Huawei Enterprise A Better Way

数据中心网络FCoE解决方案

www.huawei.com

HUAWEI

目录

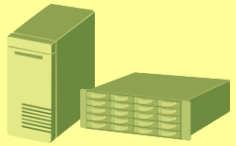
从存储技术的发展历史说起...

1 嵌入式存储（ES）

- **部署方式**：部署在服务器内部
- **特　　点**：无法共享存储，扩展性差
- **典型应用**：PC、服务器内置硬盘

2 直接存储（DS）

- **部署方式**：服务器直接连接到存储阵列
- **特　　点**：无法共享，扩展性受限于存储设备的接入能力
- **典型应用**：服务器扩展本地存储时使用

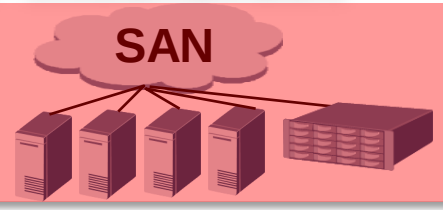


3 联网存储（NAS）

- **部署方式**：服务器通过IP网络连接到存储阵列
- **特　　点**：可共享存储，可扩展，文件式存储
- **典型应用**：配合文件服务器使用

4 存储区域网络（SAN）

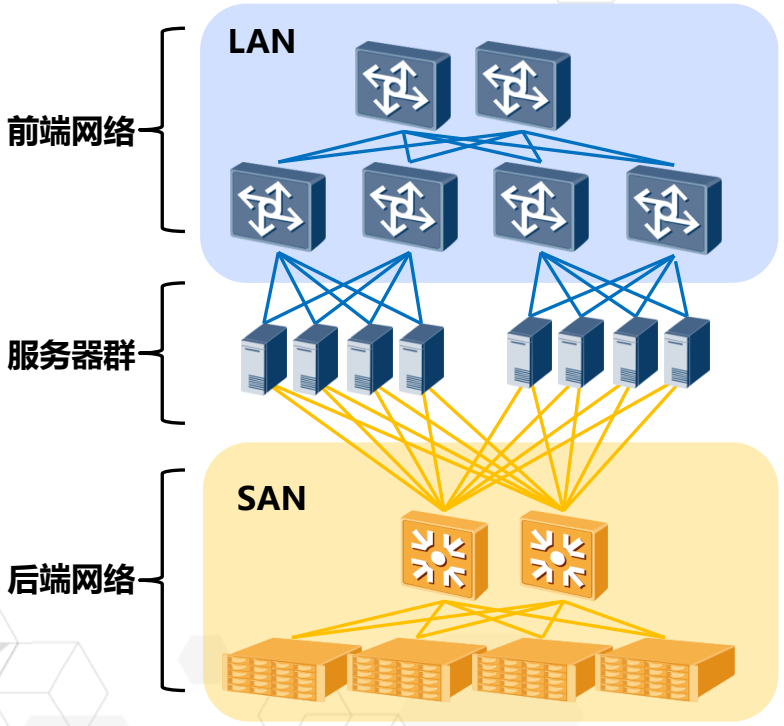
- **部署方式**：服务器通过IP或FC网络连接到存储阵列
- **特　　点**：可共享存储，可扩展，高效率块存储
- **典型应用**：配合数据库服务器使用



