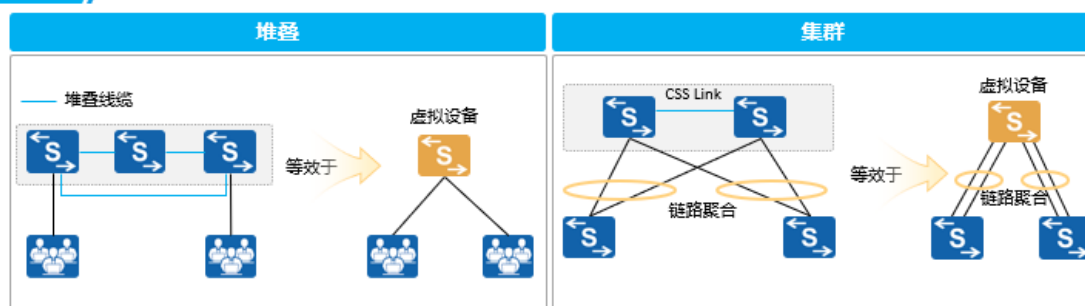


交换机堆叠与集群

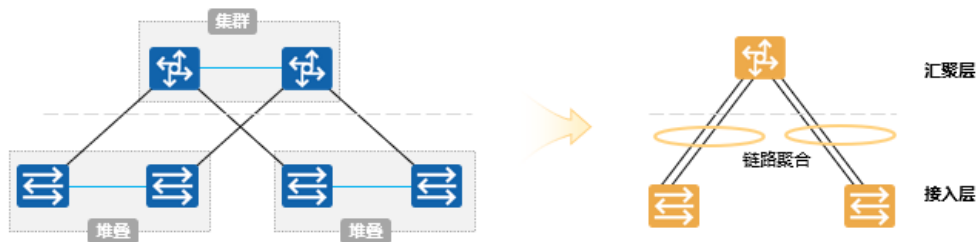
- 随着企业的发展，企业网络的规模越来越大，这对企业网络提出了更高的要求：更高的可靠性、更低的故障恢复时间、设备更加易于管理等。
- 传统的园区网高可靠性技术出现故障时切换时间很难做到毫秒级别、实现可靠性的方案通常为一主一备，存在着严重的资源浪费。同时随着网络设备的越来越多，管理将会变得越加复杂。为构建可靠、易管理、资源利用率高、易于扩展的交换网络，引入了交换机堆叠、集群技术。
- 本章节中将介绍交换机堆叠、集群技术的技术原理、组网方式等。

堆叠、集群简介



- 堆叠 (iStack)，将多台支持堆叠特性的交换机通过堆叠线缆连接在一起，从逻辑上虚拟成一台交换设备，作为一个整体参与数据转发。
- 集群 (Cluster Switch System, CSS)，将两台支持集群特性的交换机设备组合在一起，从逻辑上虚拟成一台交换设备。
- 集群只支持两台设备，一般高端框式交换机支持CSS、盒式设备支持iStack。
- 通过使用堆叠、集群技术结合链路聚合技术可以简单构建高可靠、无环的园区网络。

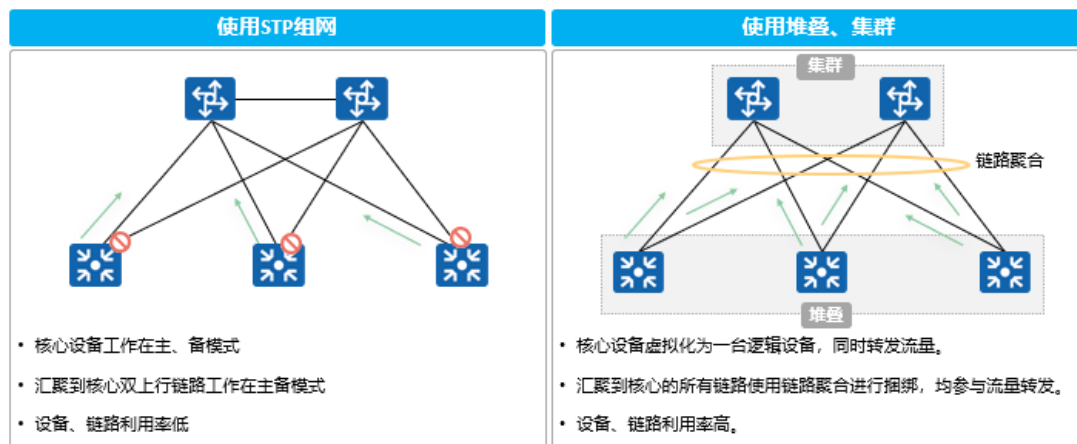
堆叠、集群架构



- 使用堆叠、集群技术将独立的交换机虚拟化成一台逻辑的交换机，一般接入、汇聚层盒式交换机采用堆叠技术，汇聚、核心层交换机采用集群技术。
- 在逻辑交换机之间使用链路聚合技术，无需部署STP、VRRP实现高可靠性。
- 实现高可靠性的同时设备之间的链路可以同时传输流量，链路利用率得以提升。

堆叠、集群的优势 (1)

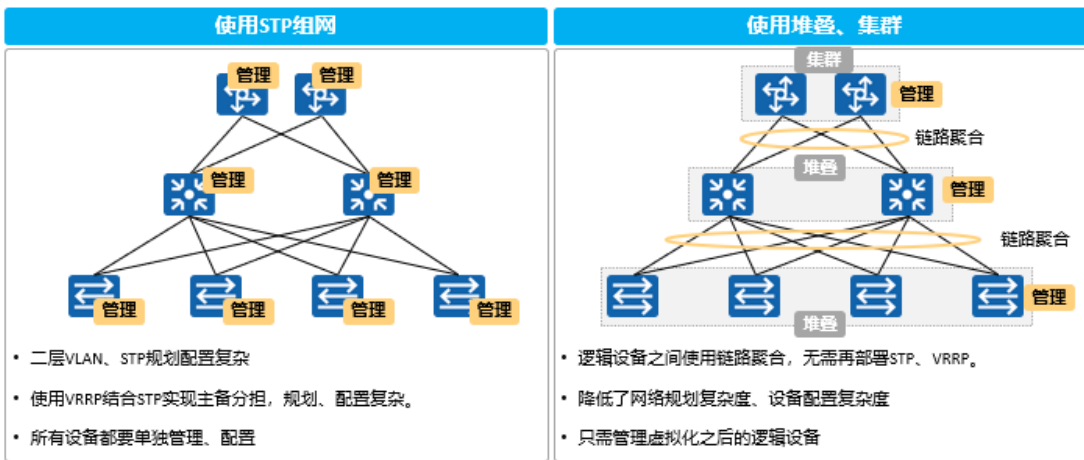
使用堆叠、集群可有效提高资源利用率，获得更高的转发性能、链路带宽。





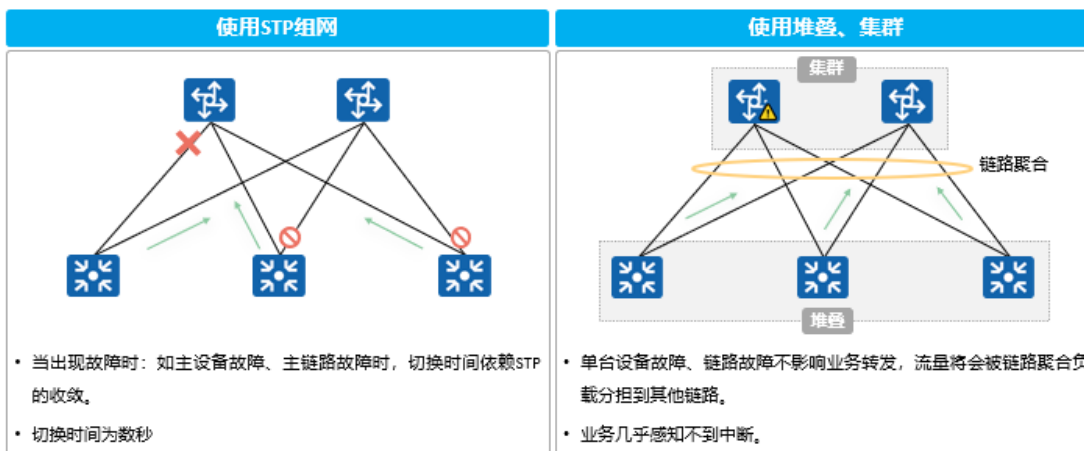
堆叠、集群的优势 (2)

