

T/ISC

团 体 标 准

T/XXX XXXX—2026

全光数据中心 OCS 组网架构设计与场景应用指南

Guidelines for All-Optical Data Center OCS Networking Architecture
Design and Scenario Applications

征求意见稿

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

2026 – XX – XX 发布

2026 – XX – XX 实施

中国互联网协会 发布

目 次

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 符号和缩略语 1

5 架构设计模型 2

6 架构设计原则 2

7 架构设计技术要求 3

8 架构关键设计流程 4

9 场景类型 5

10 实现要求 7

参考文献 9

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由××××提出。

本文件由中国互联网协会归口。

本文件起草单位：

本文件主要起草人：

全光数据中心 OCS 组网架构设计与场景应用指南

1 范围

本文件规定了全光数据中心OCS组网架构的设计模型、设计原则、设计技术要求、设计流程、场景应用要求及实现要求。

本文件适用于全光数据中心OCS组网的总体架构设计与应用场景。

本文件不涉及OCS设备内部光交换元器件的具体设计与制造要求。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 28827.4—2019 信息技术服务 运行维护 第4部分：数据中心服务要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据中心 data center

由计算机场地（机房），其他基础设施、信息系统软硬件、信息资源（数据）和人员以及相应的规章制度组成的实体。

[来源：GB/T 32910.1—2017，2.1]

3.2

Pod

数据中心内由计算节点（如GPU/NPU服务器）、存储设备和网络互连设备组成的一个逻辑或物理集群单元，作为资源调度和网络扩展的基本模块。多个Pod通过OCS或电交换网络互连形成更大的算力集群。

3.3

全光数据中心 all-optical data center

以光互连技术为核心，在数据中心内部或数据中心之间大量采用光交换、光传输等技术，并支持光电协同工作的数据中心网络架构。

3.4

网络重构 network reconfiguration

通过OCS设备在光域直接动态调整光信号的物理传输路径，实现网络拓扑结构、连接关系或光路映射的重新配置的过程。

3.5

透明传输 transparent transmission

在网络传输过程中，不对承载信号的协议、帧结构或编码方式进行解析或处理，直接以物理层信号形式进行转发和传输的能力。

3.6

资源池化 resource pooling

通过对计算、存储、网络等资源进行统一管理和调度，实现资源的解耦、共享和按需分配的一种系统组织方式。

4 符号和缩略语

下列缩略语适用于本文件。

AI: 人工智能 (Artificial Intelligence)
 DCI: 数据中心互连 (Data Center Interconnect)
 DCNI: 数据中心网络互连 (Data Center Network Interconnection)
 GPU: 图形处理器 (Graphics Processing Unit)
 JCT: 任务完成时间 (Job Completion Time)
 MoE: 混合专家模型 (Mixture of Experts)
 NPU: 神经网络处理器 (Neural Processing Unit)
 OCS: 光路交换 (Optical Circuit Switching)
 OTN: 光传送网 (Optical Transport Network)
 OXC: 光交叉连接 (Optical Cross-Connect)
 PCIe: 高速外设互连总线 (Peripheral Component Interconnect Express)
 PUE: 电能利用效率 (Power Usage Effectiveness)
 WDM: 波分复用 (Wavelength Division Multiplexing)

5 设计模型

5.1 总体逻辑架构

为提升架构设计的系统性，全光数据中心OCS组网架构设计宜按照分层组织。典型的架构层次包括但不限于以下内容。

- a) 物理资源层：包括计算资源、光交换设备、电交换设备及相关链路资源，是 OCS 组网架构运行的基础支撑层，主要关注资源的类型、规模和互连能力。
- b) 网络与互连层：用于描述光交换与电交换协同构建的网络结构，关注网络拓扑组织方式、互连模式及光电协同关系，为上层业务提供高带宽、低时延的互连能力。
- c) 管控与调度层：对物理资源进行统一管理和协同调度，支持网络状态感知、资源配置、拓扑调整及策略执行。
- d) 业务与应用层：面向数据中心的具体应用场景，描述业务需求对网络性能、连接方式和调度能力的要求，为架构设计提供需求牵引和优化目标。