从现有的存储技术来看，FC SAN能够针对共享存储提供高速的访问，而FC SAN却无法与以太兼容。

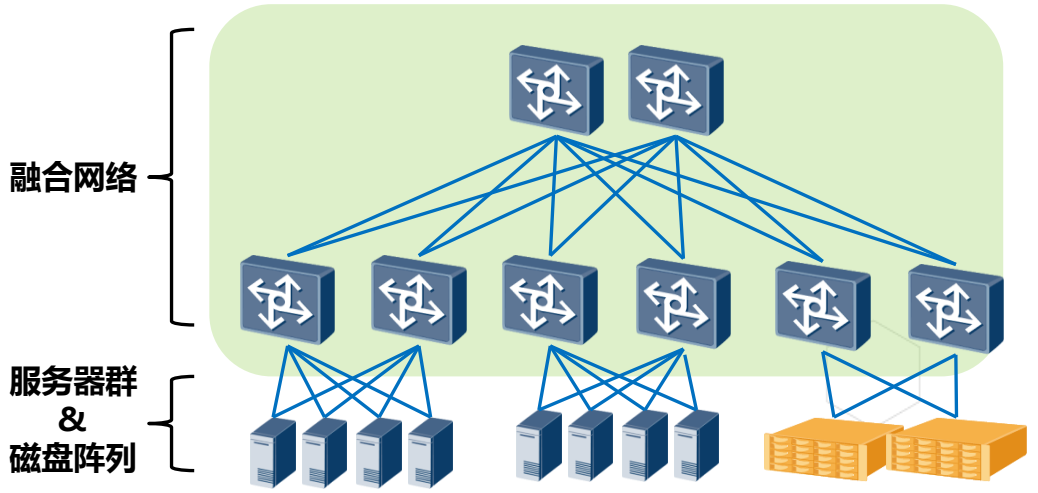
数据中心的融合网络

传统的数据中心架构



融合后

数据中心融合网络架构



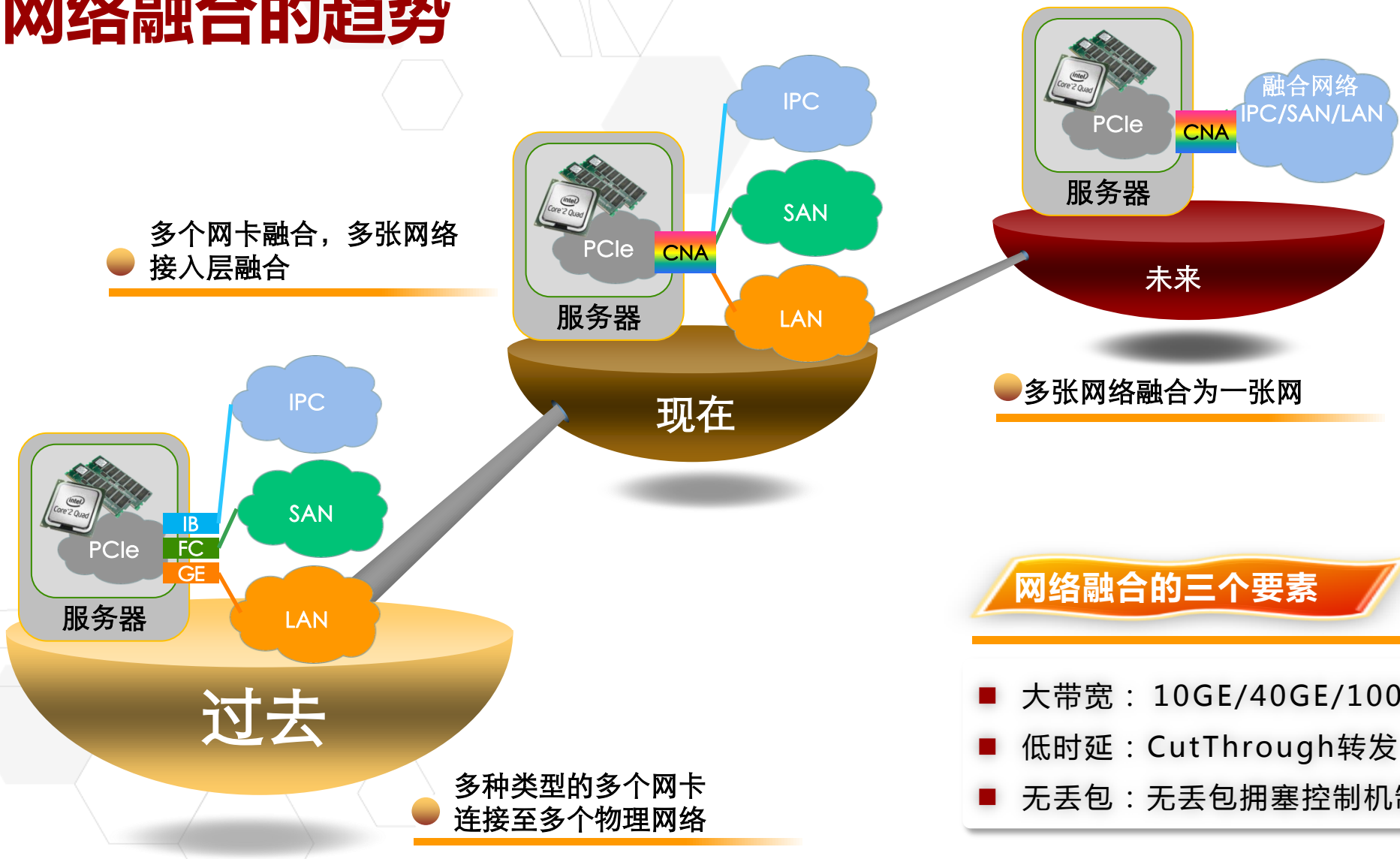
当前数据中心架构的问题

- **网络复杂，**LAN/SAN独立部署，扩展困难
- **能效比低，**服务器上至少配置4~6块网卡，增加功耗

融合网络

- **网络简化，**LAN/SAN融合，统一交换
- **低TCO，**服务器配置CNA融合网卡

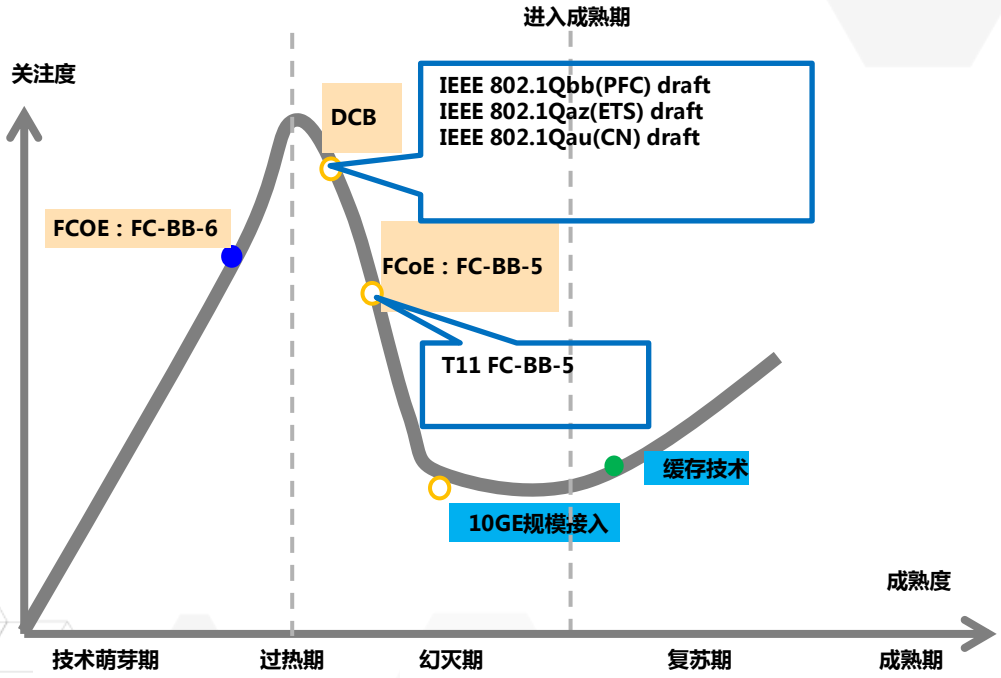
网络融合的趋势



网络融合的三个要素

- 大带宽：10GE/40GE/100GE的以太融合网络
- 低时延：CutThrough转发
- 无丢包：无丢包拥塞控制机制

技术标准进展



10GE规模接入

- **10GE融合网卡,**技术成熟业界已有相关产品推出
- 40GE/100 GE 网络产品已有推出

无丢包增强以太网

- **DCB,**借鉴FC BB-Credit和EE-Credit机制定义以太网基于业务的拥塞反压机制，标准已发布

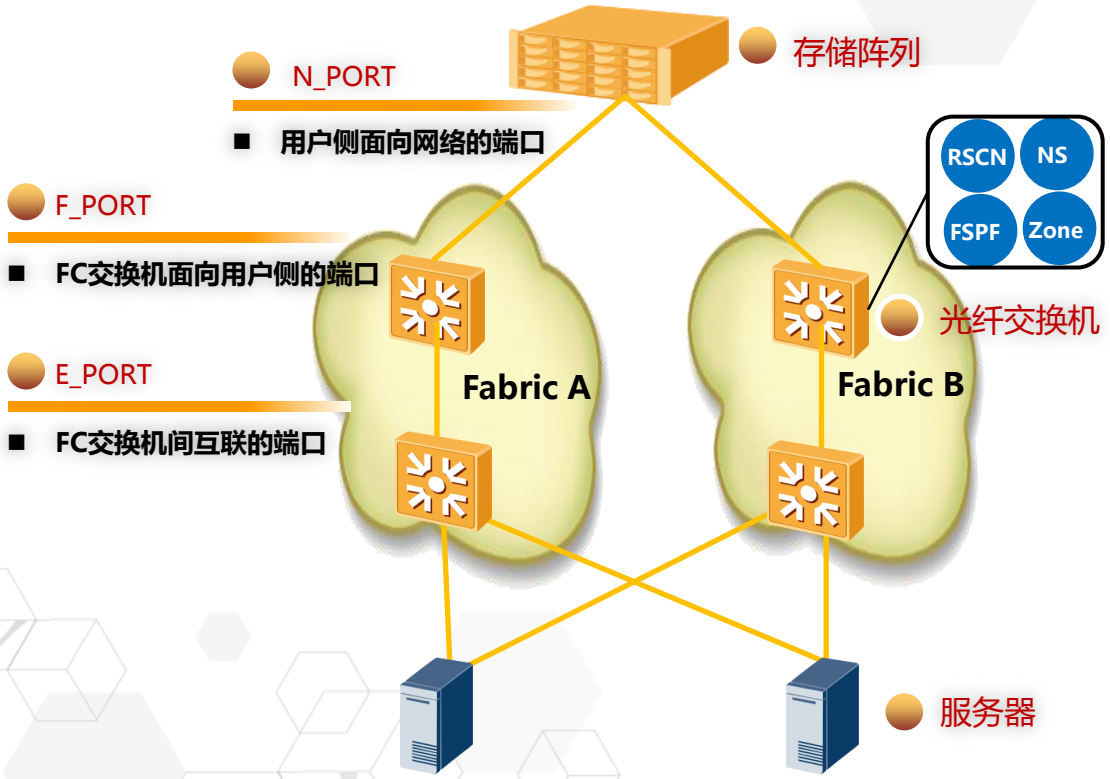
以太融合技术FCoE

- **FC-BB-5,**定义了FC over Ethernet的标准，FC协议在以太中的封装格式、信令控制和转发模型，该标准已发布。
- **FC-BB-6,**优化了本地转发，定义了FDF本地转发实体，标准仍在草案阶段。

从标准状态上看，10GE以太接入，无丢包增强以太网技术已经标准化，以太融合FCoE技术尚在完善中。

目录

FC SAN组网和相关概念



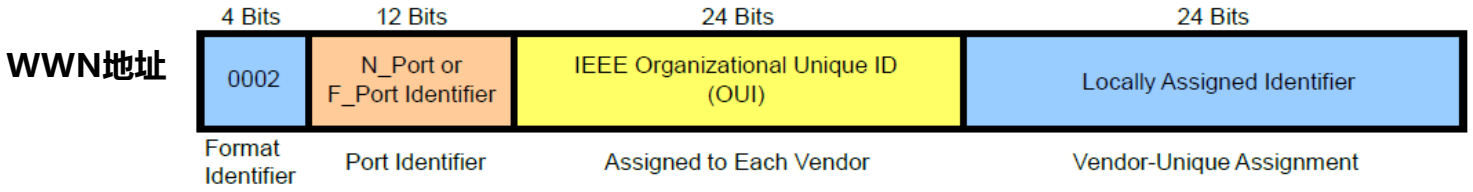
SAN的组网特点

- **双平面保证可靠性**：SAN则采用独立的双平面(双Fabric网络)，由服务器和存储设备，维护两个独立的连接，当任意节点或者链路故障时，由服务器的软件进行切换。
- **终端的自动化注册**：交换机需要向用户（服务器、存储）提供注册，服务，状态维护等功能
- **分布式Fabric服务**：交换机间需交互并维护所在Fabric内的用户状态(RSCN)，名字服务(NS)，路由信息(FSPF), 区域信息（Zone）

FC的编址

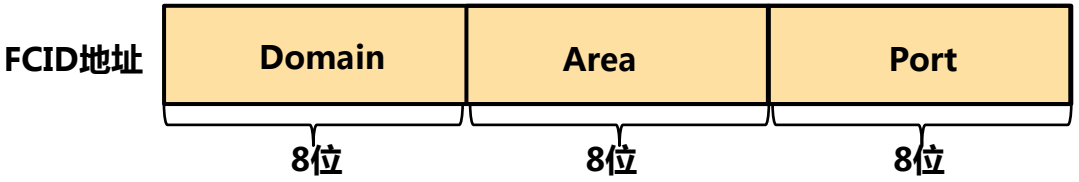
WWN地址, 全球惟一名字

- **WWN，**World Wide Names，（类似以太中的MAC)是由设备商指定, 标识FC网络中所有设备（服务器FC网卡，交换机端口）的64位全球唯一地址。