使用堆叠、集群可以降低网络规划的复杂度，方便对于网络的管理。



堆叠、集群的优势 (3)

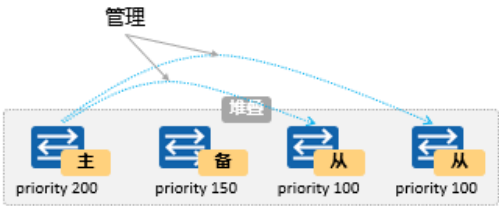
使用堆叠、集群可以大大降低故障导致的业务中断时间。





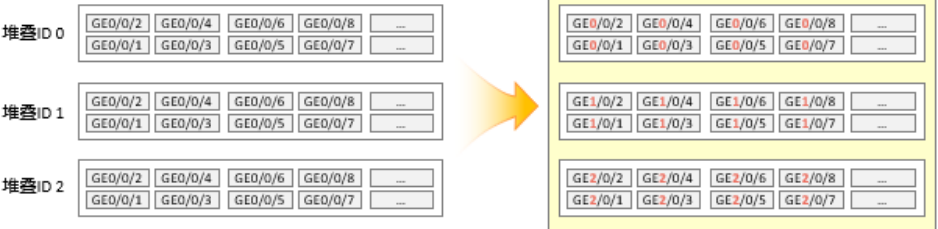
堆叠基本概念

- 堆叠系统中所有的单台交换机都称为成员交换机，按照功能不同，可以分为三种角色：
 - 主交换机（Master）：主交换机负责管理整个堆叠。堆叠系统中只有一台主交换机。
 - 备交换机（Standby）：备交换机是主交换机的备份交换机。堆叠系统中只有一台备交换机。当主交换机故障时，备交换机会接替原主交换机的所有业务。
 - 从交换机（Slave）：从交换机用于业务转发，堆叠系统中可以有多台从交换机。从交换机数量越多，堆叠系统的转发带宽越大。除主交换机和备交换机外，堆叠中其他所有的成员交换机都是从交换机。当备交换机不可用时，从交换机承担备交换机的角色。
- 堆叠优先级：堆叠优先级是成员交换机的一个属性，主要用于角色选举过程中确定成员交换机的角色，优先级值越大表示优先级越高，优先级越高当选为主交换机的可能性越大。



堆叠ID

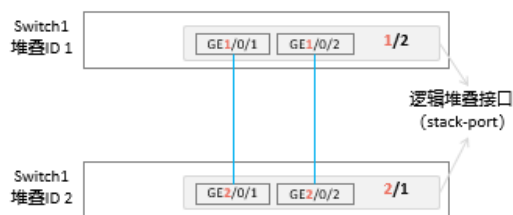
- 堆叠ID，即成员交换机的槽位号（Slot ID），用来标识和管理成员交换机，堆叠中所有成员交换机的堆叠ID都是唯一的。
- 设备堆叠ID缺省为0。堆叠时由堆叠主交换机对设备的堆叠ID进行管理，当堆叠系统有新成员加入时，如果新成员与已有成员堆叠ID冲突，则堆叠主交换机从0 ~ 最大的堆叠ID进行遍历，找到第一个空闲的ID分配给该新成员。
- 在建立堆叠时，建议提前规划好设备的堆叠ID。





堆叠逻辑接口

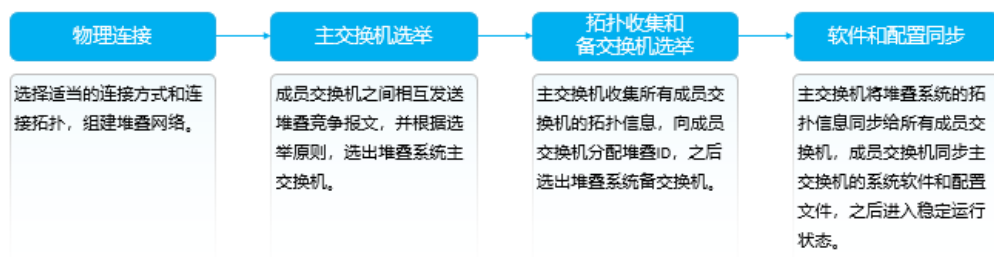
- 堆叠逻辑接口：交换机之间用于建立堆叠的逻辑接口，每台交换机支持两个逻辑堆叠端口，分别为stack-port n/1和stack-port n/2，其中n为成员交换机的堆叠ID。
- 一个逻辑堆叠端口可以绑定多个物理成员端口，用来提高堆叠的可靠性和堆叠带宽。
- 堆叠成员设备之间，本端设备的逻辑堆叠端口stack-port n/1必须与对端设备的逻辑堆叠端口stack-port m/2相连。



- 如果两端设备对应的逻辑堆叠端口（本端的 stack-port n/1 与对端的 stack-port m/2）内包含多个物理成员端口，对物理成员端口的连接无对应端口号的要求。



堆叠系统组建过程

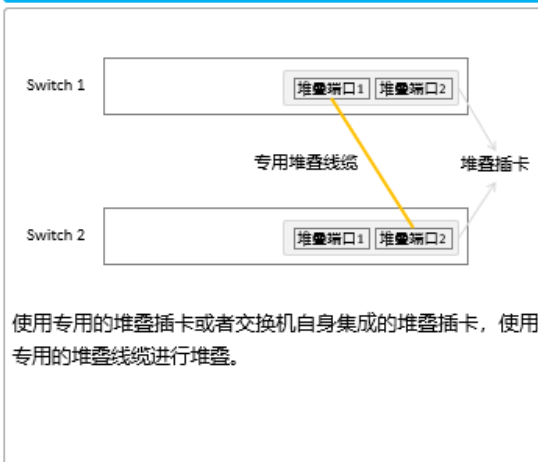




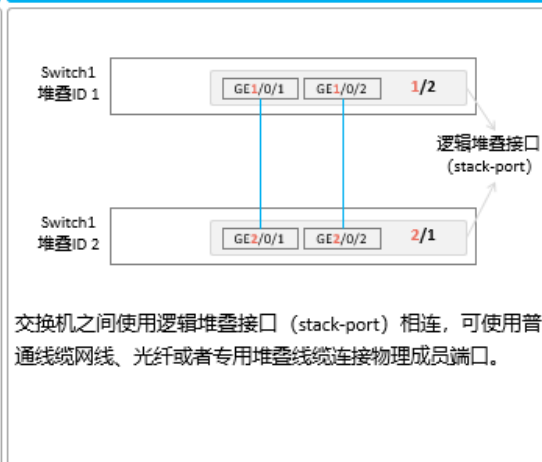
堆叠方式

物理连接 主交换机选举 拓扑收集和备交换机选举 软件和配置同步

堆叠卡堆叠



业务口堆叠



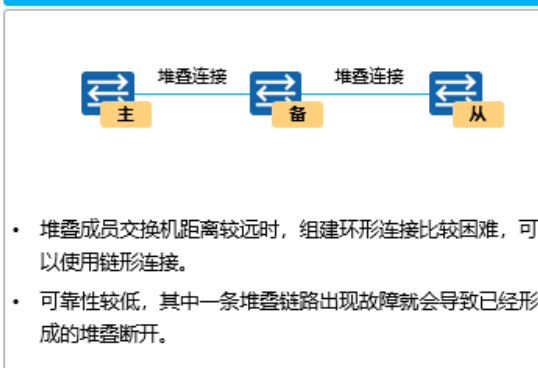
- 每台交换机支持两个逻辑堆叠端口，分别为 stack-port n/1 和 stack-port n/2，其中 n 为成员交换机的堆叠 ID。

