HASH选路。这一原则的出发点在于，尽可能减少对IRF本身堆叠链路的带宽占用。对于组播，只选择一个互联链路发送一份组播拷贝，当同一PE下有多个用户时，实际复制工作在PE进行。

当然，为了提升系统的转发性能和减小延迟，某些PE也可提供本地流量转发功能。

# VCF架构的特点

## VCF部署中的多级冗余和高可靠性

VCF支持多种CB和PE设备。就CB来说，包括框式和盒式两种类型，都支持服务器跨PE冗余接入。特别是框式CB，由于CB通过IRF横向堆叠组网，天然支持跨框和跨板聚合，从而为网络冗余设计提供丰富的选择。同时，VCF方案不仅支持虚拟接入层的冗余，而且也支持核心（汇聚）层的冗余，能更全面提升系统级的可靠性。

## VCF技术的高可扩展性

VCF中CB可通过IRF横向堆叠建立，即可由多台设备组成。例如H3C框式高端设备最多可构建4台设备组成的IRF堆叠，这些设备充当CB角色时也是如此。盒式设备则可构建更大规模的IRF和VCF系统。

VCF中CB角色通过不同组合可提供更大的灵活性和更好的扩展性。这一优势能更好地支持企业或云服务运营商根据自身业务发展规模进行IT设施的平滑扩容。

## L2/L3流量线速转发

VCF中承担CB角色的所有设备，包括框式设备和盒式设备均支持二层、三层流量的线速转发。既不需要增加额外的板卡，而且L2/L3流量完全线速。

## PE设备支持双模式和保护用户投资

H3C的PE设备支持两种运行模式，即标准交换模式和PE模式。两者模式可以通过命令行或者网管进行切换。设备出厂缺省设置为标准交换模式；当和支持VCF纵向管理的设备互连且纵向特性开启的情况下，设备可自动感知切换到PE模式，也就是支持即插即用。

双模式特性使得用户可根据自身网络系统建设组网的需要进行选择，在不牺牲纵向设备“即插即用”等简化管理功能的情况下，很好地保护了用户投资。

# 总结

IT基础设施虚拟化既是当前热点，也是今后一段时期的趋势。VCF技术为网络虚拟化以及支持数据中心服务器虚拟化提供了一种思路。VCF纵向扩展技术将有助于构建大规模虚拟化网络和简化管理*。*