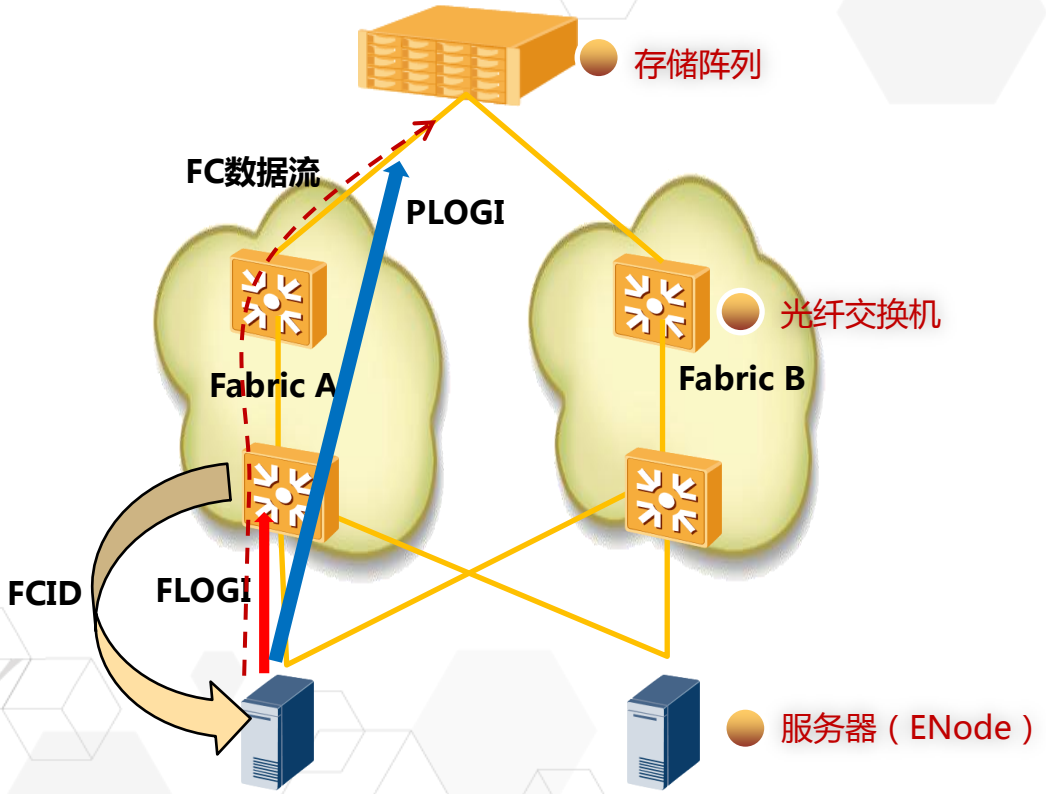
FCID地址，光纤通道标识

- **FCID,（**类似IP地址**）**24位FCID地址, FC网络在转发中使用的地址。标识每一个用户节点，并由FC交换机在用户注册时动态分配。
- **FCID的组成**，共24位，由**domain**，**Area**，**port**组成，其中Domain 用来标识FC交换机，最多可有236个Domain ID(可类比IP网段掩码)。



WWN的64位地址用来做业务与设备绑定时使用，而FCID主要用在Fabric转发时使用。

FC终端的注册过程



用户注册FLOGI（Fabric Login）

- **FLOGI的用途，**用户在该Fabric网络中上线时，请求FC交换机提供服务，并得到由FC交换机分配的FCID地址。

用户注册PLOGI（Port Login）

- **PLOGI的用途**，用户向目标节点发起的会话请求，会话建立后可访问对端设备。

FC通讯

- 服务器与存储阵列通过PLOGI建立会话后，就可以传输FC数据，传输过程中FC交换机依据两端的FCID进行路径的查找和转发。

FC分布式网络服务（Fabric服务）

NS

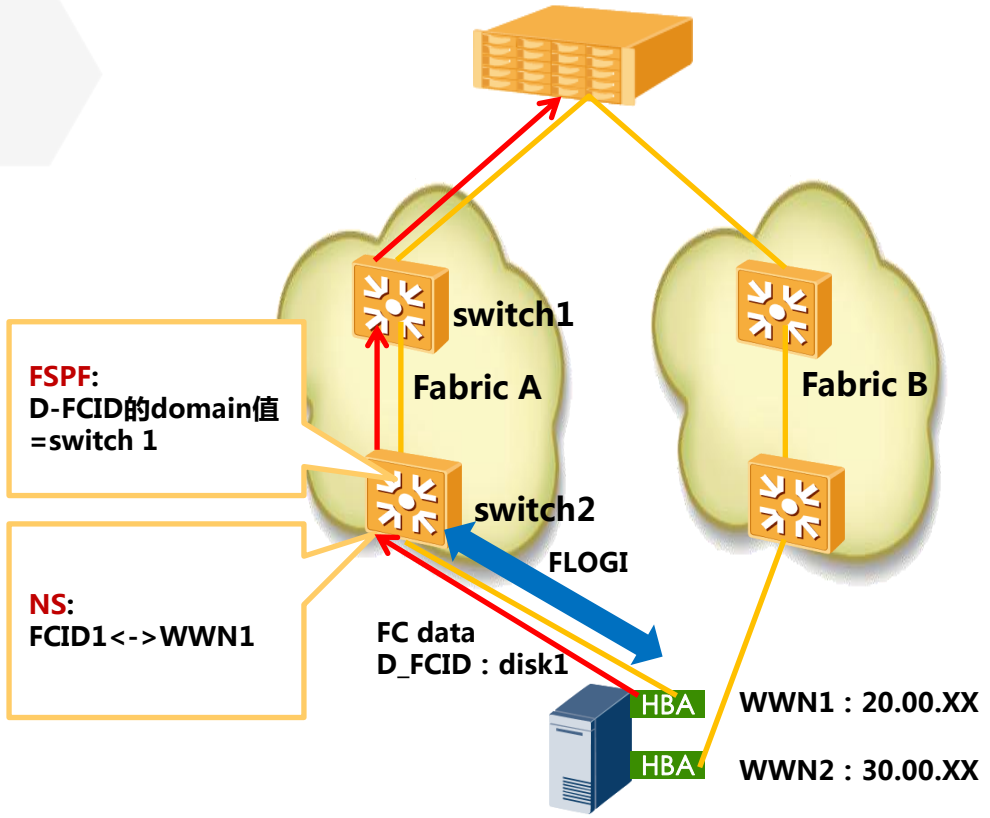
- **Name Service**：FC交换机为用户提供的FCID与WWN映射关系等信息的查询。

FSPF

- **Fabric 最短路径优先转发**：类似OSPF路由协议，针对目的FCID中的domain ID字段，进行选路转发的控制协议。

RSCN

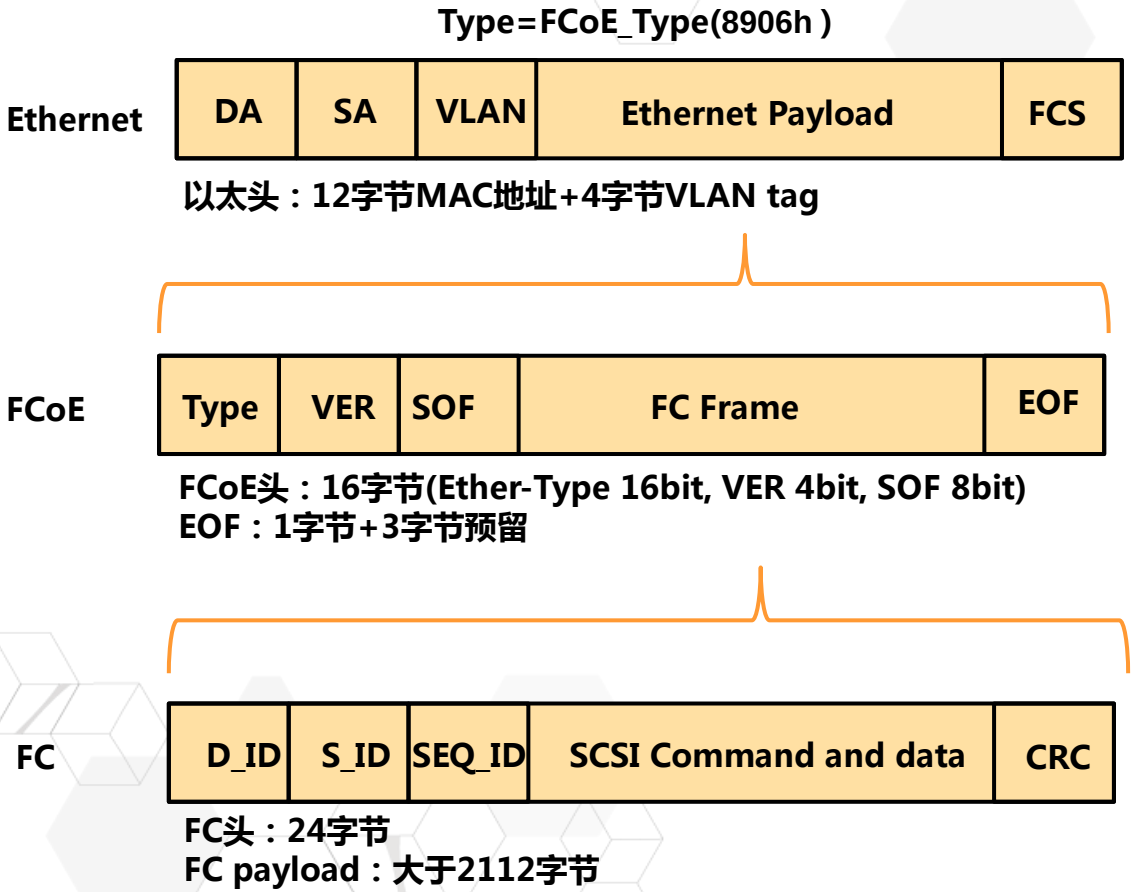
- **用户状态变更通知**：当用户在线状态发生变化时，FC交换机需要通过RSCN将该状态变化通告Fabric网络。