物理连接 主交换机选举 拓扑收集和备交换机选举 软件和配置同步

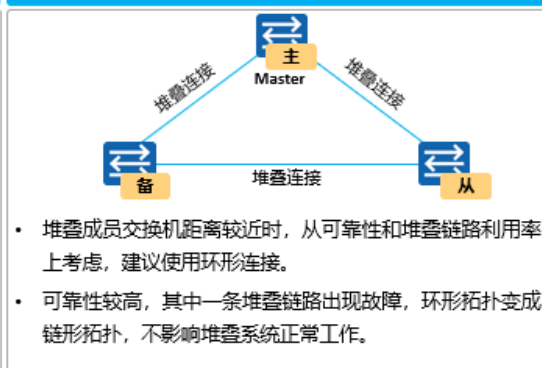


堆叠连接拓扑

链形连接



环形连接





主交换机选举

- 确定出堆叠的连接方式和连接拓扑，完成成员交换机之间的物理连接之后，所有成员交换机上电。此时，堆叠系统开始进行主交换机的选举。

主交换机的选举规则

- 1 运行状态比较，已经运行的交换机比处于启动状态的交换机优先竞争为主交换机
堆叠主交换机选举超时时间为20s，堆叠成员交换机上电或重启时，由于不同成员交换机所需的启动时间可能差异比较大，因此不是所有成员交换机都有机会参与主交换机的第一次选举。
- 2 堆叠优先级高的交换机优先竞争为主交换机
- 3 堆叠优先级相同时，MAC地址小的交换机优先竞争为主交换机



备交换机选举

拓扑搜集、堆叠ID分配及Standby交换机选举

主交换机选举完成后，主交换机会收集所有成员交换机的拓扑信息，并向所有成员交换机分配堆叠ID。之后进行备交换机的选举，作为主交换机的备份交换机。

除主交换机外最先完成设备启动的交换机优先被选为备份交换机。

当除主交换机外其它交换机同时完成启动时，备交换机的选举规则如下（依次从第一条开始判断，直至找到最优的交换机才停止比较）：

- 1 堆叠优先级最高的交换机成为备交换机
- 2 堆叠优先级相同时，MAC地址最小的成为备交换机



软件、配置同步

软件和配置同步、稳定运行

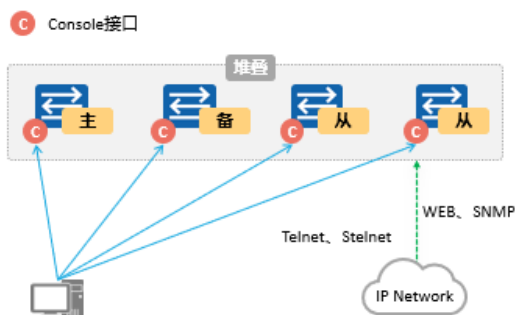
角色选举、拓扑收集完成之后，所有成员交换机会自动同步主交换机的系统软件和配置文件：

- 1 堆叠具有自动加载系统软件的功能，待组成堆叠的成员交换机不需要具有相同软件版本，只需要版本间兼容即可。
当备交换机或从交换机与主交换机的软件版本不一致时，备交换机或从交换机会自动从主交换机下载系统软件，然后使用新系统软件重启，并重新加入堆叠。
- 2 堆叠具有配置文件同步机制，备交换机或从交换机会将主交换机的配置文件同步到本设备并执行，以保证堆叠中的多台设备能够像一台设备一样在网络中工作，并且在主交换机出现故障之后，其余交换机仍能够正常执行各项功能。



堆叠管理与配置文件

堆叠管理



- 登录任意成员的Console口即可管理整个堆叠系统。
- 与堆叠系统上的任意IP地址路由可达，即可通过Telnet、Stelnet、WEB、SNMP等方式登录、管理整个堆叠系统。

配置文件



- 堆叠建立后，竞争为主交换机的配置文件生效，整个堆叠系统的配置都保存在主交换机的配置文件中。
- 备交换机会备份主交换机的配置文件，当主交换机故障时备份交换机的配置继续运行。

- 堆叠系统的配置文件包括：
- 全局配置：例如 IP 地址、STP、VLAN、SNMP 等，适用于所有堆叠成员交换机。新加入堆叠系统的交换机使用堆叠系统的全局配置，不再使用交换机自己的全局配置。交换机离开堆叠系统时，将继续使用堆叠系统的配置直到加入新的堆叠系统。
- 接口配置：适用于接口所在成员交换机。接口上的配置和堆叠 ID 有关，当堆叠 ID 改变时：
- 如果新 ID 在配置文件中不存在对应的接口配置，则新 ID

的接口配置使用默认配置。

- 如果新 ID 在配置文件中存在对应的接口配置，则新 ID 的接口配置使用对应的配置。
- 堆叠相关配置不会丢失，即从设备离开堆叠之后，堆叠相关配置仍会继续存在，同样会进行堆叠的选举，如果需要彻底清除堆叠配置需要通过命令手动清除，清除的配置内容包括：

- 交换机槽位号
- 堆叠优先级
- 堆叠保留 VLAN
- 系统 MAC 切换时间
- 堆叠口配置
- 堆叠口速率配置

执行该命令会导致设备重启



堆叠成员退出

堆叠成员退出是指成员交换机从堆叠系统中离开。根据退出成员交换机角色的不同，对堆叠系统的影响也有所不同：

- 当主交换机退出，备份交换机升级为主交换机，重新计算堆叠拓扑并同步到其他成员交换机，指定新的备交换机，之后进入稳定运行状态。
- 当备交换机退出，主交换机重新指定备交换机，重新计算堆叠拓扑并同步到其他成员交换机，之后进入稳定运行状态。
- 当从交换机退出，主交换机重新计算堆叠拓扑并同步到其他成员交换机，之后进入稳定运行状态。



管理员拆除堆叠线缆

- 堆叠成员交换机退出的过程，主要就是拆除堆叠线缆和移除交换机的过程：
- 对于环形堆叠：成员交换机退出后，为保证网络的可靠性还需要把退出交换机连接的两个端口通过堆叠线缆进行连接。
- 对于链形堆叠：拆除中间交换机会造成堆叠分裂。这时需要在拆除前进行业务分析，尽量减少对业务的影响。

堆叠成员加入

堆叠成员加入是指向已经稳定运行的堆叠系统添加一台新的交换机：

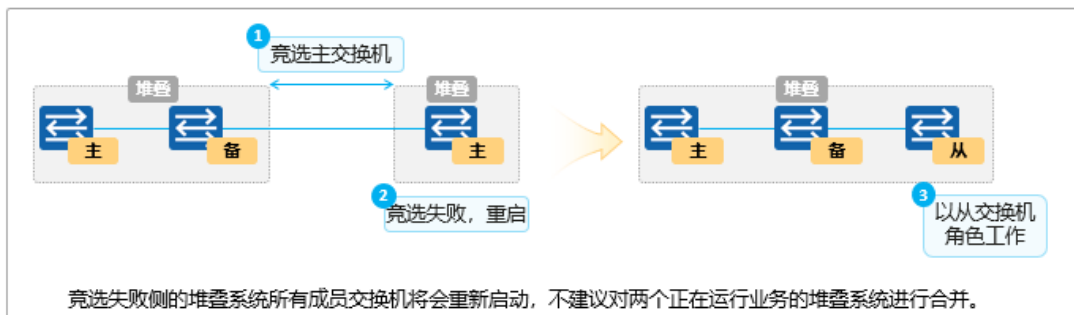
- 将未上电的交换机连线加入堆叠之后再上电启动，新加入的交换机会选举为从交换机，堆叠系统中原有主备从角色不变。
- 角色选举结束后，主交换机更新堆叠拓扑信息，同步到其他成员交换机上，并向新加入的交换机分配堆叠ID（新加入的交换机没有配置堆叠ID或配置的堆叠ID与原堆叠系统的冲突时）。
- 新加入的交换机更新堆叠ID，并同步主交换机的配置文件和系统软件，之后进入稳定运行状态。



- 未上电的交换机加入堆叠：
- 如果是链形连接，新加入的交换机建议添加到链形的两端，这样对现有的业务影响最小。
- 如果是环形连接，需要把当前环形拆成链形，然后在链形的两端添加设备。

堆叠合并

- 堆叠合并是指稳定运行的两个堆叠系统合并成一个新的堆叠系统。
- 例如：已上电的一台交换机并且配置了堆叠（已形成单机堆叠），通过堆叠线缆与已经在运行的堆叠系统连接。该过程为堆叠合并，与堆叠加入不同。

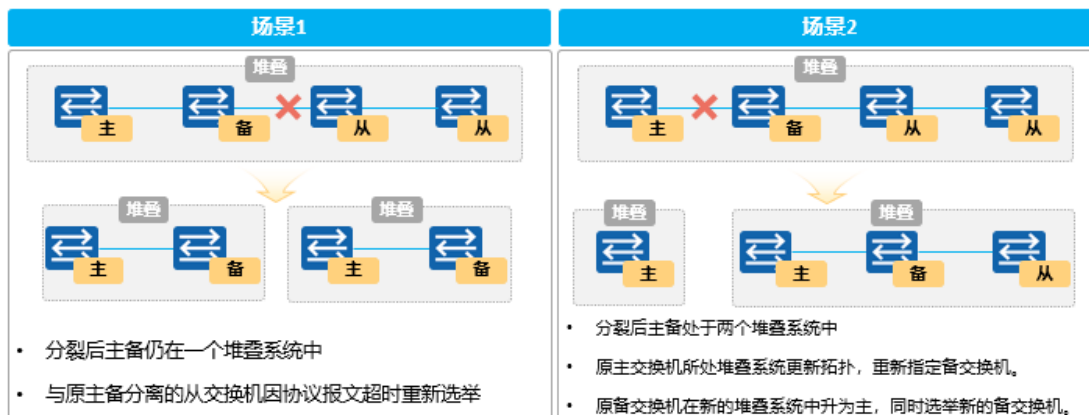


- 单机堆叠：单机堆叠就是一台交换机使能了堆叠功能，整个堆叠系统中只有一台主交换机。只有使能了堆叠功能的交换机才可以加入堆叠系统或与其他使能了堆叠功能的交换机组建堆叠。
- 堆叠合并过程如下：

- 两个堆叠系统的主交换机通过竞争，选举出一个更优的作为新堆叠系统的主交换机。
- 竞争成功的主交换机所在的堆叠系统将保持原有主备从角色和配置不变，业务也不会受到影响
- 而另外一个堆叠系统的所有成员交换机将重新启动，以从交换机的角色加入到新堆叠系统，其堆叠 ID 将由新主交换机重新分配，并将同步新主交换机的配置文件和系统软件，该堆叠系统的原有业务也将中断。

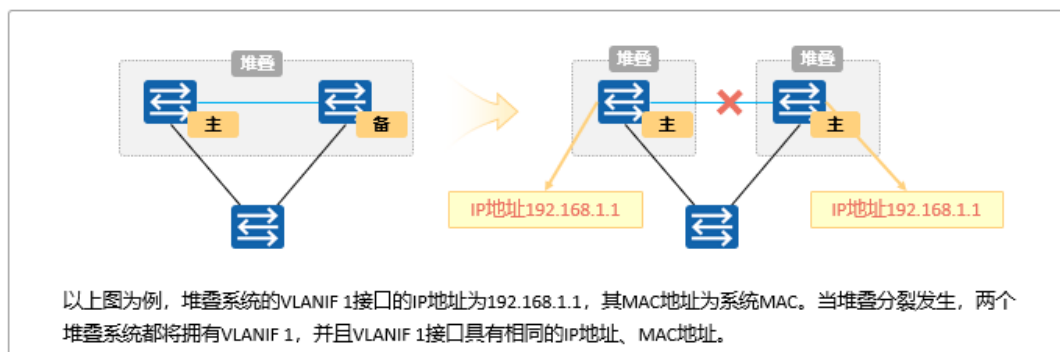
堆叠分裂

堆叠分裂是指稳定运行的堆叠系统中带电移出部分成员交换机，或者堆叠线缆多点故障导致一个堆叠系统变成多个堆叠系统。



堆叠分裂引起的问题

由于堆叠系统中所有成员交换机都使用同一个IP地址（VLANIF接口地址）和MAC地址（堆叠系统MAC），一个堆叠系统分裂后，可能产生多个具有相同IP地址和MAC地址的堆叠系统，从而引起网络故障，为此必须进行IP地址和MAC地址的冲突检查。



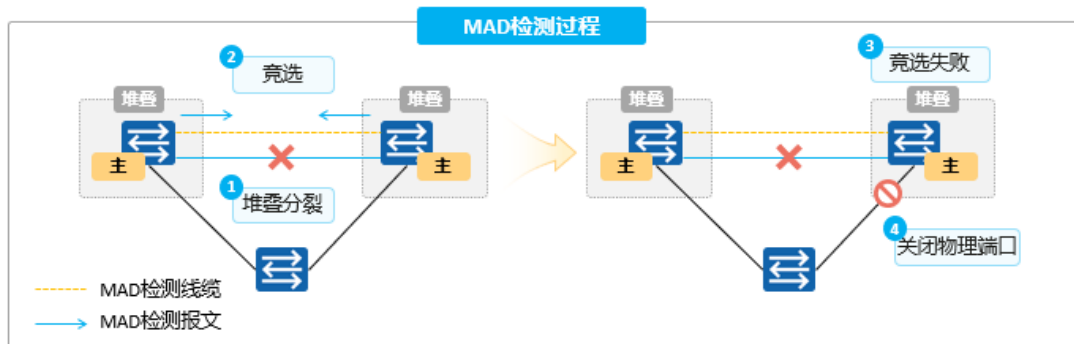
- 堆叠系统作为一台设备与网络中其他设备通信，具有唯

一的 MAC 地址，称为堆叠系统 MAC 地址。通常情况下使用主交换机的 MAC 地址作为堆叠系统 MAC 地址。

- 当堆叠系统的 MAC 地址是主交换机的 MAC 地址，主交换机故障或者离开堆叠系统，在默认情况下堆叠系统 MAC 地址会延时 10 分钟切换，即在 10 分钟内两个分裂的堆叠系统的 MAC 相同。

MAD检测

- 多主检测MAD（Multi-Active Detection）：一种检测和处理堆叠分裂的协议，链路故障导致堆叠系统分裂后，MAD可以实现堆叠分裂的检测、冲突处理和故障恢复，降低堆叠分裂对业务的影响。
- MAD检测方式有两种：直连检测方式和代理检测方式。在同一个堆叠系统中，两种检测方式互斥，不可以同时配置。

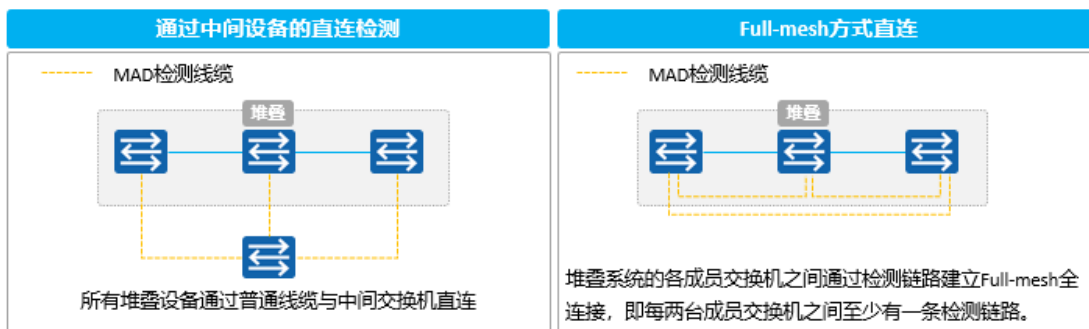


- 分裂后的堆叠系统通过 MAD 检测线缆（普通线缆，手动配置为 MAD 检测链路）发送 MAD 检测报文进行竞选，竞选失败的堆叠系统会关闭所有的物理端口（手动配置的保留端口除外）以保证不会因 IP、MAC 冲突对业务产生影响。
- MAD 检测过程如下：
- 堆叠分裂之后，分裂后的堆叠系统存在相同的 IP 地址、MAC 地址（堆叠系统 MAC），将会导致网络中表项错误，如下游设备的 ARP 表项、MAC 地址表项，从而导致业务异常。
- 为此分类后的堆叠系统会通过 MAC 检测线缆进行竞选。
- 竞选失败者关闭自身物理接口，从而不会和竞选成功地堆叠系统产生 IP 地址、MAC 地址冲突。