Fabric服务由光纤交换机（或FCoE中定义的FCF）提供分布式的管理和同步，目前不同厂商的Fabric服务的同步协调并不理想。

目录

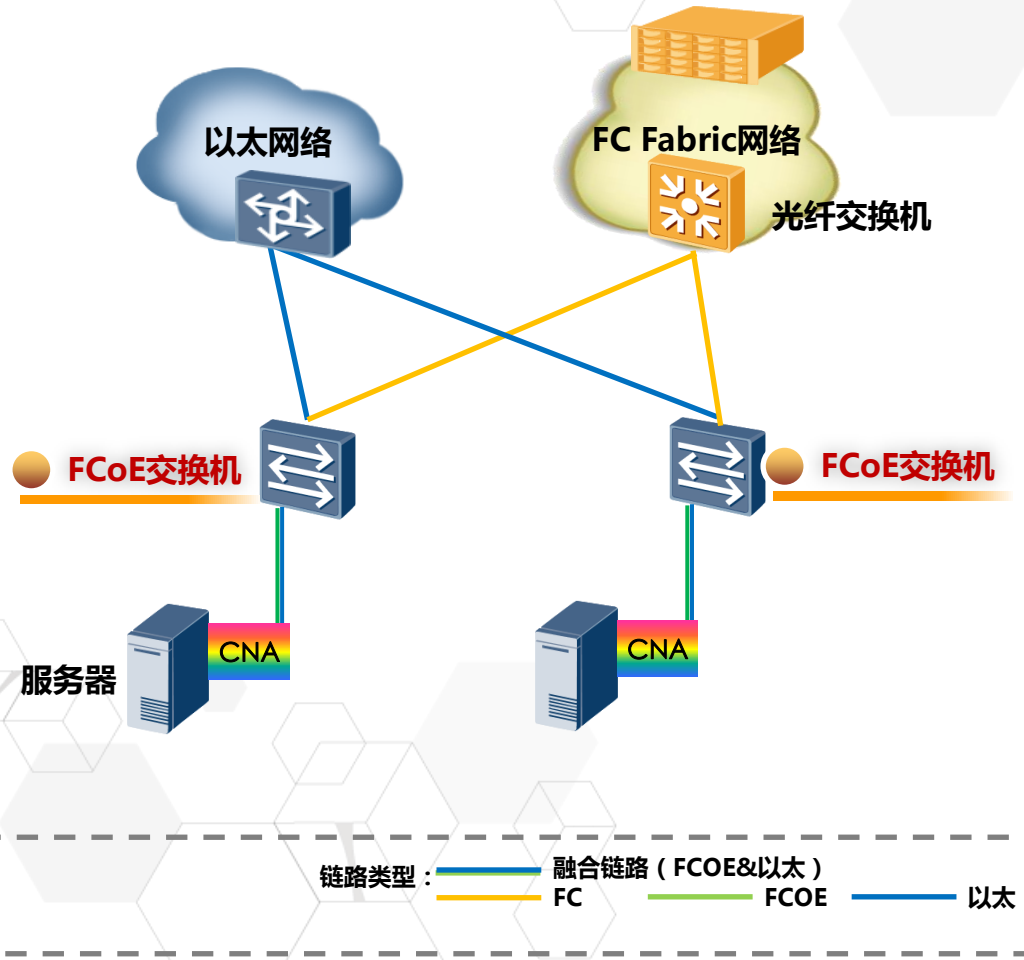
FCoE的封装



FCoE报文格式

- DA：目的MAC，单播为下一跳FCF的MAC，组播为保留MAC。
- SA：源MAC，为每一跳FCF的MAC或终端MAC。
- VLAN：FIP协议中指定 VLAN或FCoE数据VLAN。
- Type：以太类型，取值为8906h时，对应的是FCoE报文。
- FCS：以太帧校验。
- VER:版本号
- SOF:开始标志
- EOF:结束标志
- D_ID：目的FCID地址。
- S_ID：源FCID地址。
- SEQ_ID：Sequence号。

FCoE组网和FCoE交换机介绍



FCoE组网

- 服务器通过融合网卡接入到FCoE交换机。
- FCoE交换机对以太业务和FC业务进行分流。
- 通过FIP协议，由交换机完成用户的FCoE的初始化，分配FCID和MAC地址。

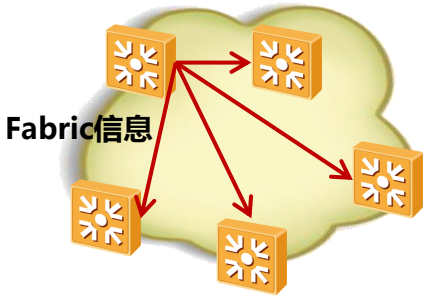
FIP协议

- **FCoE初始化协议，**为申请上线的用户分配MAC,VLAN.

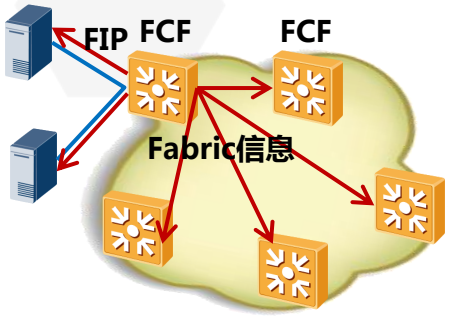
FCoE交换机的分类

- **分类**：按照不同功能分为FCF,NPV,FDF,FSB。其中FCF为FCoE组网的核心组件，提供FIP协议，Fabric服务等关键功能。

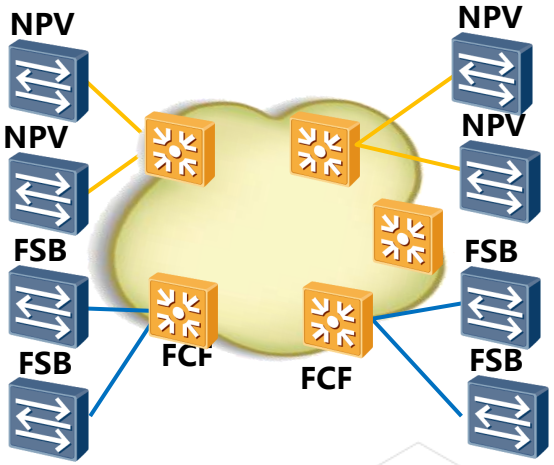
FCoE 网络的演进



过渡到FCoE



网络规模扩展



传统FC规模和性能问题

- **Domain ID的限制**：每台光纤交换机占用1个domain ID, 限制了FC SAN网络的交换节点理论上不大于236个。
- **Fabric数据的同步带来的性能问题**：FC交换机间需要同步用户信息，路由信息，交换节点越多同步的工作量越大，整网的性能越低。

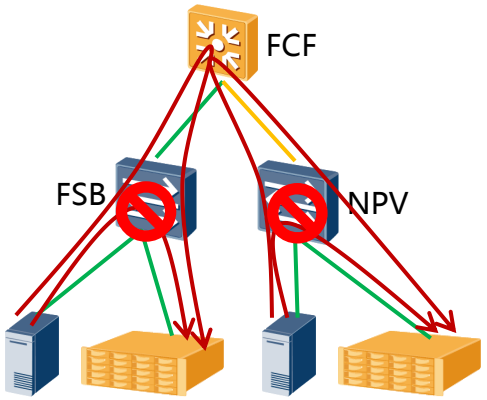
FCF的功能

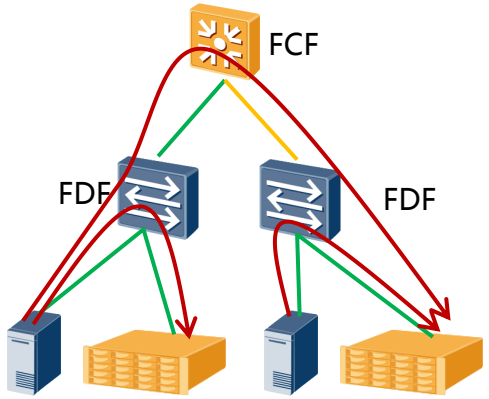
- **提供了完整的FCoE功能**：通过FIP协议对用户进行FCoE初始化并分配FCID地址和MAC地址。
- **继承了光纤交换机的缺点**：FCF能够提供与光纤交换机类似的Fabric服务，并响应用户的接入注册请求，同时可以与传统的光纤交换机互通。但这些功能会占用Domain ID并同步Fabric数据。

丰富FCoE接入技术，扩展网络

- **网络规模扩大**：Fabric网络外围增设多种接入交换机。
- **NPV**：提供FC端口接入，但不参与Fabric服务，也不占用Domain ID。
- **FSB**：提供以太端口接入，不参与Fabric服务，也不占用Domain ID。

FCoE转发路径的优化





FSB，NPV对FCoE网络扩展时存在回绕流量

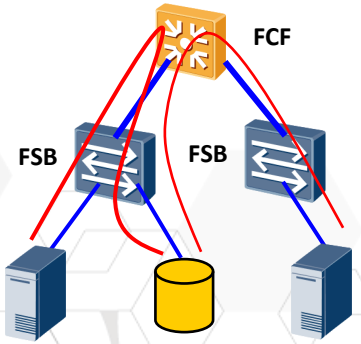
- FC业务基于FCID转发，只能由FCF或光纤交换机完成。
- FSB或NPV本地直连的服务器和存储设备，无法直接通信，必须经过FCF或光纤交换机转发。

FDF对FCoE网络扩展时支持本地转发

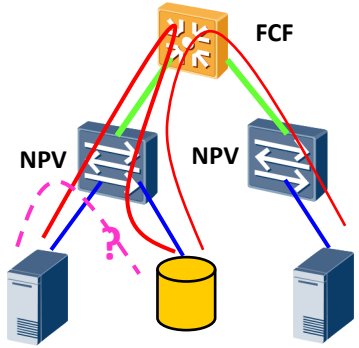
- FDF不占用domain ID，通过FDF对Fabric网络扩展。
- FDF支持本地转发，非本地直连的用户仍需要通过FCF进行转发。