MAD检测 - 直连检测

直连检测方式是指堆叠成员交换机间通过普通线缆直连的专用链路进行多主检测。在直连检测方式中，堆叠系统正常运行时，不发送MAD报文；堆叠系统分裂后，分裂后的两台交换机以1秒为周期通过检测链路发送MAD报文进行多主冲突处理。



- 通过中间设备直连可以实现通过中间设备缩短堆叠成员交换机之间的检测链路长度，适用于成员交换机相距较远的场景。与通过中间设备直连相比，Full-mesh方式直连可以避免由中间设备故障导致的MAD检测失败，但是每两台成员交换机之间都建立全连接会占用较多的接口，所以该方式适用于成员交换机数目较少的场景。



MAD检测 - 代理检测

代理检测方式是在堆叠系统Eth-Trunk上启用代理检测，在代理设备上启用MAD检测功能。此种检测方式要求堆叠系统中的所有成员交换机都与代理设备连接，并将这些链路加入同一个Eth-Trunk内。与直连检测方式相比，代理检测方式无需占用额外的接口，Eth-Trunk接口可同时运行MAD代理检测和其他业务。



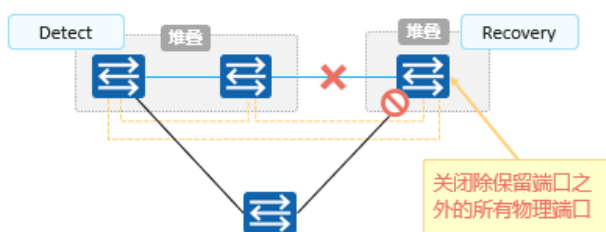
- 在代理检测方式中，堆叠系统正常运行时，堆叠成员交换机以30s为周期通过检测链路发送MAD报文。堆叠成员交

交换机对在正常工作状态下收到的 MAD 报文不做任何处理；堆叠分裂后，分裂后的两个堆叠系统以 1s 为周期通过检测链路发送 MAD 报文进行多主冲突处理。

MAD冲突处理 (1)

堆叠分裂后，MAD冲突处理机制使用MAD报文进行MAD竞争，竞争结果为堆叠系统处于Detect状态或者Recovery状态：

- Detect：竞争成功，堆叠系统将处于正常工作状态。
- Recovery：竞争失败，堆叠系统将状态处于禁用状态，关闭除手动配置的保留端口以外的其它所有物理端口。



MAD冲突处理 (2)

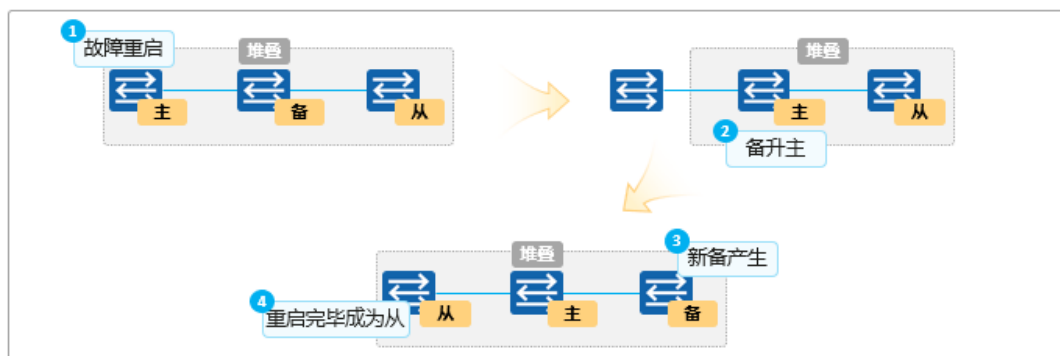
MAD竞争原则

MAD竞争原则与主交换机的竞争原则类似：

- 1 先比较启动时间，启动完成时间早的堆叠系统成为Detect状态。启动完成时间差在20秒内则认为堆叠的启动完成时间相同。
- 2 启动完成时间相同时，比较堆叠中主交换机的优先级，优先级高的堆叠系统成为Detect状态。
- 3 优先级相同时，比较堆叠系统的MAC，MAC小的堆叠系统成为Detect状态。

堆叠主备倒换

- 如果堆叠系统当前的主交换机不是用户期望的，此时可以通过配置主备倒换实现将堆叠备交换机升为主交换机。
- 除了用户通过命令执行的主备倒换之外，主交换机故障重启也会引起主备倒换。

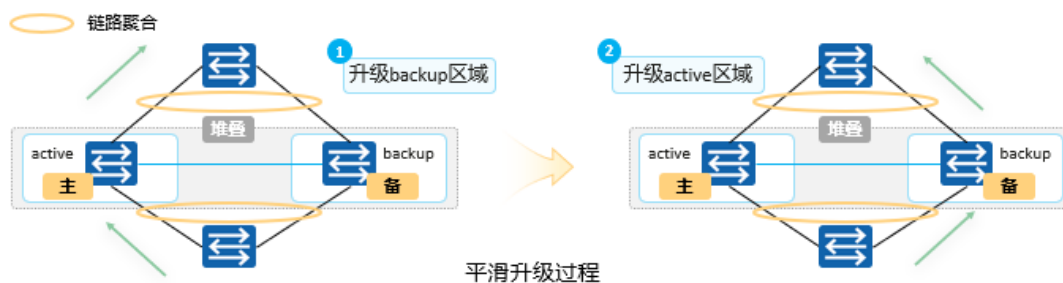


- 原来的备交换机升为主交换机
- 新主交换机重新指定备交换机
- 原来的主交换机重启后重新加入堆叠系统，并被选举为从交换机。

堆叠升级

堆叠升级方式有三种：智能升级、传统升级和平滑升级。

- 智能升级：堆叠建立或者新的交换机加入堆叠时会自动和主交换机的版本进行同步。
- 传统升级：和普通设备升级一样，指定下次启动版本，重启整个堆叠系统进行升级，会造成较长时间的业务中断。
- 平滑升级：将堆叠系统划分成为active、backup区域，可以分区域升级，整个堆叠系统的上下行采用备份组网，主、备链路分别处于active、backup区域，可以实现升级时的业务不中断。

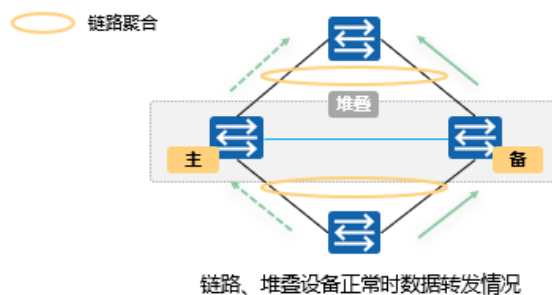


- Active 区域：主交换机所在的区域
- 平滑升级的三个阶段：
- 主交换机下发命令触发整个堆叠系统进入平滑升级状态，backup 区各个成员交换机以新的系统软件进行启动。