FCoE交换机概览和对比

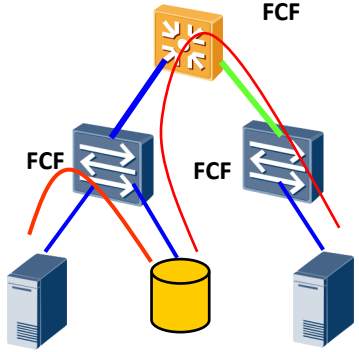
	支持传统FC	转发机制	最优路径转发	提供FC 分布式Fabric服务	是否使用Domain ID	扩展性	复杂性
FSB	否	MAC	否	否	否	好	Low
NPV	是	FCID	否	否	否	好	medium
FCF	是	FCID	是	是	独占	差	High
FDF	是	FCID	是	否	共享	好	Medium



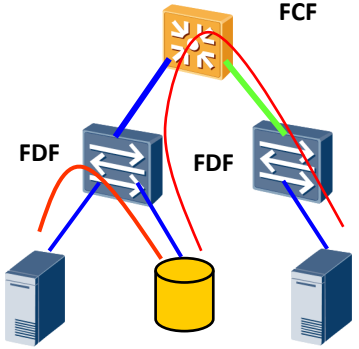
FCF+FSB组网的情况
无路径优化



FCF+NPV组网的情况
无路径优化

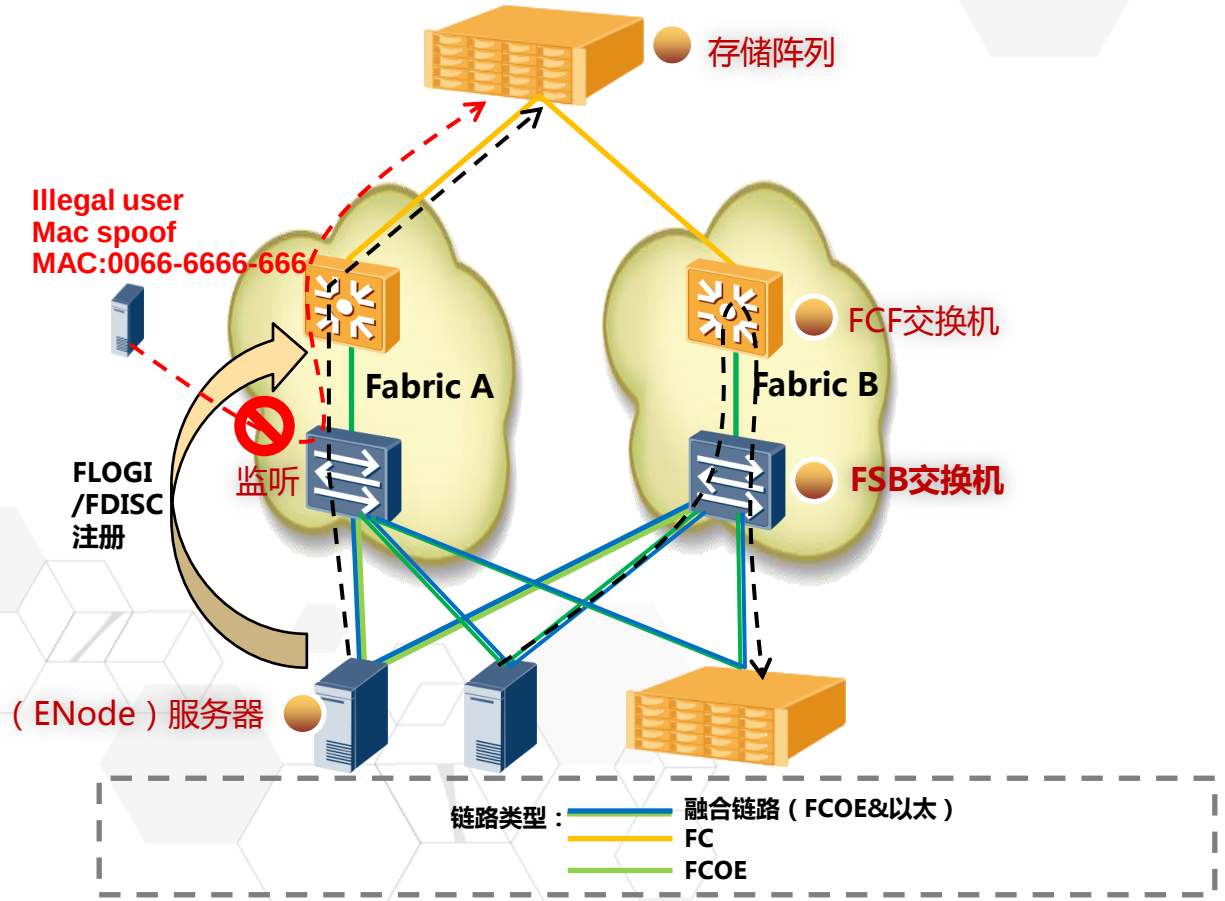


FCF组网的情况
路径优化



FCF+FDF组网的情况
路径优化

FCoE交换机：FSB（FIP Snooping Bridge）



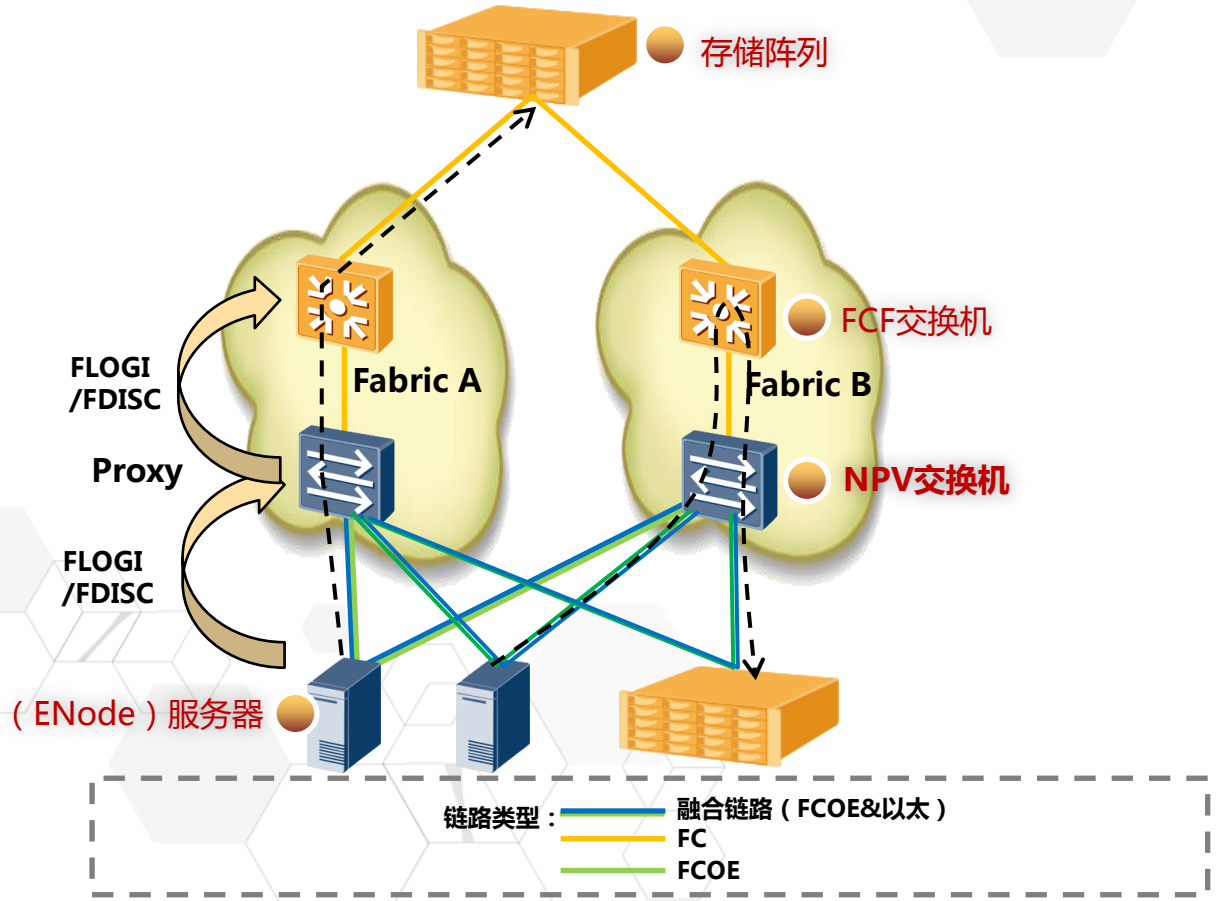
FSB功能介绍

- FSB（FIP Snooping Bridge）FIP监听桥
- **控制面**：不参与FIP控制协议，但监听FIP消息，并依据监听的消息控制链路的访问权，增强安全性。
- **转发面**：通过MAC进行转发。
- **作用**：提供10GE FCoE融合接入并进行以太透传。

FSB的优缺点

- **优点**: 部署简单，基于MAC转发，效率高。
- **缺点**：不能独立组网，仅支持FCoE端口，不能兼容FC网络接入，从Source到Target的存储流量必须经过FCF，流量回绕。

FCoE交换机： NPV（N_Port Virtualizer）



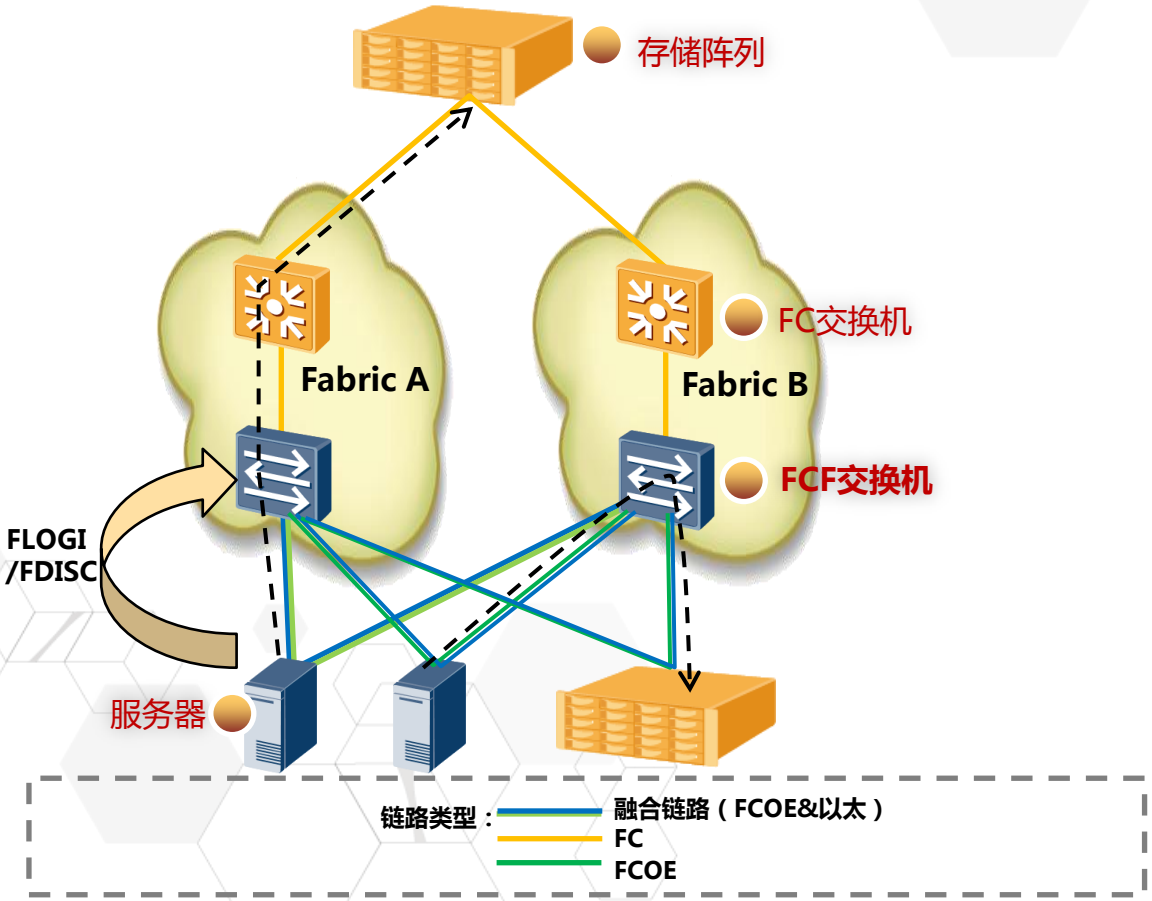
NPV功能介绍

- NPV（ N_Port Virtualization ）N_Port代理。
- **控制面**：部分参与FIP控制协议，代理服务器侧的FIP请求向。
- **转发面**：通过FCID进行转发。
- **作用**：提供FCoE网络融合接入，并且通过提供的FC端口支持与传统FC网络对接。

NPV的优缺点

- **优点**：支持与传统FC网络对接
- **缺点**：无法独立组网需要与FCF配合，流量回绕。

FCoE交换机： FCF（ FC Forwarder ）



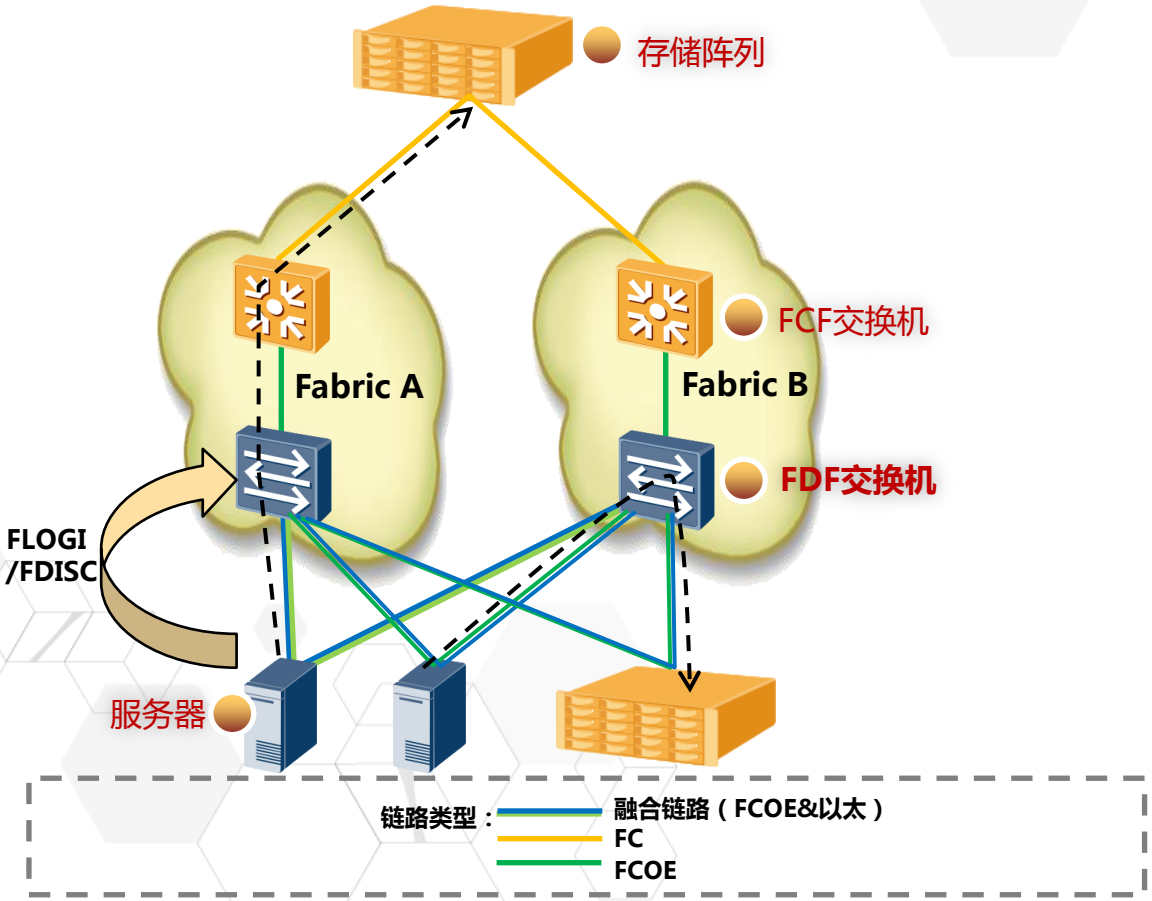
FCF功能介绍

- FCF（ FC Forwarder ）FC转发器
- **控制面**：提供FIP注册服务，并向所在Fabric网络提供NS（名字服务），RSCN（用户状态变更通知），FSPF（FC最短路径转发）。
- **转发面**：通过FCID进行转发，支持本地转发。
- **作用**：提供FCoE初始化功能（FIP），并提供Fabric服务。

FCF的优缺点

- **优点**：支持与传统FC网络对接，能够独立组网。
- **缺点**：受限于domain ID的规格，限制了网络规模。

FCoE交换机： FDF（ FCoE Data Forwarder ）



FDF功能介绍

- FDF（ FCoE Data Forwarder ）FCoE数据转发器
- **控制面**：部分参与FIP协议，但不提供Fabric服务。
- **转发面**：通过FCID进行转发，并支持本地转发。
- **作用**：提供FCoE接入，并且通过提供的FC端口支持与传统FC网络对接，同时支持本地转发。