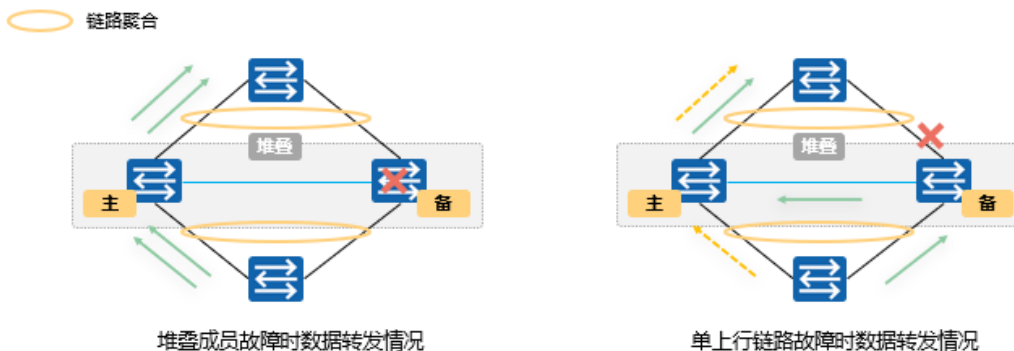
- backup 区以新版本建立一个独立的堆叠系统，并通知 active 区进入升级阶段，主控权由 active 区的主交换机转移到 backup 区的主交换机，backup 区负责流量传输，active 区进入升级过程。
- active 区以新系统软件重新启动并加入 backup 区的堆叠系统，backup 区的主交换机根据最终堆叠建立的结果发布升级的结果。

跨设备链路聚合 (1)

- 堆叠支持跨设备链路聚合技术，堆叠后成为逻辑上的一台交换机，支持将Eth-Trunk的成员接口分布在不同的成员交换机上。
- 当其中一条聚合链路故障或堆叠中某台成员交换机故障时，Eth-Trunk接口通过堆叠线缆将流量重新分布到其他聚合链路上，实现了链路间和设备间的备份，保证了数据流量的可靠传输。



跨设备链路聚合 (2)



- 如上图所示，在上行链路出现故障时跨设备链路聚合技

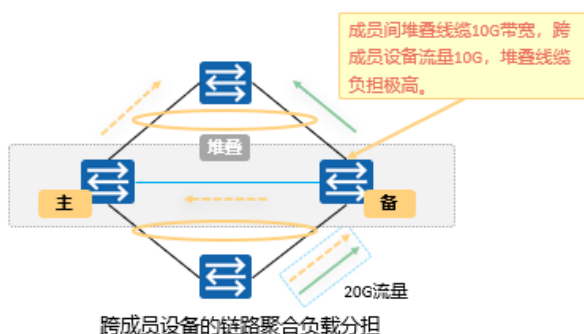
术能够通过堆叠线缆将流量负载分担到其他成员交换机上的成员接口，从而提高网络可靠性。

- 跨成员设备的流量会经过堆叠线缆，如果跨成员设备流量较大，将会大大增加堆叠线缆的带宽压力。



流量本地优先转发 (1)

链路聚合的负载分担算法根据流量特征将报文分担在不同的成员链路上，对于跨设备链路聚合极有可能会出现报文的出接口和入接口不在同一台成员设备之上的情况，此时堆叠成员之间将会通过堆叠线缆转发流量，这增加了堆叠线缆的流量负担，同时也降低了转发效率。

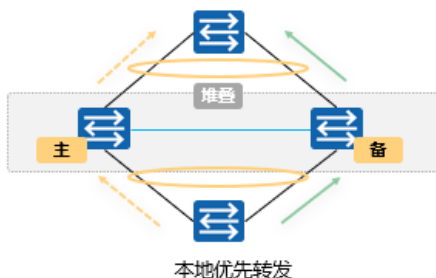


- 如上图所示，当链路聚合将流量分担到其他成员设备上时，部分流量将会经过堆叠线缆转发，这会大大增加堆叠线缆的带宽承载压力。如果通过堆叠线缆需要转发的流量甚至超过了堆叠线缆的带宽，报文将无法及时得到传输。



流量本地优先转发 (2)

为保证流量转发效率、降低堆叠线缆带宽负载，设备可以开启流量本地优先转发，从本设备进入的流量优先从本地转发出去，当本设备无出接口或者出接口全部故障，才会从其它成员交换机的接口转发出去。





集群介绍

- 集群交换机系统CSS (Cluster Switch System)，又称为集群，是指将两台支持集群特性的交换机设备组合在一起，从逻辑上虚拟成一台交换设备。
- CSS与iStack的区别在于，一般框式交换机堆叠称为CSS，盒式交换机称为堆叠，堆叠与集群两者只是叫法和实现有些差异，但是功能是一样的。
- 本小节着重介绍集群与堆叠存在的差异之处。

集群内物理设备转发平面合一，转发信息共享并实时同步。

集群交换机对外表现为一台逻辑交换机，控制平面合一，统一管理。

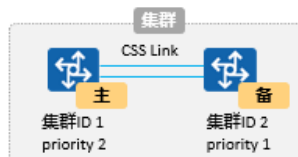
跨集群内物理设备的链路被聚合成一个Eth-Trunk端口和下游设备实现互联。

集群特点



集群基本概念

- 集群中的单台交换机称为集群成员交换机，按照功能不同，可以分为两种角色：
 - 主交换机 (Master)：主交换机，即Master，负责管理整个集群。
 - 备交换机 (Standby)：备交换机，即Standby，是主交换机的备份交换机。
- 集群ID：即CSS ID，用来标识成员交换机，集群中成员交换机的集群ID是唯一的。
- CSS Link：集群链路，专门用于组建集群，实现主交换机和备交换机之间数据通信。
- 集群优先级：即CSS Priority，主要用于角色选举过程中确定成员交换机的角色。优先级值越大优先级越高。



- 备交换机，即 Standby，是主交换机的备份交换机。当主交换机故障时，备交换机会接替原主交换机的所有业务。集群中只有一台备交换机。
- 集群链路可以是一条链路，也可以是捆绑在一起的多条链路。
- 缺省情况下，交换机的集群 ID 为 1。相同 ID 的两台交换机不能建立集群，所以在建立集群前，需要手工配置集群中一台交换机的集群 ID 为 2。
- 交换机的集群优先级越高当选为主交换机的可能性越大，

但选举主交换机比较的不止优先级。

集群控制平面

- 两台交换机使用集群线缆连接好，分别使能集群功能并完成配置后重启，集群系统会自动建立。
- 集群系统建立后，在控制平面上，主交换机的主用主控板成为集群系统的控制平面，作为整个系统的管理主角色。备交换机的主用主控板成为集群系统的备用控制平面，作为系统的管理备角色。主交换机和备交换机的备用主控板作为集群系统候选备用主控板。



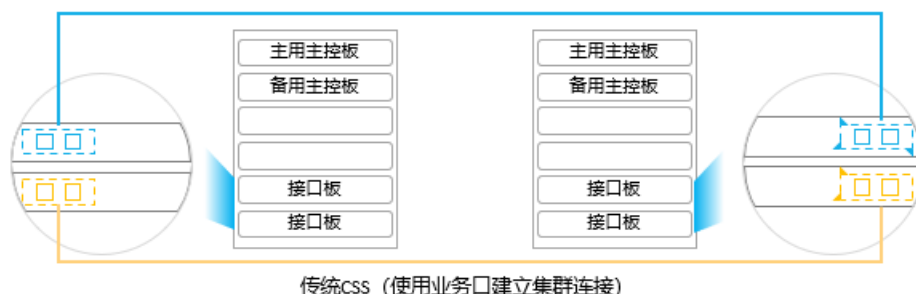
- 集群主交换机选举规则与堆叠选举规则一致，另外一台交换机则是备交换机。

集群物理连接

根据集群技术发展阶段不同，集群物理连接方式也存在区别：

- 传统CSS：使用主控板上的集群卡建立集群连接，或者使用业务口建立集群连接。
- CSS2：第二代集群交换机系统，专指使用交换网板上的集群卡方式建立集群连接的集群。

□ 物理成员端口 □ 逻辑集群端口 — 集群连接

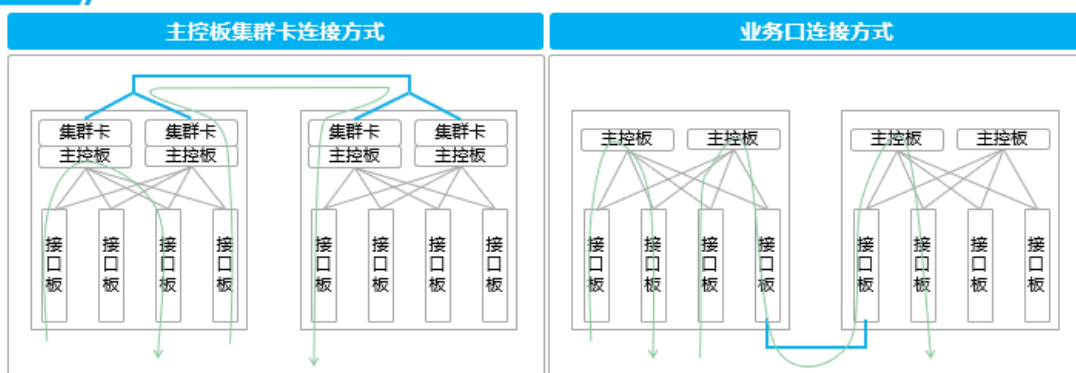


- 使用业务口连接时两端成员接口数量、类型相同，接线顺序无限制。
- 业务口集群按照链路的分布，有两种组网形式：
- 1+0 组网：每台成员交换机配置一个逻辑集群端口，物理成员端口分布在一块业务板上，依靠一块业务板上物理成员端口与对端的物理成员端口实现集群连接。

- 1+1 组网：每台成员交换机配置两个逻辑集群端口，物理成员端口分布在两块业务板上，如图所示，不同业务板上的集群链路形成备份。
- CSS2：专指交换网板上通过集群卡方式建立的交换网硬件集群，并且在原有集群技术的基础上，增加了集群主控 1+N 备份等技术。



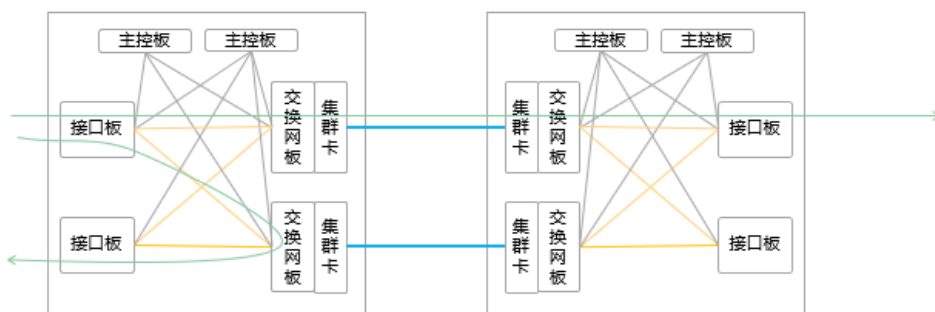
传统CSS



- 对于只支持CSS构架的框式交换机，框内接口板之间流量、跨框流量必须经过主控板。
- 单框上没有正常工作的主控板时流量无法从一个接口板转发到另外一个接口板，同时也无法跨框转发到另一个框。



CSS2



- 支持CSS2构架的框式交换机采用转控分离的构架，单框内接口板之间流量、跨框流量无需经过主控板，集群系统内单台框无能够正常工作的主控板不影响该框的流量转发。
- CSS2支持任意一个框式交换机内存在一个主控板运行正常，集群的两个框式交换机上的接口板都可以正常转发报文，该特性被称为“集群主控1+N备份”。



堆叠配置 (1)

1. 创建堆叠逻辑接口，绑定物理成员端口到堆叠接口中

```
[Huawei] interface stack-port member-id/port-id  
[Huawei-stack-port0/1] port interface { interface-type interface-number1 [ to interface-type interface-number2 ] } <1-10>  
enable
```

Member-id为设备的堆叠ID，port-id为本地的堆叠逻辑接口编号，只能为1或者2。

2. 配置设备的堆叠ID

```
[Huawei] stack slot slot-id renumber new-slot-id
```

缺省情况下，设备的堆叠ID为0。修改后的堆叠ID在保存当前配置并重启之后才会生效。

3. 配置设备的堆叠优先级

```
[Huawei] stack slot slot-id priority priority
```

缺省情况下，成员交换机的堆叠优先级为100。



堆叠配置 (2)

4. 配置直连方式多主检测

```
[Huawei-GigabitEthernet0/0/1] mad detect mode direct
```

在接口视图下配置接口的直连多主检测功能。

5. 代理方式多主检测，当代理设备为一台交换机时，在堆叠系统上

```
[Huawei] interface eth-trunk trunk-id  
[Huawei-Eth-Trunk1] mad detect mode relay
```

在与代理设备互联的Eth-Trunk中开启MAD检测。

6. 代理方式多主检测，当代理设备为一台交换机时，在指定的代理设备上

```
[Huawei] interface eth-trunk trunk-id  
[Huawei-Eth-Trunk1] mad relay
```

代理设备上与堆叠系统互联的Eth-Trunk中开启MAD代理。



堆叠配置 (3)

7. 代理方式多主检测，两个堆叠系统互为代理，配置堆叠系统MAD域值

```
[Huawei] mad domain domain-id
```

缺省情况下，堆叠系统MAD域值为0，堆叠系统互为代理时需为两套堆叠系统配置不同的MAD域值。

8. 代理方式多主检测，两个堆叠系统互为代理，开启MAD检测与MAD代理

```
[Huawei] interface eth-trunk trunk-id  
[Huawei-Eth-Trunk1] mad detect mode relay  
[Huawei-Eth-Trunk1] mad relay
```

在与代理设备互联的Eth-Trunk中开启MAD检测。

9. 堆叠主备倒换

```
[Huawei] slave switchover
```

执行堆叠主备倒换。



堆叠配置 (4)

10. 配置堆叠系统MAC地址切换时间

```
[Huawei] stack timer mac-address switch-delay delay-time
```

缺省情况下，系统MAC地址的切换时间为10分钟，堆叠系统MAC地址切换时间设置为0时，表示不切换，执行undo stack timer mac-address switch-delay命令表示立即切换。

11. 清除堆叠的所有配置

```
[Huawei] reset stack configuration
```

执行该命令将清除堆叠的所有配置，包括：交换机槽位号、堆叠优先级、堆叠保留VLAN、系统MAC切换时间、堆叠口配置、堆叠口速率配置。例如，交换机插入专用堆叠线缆进行堆叠时，为了根据专用堆叠线缆的连线顺序自动生成槽位号，需要先执行该命令清除堆叠的配置。

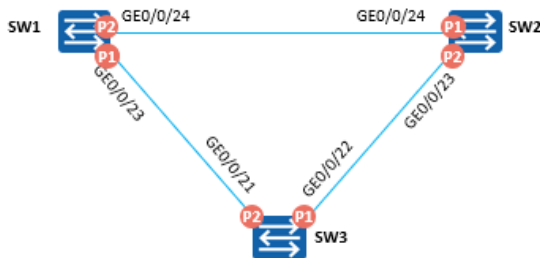
执行该命令后会导致原有堆叠系统分裂，设备重启。



堆叠配置案例 (1)

P1 Stack-port 1

P2 Stack-port 2



- SW1、SW2和SW3三台接入交换机采用环形堆叠组网。
- SW1、SW2和SW3的角色分别为主、备、从，堆叠ID分别为0、1、2，优先级分别为200、100、100。