FDF功能介绍

- **优点**：支持与传统FC网络对接，可本地转发，不存在流量回绕。
- **缺点**：不能独立组网。

FDF目前在FC-BB-6有相关定义，但尚未标准化

目录

以太技术的完善

传统以太

- **拥塞控制**：基于TCP的丢包重传以及滑动窗口机制和于目的端的流控
- **Pause机制**：以太现有技术，通过Pause帧通告下游停止发送。

传统FC

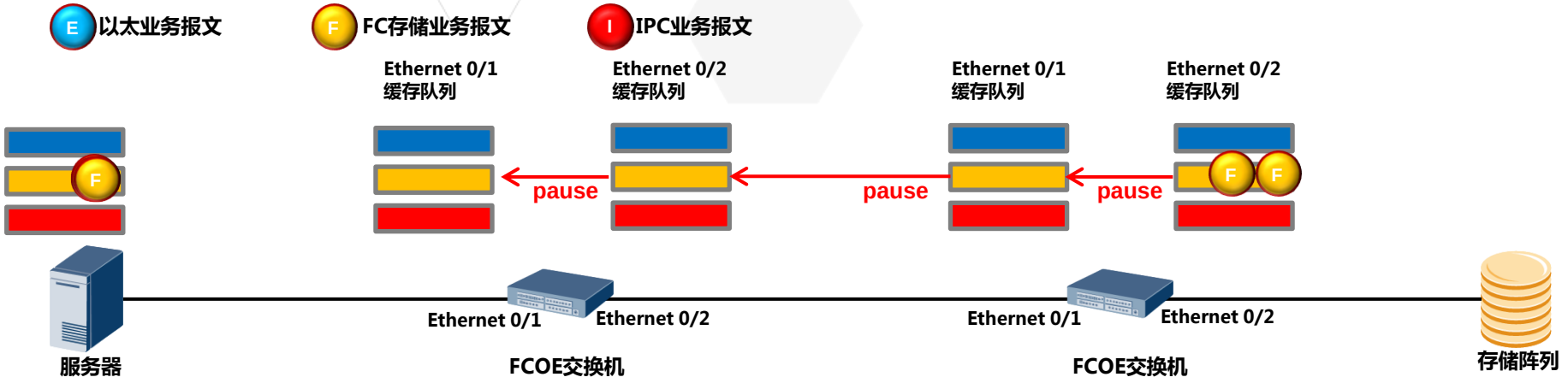
- **拥塞控制**：基于传输节点间BB_Credit和端到端EE_Credit的流量拥塞控制(类似令牌机制，针对带宽传输能力进行传输节点间的流控)。

增强以太网

- **拥塞控制**：利用以太Pause机制，并参考FC的流控实现方法，定义了基于PFC,ETS的逐级流量拥塞控制，和基于CN的端到端流量拥塞控制机制。

存储业务要求在网络传输中不丢包，传统的FC通过Credit机制来保证，而传统的以太技术只能在传输中做到“尽力而为”，通过引入DCB技术，改进以太网在上述方面的不足，为FCoE融合业务提供了传输质量保证。

PFC(Priority Flow Control)



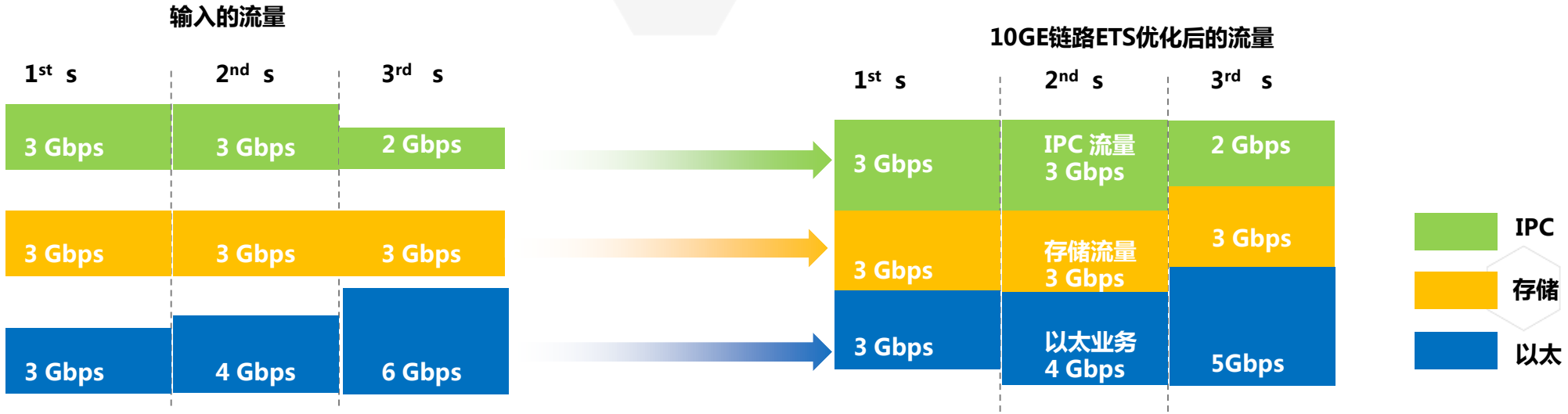
PFC业务识别

- 网络规划设计时，预先定义不同业务的优先级，并为不同的业务设定预定队列阈值。
- 基于业务优先级（0~7）的拥塞控制，超过既定阈值发生拥塞时不影响其他业务的正常处理和转发。

逐级反压

- 通过以太Pause机制，在发生拥塞时，针对该类业务向下游发送反压信号。

ETS(Enhanced Transmission Selection)



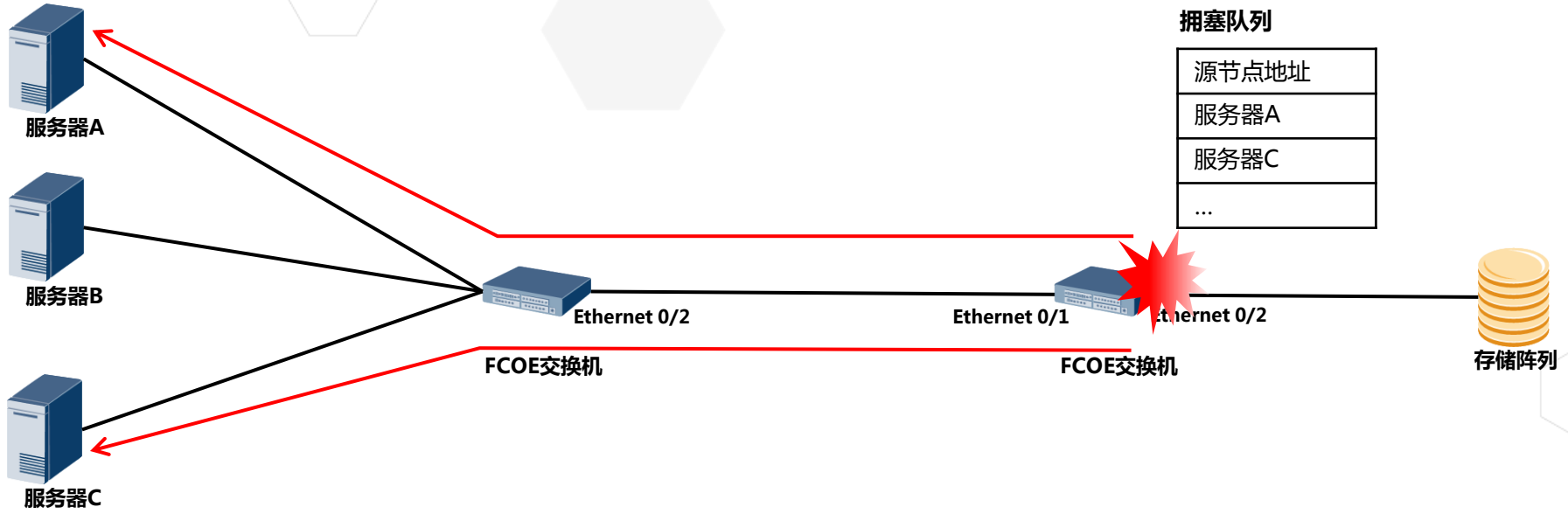
ETS业务识别

- 网络规划设计时，预先定义不同业务的优先级，并为不同的业务设定调度方式和带宽。
- 对于时延高的IPC业务设定优先级调度，对SAN,LAN设定轮询调度。

配合CN,PFC反压机制

- 低时延、高可靠类业务比如IPC,SAN，在超出预设带宽时造成的队列拥塞，通过PFC或CN向下游反压减缓流量压力。
- 对于时延，丢包不敏感的LAN业务，不进行反压，由TCP/IP保证业务重传。

CN(Congestion Notification)



CN实现原理

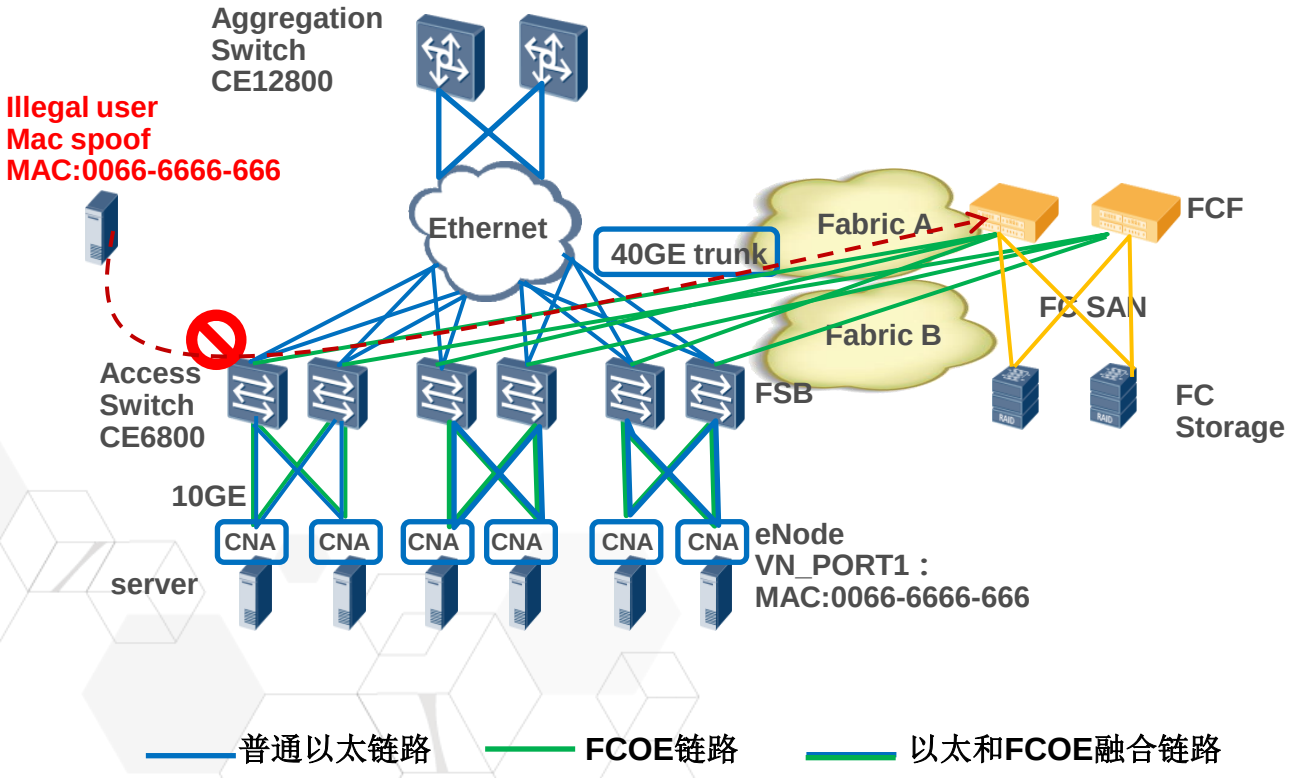
- **实现原理**：发生拥塞时，从拥塞队列中查找到引发拥塞的报文源节点，直接向源节点发送拥塞通知，直到拥塞解除。