配置逻辑端口并加入物理成员端口：

SW1的配置如下：

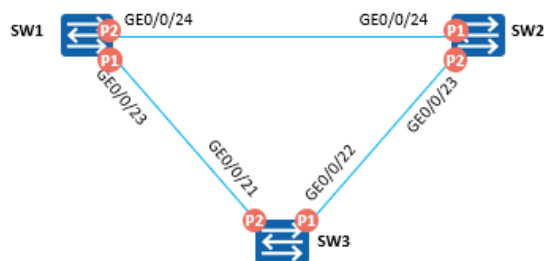
```
[SW1] interface stack-port 0/1
[SW1-stack-port0/1] port interface gigabitethernet 0/0/23 enable
[SW1] interface stack-port 0/2
[SW1-stack-port0/2] port interface gigabitethernet 0/0/24 enable
```

SW2的配置如下：

```
[SW2] interface stack-port 0/1
[SW2-stack-port0/1] port interface gigabitethernet 0/0/24 enable
[SW2] interface stack-port 0/2
[SW2-stack-port0/2] port interface gigabitethernet 0/0/23 enable
```

堆叠配置案例 (2)

- P1 Stack-port 1
- P2 Stack-port 2



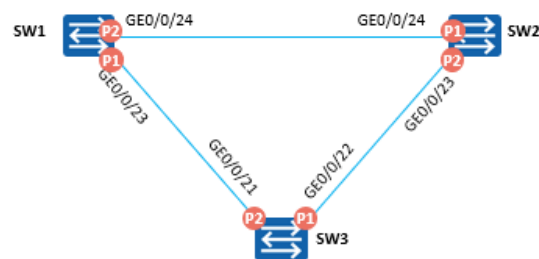
SW3的配置如下:

```
[SW3] interface stack-port 0/1
[SW3-stack-port0/1] port interface gigabitethernet 0/0/22 enable
[SW3] interface stack-port 0/2
[SW3-stack-port0/2] port interface gigabitethernet 0/0/21 enable
```

- SW1、SW2和SW3三台接入交换机采用环形堆叠组网。
- SW1、SW2和SW3的角色分别为主、备、从，堆叠ID分别为0、1、2，优先级分别为200、100、100。

堆叠配置案例 (3)

- P1 Stack-port 1
- P2 Stack-port 2



配置堆叠ID和堆叠优先级:

SW1的配置如下:

```
[SW1] stack slot 0 priority 2
```

SW2的配置如下:

```
[SW2] stack slot 0 rename 1
```

SW3的配置如下:

```
[SW3] stack slot 0 rename 2
```

- SW1、SW2和SW3三台接入交换机采用环形堆叠组网。
- SW1、SW2和SW3的角色分别为主、备、从，堆叠ID分别为0、1、2，优先级分别为200、100、100。

SW1、SW2、SW3下电，使用SFP+电缆连接后再上电，下电之前使用命令save保存配置。

本设备的stack-port 0/1必须连接邻设备的stack-port 0/2，否则堆叠组建不成功。

集群配置 (1)

1. 配置交换机的集群ID

```
[Huawei] set css id new-id
```

缺省情况下，交换机的集群ID都为1，所以在建立集群前，需要手工配置集群中一台交换机的集群ID为2，相同ID的两台交换机不能建立集群。

2. 创建集群逻辑接口，绑定物理成员端口到堆叠接口中

```
[Huawei] interface css-port port-id  
[Huawei-css-port1] port interface { interface-type interface-number1 [ to interface-type interface-number2 ] } &<1-10> enable
```

配置业务口为物理成员端口，并将物理成员端口加入到逻辑集群端口中。

3. 配置设备的集群优先级

```
[Huawei] set css priority priority
```

缺省情况下，设备的集群优先级为1。

集群配置 (2)

4. 使能交换机集群功能

```
[Huawei] css enable
```

缺省情况下，交换机的集群功能未使能，使能集群功能后，系统会提示立即重启使配置生效。需要在两台成员交换机上分别使能集群功能。

5. 设置设备的集群连接方式

```
[Huawei] set css mode { lpu | css-card }
```

缺省情况下设备的集群连接方式与设备型号相关。

集群配置案例 (1)



- SW1、SW2分别安装业务板并连接集群线缆。为了增强带宽和可靠性，两台交换机之间通过两块业务板上的四对业务口进行连接。
- 在SW1和SW2上分别配置集群连接方式，配置集群ID分别为1和2，配置集群优先级分别为100和10，以提高SW1成为主交换机的可能。

配置集群连接方式、集群ID及集群优先级

SW1的配置如下：

```
[SW1] set css mode lpu  
[SW1] set css id 1  
[SW1] set css priority 100
```

SW2的配置如下：

```
[SW2] set css mode lpu  
[SW2] set css id 2  
[SW2] set css priority 10
```


集群配置案例 (2)



- SW1、SW2分别安装业务板并连接集群线缆。为了增强带宽和可靠性，两台交换机之间通过两块业务板上的四对业务口进行连接。
- 在SW1和SW2上分别配置集群连接方式，配置集群ID分别为1和2，配置集群优先级分别为100和10，以提高SW1成为主交换机的可能。

配置逻辑集群端口

SW1的配置如下：

```
[SW1] interface css-port 1
[SW1-css-port1] port interface xgigabitethernet 1/0/1 to
xgigabitethernet 1/0/2 enable
[SW1-css-port1] quit
[SW1] interface css-port 2
[SW1-css-port2] port interface xgigabitethernet 2/0/1 to
xgigabitethernet 2/0/2 enable
[SW1-css-port2] quit
```

配置SW1、SW2的业务口XGE1/0/1 ~ XGE1/0/2为集群物理成员端口并加入集群端口1，XGE2/0/1 ~ XGE2/0/2为集群物理成员端口并加入集群端口2。

SW2与SW1配置相同，省略。

集群配置案例 (3)



- SW1、SW2分别安装业务板并连接集群线缆。为了增强带宽和可靠性，两台交换机之间通过两块业务板上的四对业务口进行连接。
- 在SW1和SW2上分别配置集群连接方式，配置集群ID分别为1和2，配置集群优先级分别为100和10，以提高SW1成为主交换机的可能。

使能集群功能

```
[SW1] css enable
Warning: The CSS configuration will take effect only after the
system is rebooted. T
he next CSS mode is LPU. Reboot now? [Y/N]:y
[SW2] css enable
Warning: The CSS configuration will take effect only after the
system is rebooted. T
he next CSS mode is LPU. Reboot now? [Y/N]:y
```

配置SW1、SW2的业务口XGE1/0/1 ~ XGE1/0/2为集群物理成员端口并加入集群端口1，XGE2/0/1 ~ XGE2/0/2为集群物理成员端口并加入集群端口2。

SW2与SW1配置相同，省略。

思考题：

- （简答题）堆叠加入与堆叠合并有什么区别？
- （简答题）多主检测的目的是？多主检测失败的一方会采取什么操作？
- （简答题）CSS2 相比较于传统 CSS 区别在于？

答案：

- 堆叠加入是交换机未上电之前已经通过堆叠线缆连接到了已有的堆叠系统，当上电启动之后因为堆叠系统已经存在主

交换机，所以该交换机只会竞选成为从交换机。堆叠合并是两个堆叠系统通过堆叠线缆连在一起，此时两个堆叠系统会竞选新的主交换机并更新拓扑。

一台交换机已经上电之后自身会成为一个堆叠系统并且由于没有竞争者自己会成为主交换机，这种情况下把该交换机通过堆叠线缆与另外一个堆叠系统相连就属于堆叠合并，这是一种典型的堆叠合并与堆叠加入的差异性场景。

- 防止堆叠、集群分裂之后相同的 IP 地址、MAC 地址造成网络中其他网络设备的表项冲突，从而导致转发异常。检测失败的一方将会关闭除保留端口之外的所有端口，从而避免相同 IP 地址、相同 MAC 地址对其他网络设备造成表项冲突。
- CSS2 支持 1+N 主控备份，整个集群系统中只要有一个主控板处于工作状态，整个集群系统的控制平面可以正常工作的情况下，集群系统整体的数据平面就可以正常转发报文。