CN的特点

- 直接命中引发拥塞的源节点，阻止其继续向网络发送报文。

目录

华为接入融合的FCoE方案



高可靠

- **双Fabric平面**：通过服务器双CNA接入，实现双Fabric平面，保证存储可靠性。
- **多以太链路**：通过传统的以太链路捆绑技术实现单Fabric网络内多链路可靠性保证。

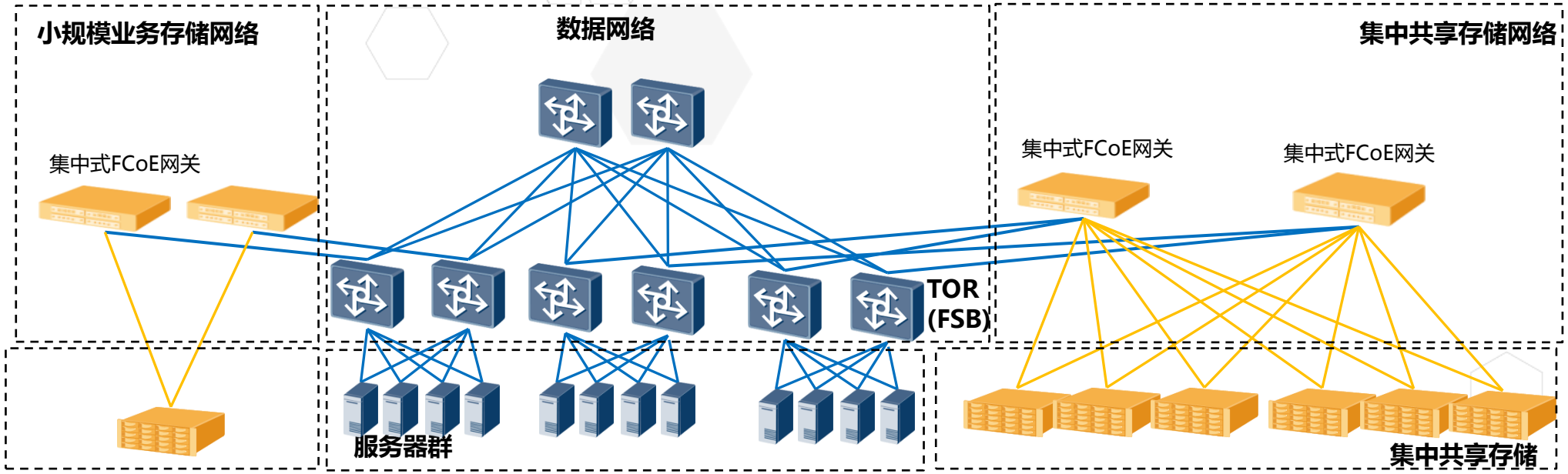
无丢包

- **DCB**：通过在服务器->FSB->FCF间部署DCB，满足FC存储业务在以太网络中传输的无丢包需求。
- **10GE接入**：新一代数据中心万兆接入交换机融合后多业务的接入带宽需求。

高安全

- **FSB**：通过监听FIP消息，仅向合法用户开放链路。

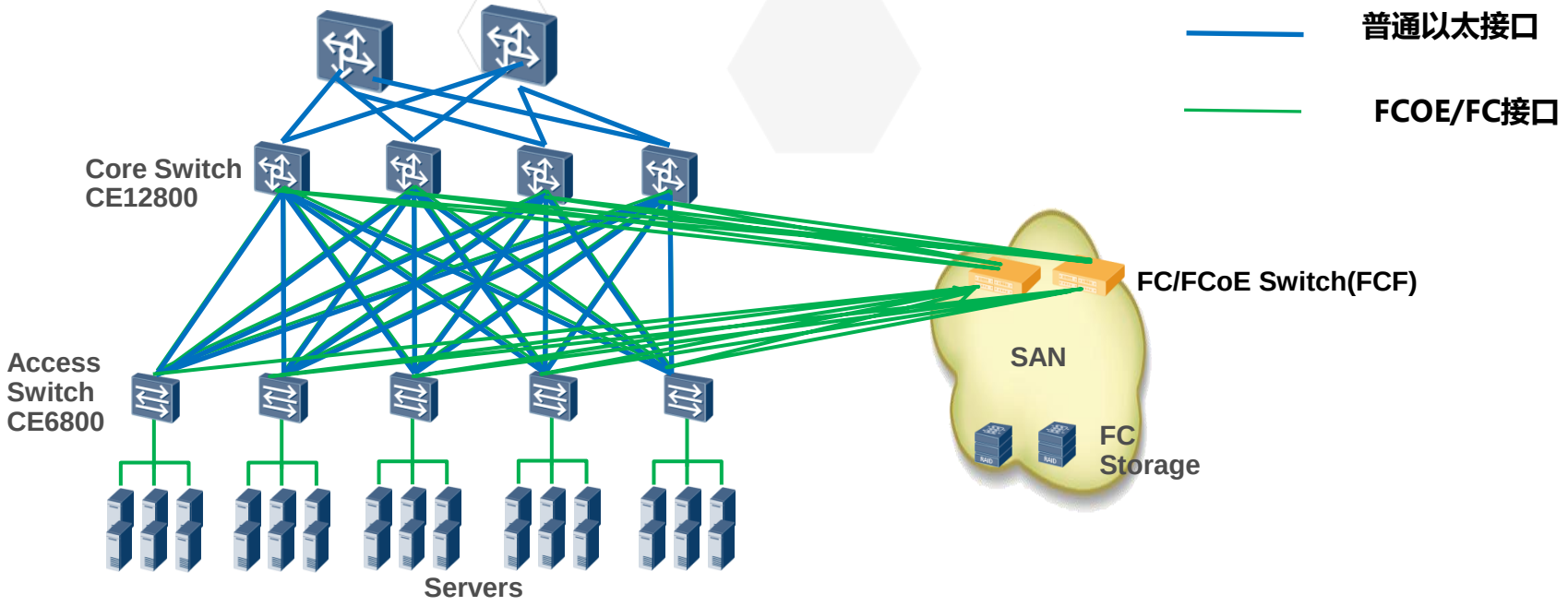
集中式FCoE网关



集中式FCOE网关位置灵活部署

- 小规模业务部署：集中式FCoE网关可根据业务的实际需要，就近部署在业务相关的服务器列头柜。
- 集中式部署：如果存在全网集中共享存储的业务需求，集中式FCoE网关可集中部署在存储汇聚区域。

华为融合网络演进思路



接入存储融合

- TOR支持FCOE（V1R1:FSB）/FC(V1R2:FCF/NPV）接口和DCB无丢包以太网。
- 存储流量在TOR直接分流到SAN存储网络。

演进

核心存储融合

- TOR/核心都支持FCOE/FC接口和DCB无丢包以太网。
- 存储流量在核心分流到SAN存储网络。

Huawei Enterprise A Better Way

附录：技术缩略语

Abbreviations缩略语	Full spelling 英文全名	Chinese explanation 中文解释
FC	Fiber Channel	光纤通道
FCOE	Fiber Channel Over Ethernet	FC在以态 上承载
FIP	FCoE Initialization Protocol	FCOE的初始化协议
FSB	FIP Snooping Bridge	FIP的侦听桥
NPV/NPIV	N_Port Virtualization/ N_Port ID Virtualization/	N端口虚拟化，本文NPV指使用NPIV技术提供FC接入代理的交换机
SAN	Storage Area Network	存储区域网络
FCID	Fiber Channel ID	光纤通道标识
WWN	World Wide Name	全球惟一名字
FCF	FC Forwarder	FC的转发器
eNode	Ethernet Node	支持FCOE的节点
NS	Name Service	名字服务
FSPF	Fabric Shortest Path First	FC网络最短路径优先路由协议
RSCN	Registered State Change Notification	注册状态变更通知