5.2 物理平面模型

为明确OCS在全光组网架构中的部署位置，可以多级Clos拓扑和3D Torus架构为基础，根据业务需求采用以下典型部署模型。

- a) 单元间 (Inter-Pod) 互连部署：以多级 Clos 拓扑为基础架构时，OCS 宜部署于核心层，用于支撑跨 Pod 算力集群的动态互连，实现计算资源的弹性池化。
- b) 单元内 (Intra-Pod) 或机架直连部署：以多级 Clos 拓扑为基础架构时，OCS 宜部署于接入层或直接连接计算节点，提供确定性物理传输平面；以 3D Torus 拓扑为基础架构时，OCS 宜部署于物理机柜 (Cube) 之间，突破电信号在长距离传输上的瓶颈。
- c) 边界与传输节点部署：OCS 宜部署于数据中心出口边界或跨地域传输路径的物理转发节点。该位置用于承载跨中心/跨地域的物理链路调度，实现光层信号的透明转发。

5.3 控制平面模型

控制平面宜实现对包括电交换节点、电信号链路、光交换节点、光链路等在内的光、电异构资源的统一编排，并符合下列要求。

- a) 通用配置接口要求：宜采用基于模型驱动的统一管理接口，实现对 OCS 物理实体的统一资源抽象与配置封装，屏蔽底层硬件差异。
- b) 拓扑与路由同步：应具备物理链路状态与逻辑网络拓扑的关联映射功能，在光层路径重构过程中支持物理连接关系与网络层拓扑信息的同步更新，确保全光组网环境下逻辑路由的可达性与拓扑状态的一致性。

6 设计原则

开展全光数据中心OCS组网架构设计应遵循以下设计原则。

- a) 深度融合原则：架构设计应实现计算并行模式、存储 I/O 特征与光层拓扑重构能力的协同，确保物理路径建立与计算任务通信模式相匹配。
- b) 管控与协同原则：架构宜支持对 OCS、电交换机以及算/存接口的统一编排，实现管控层面的逻辑统一与光电平面的联动。
- c) 灵活重构与迭代演化原则：架构设计应具备软件定义能力，支持拓扑架构随业务需求动态调整，并具备支撑 Fat-Tree、Dragonfly 等多样化拓扑演进的兼容性。
- d) 稳定可靠原则：架构应配置完善的保护倒换机制，确保在光路重构或链路故障时，业务层连接具备连续性，尽量避免因网络重构或链路故障触发计算任务超时退出。
- e) 可观测性原则：架构设计宜包含统一监控平台，实现对光物理层性能及网络逻辑层状态的实时采集与可视化管理。
- f) 绿色节能原则：架构宜利用 OCS 静态功耗低的物理特性，通过优化光路径配比降低数据中心网络子系统功耗，进而改善整体 PUE。
- g) 确定性设计原则：架构宜明确界定 OCS 承载流量边界，利用物理层透明传输特性，为高性能任务提供高确定性、低抖动的传输平面，以降低电交换缓存波动对业务性能的影响。

7 设计技术要求

7.1 性能要求

架构设计应确保高带宽透明传输与确定性时延能力，包括以下内容。

- a) 传输一致性：应支持多轨并行传输场景下的光链路延迟对齐与报文保序能力。在带宽并行扩展场景下，应具备光链路层面的时间同步能力，并协同网络协议栈实现报文顺序一致性。
- b) 物理层透明性：应支持多种链路速率的透明物理层转发，应支持调制格式透明，宜兼容 400 G、800 G、1.6 T 等多种链路速率。

7.2 规模要求

架构设计应具备高扩展性，支撑算力集群的平滑演进，并包括以下内容。

- a) 交换规模：应支持多类型交换规模，宜根据应用场景的互连需求配置不同规模的光交换节点。架构设计应根据业务规模、部署层级、互联对象数量及扩展需求选择适当端口规模的 OCS 设备。OCS 可支持小端口、中端口及大端口等多种交换规模；在接入层、Pod 内互联、局部资源池化、测试验证等场景下，可采用较小端口规模的 OCS；在核心层、Pod 间互联或大规模算力集群扩展场景下，宜采用大端口、高密度光交换节点，典型规模可达到 128×128、256×256 或更高。
- b) 集群支撑能力：宜具备支撑大规模算力节点互联的能力，并支持通过多级拓扑实现集群规模的横向扩展。

7.3 灵活性与重构要求

架构应具备按需动态调整物理连接的能力，并包括以下内容。

- a) 拓扑重构：应支持根据业务特征、通信模式及流量特征进行物理拓扑的重构。
- b) 按需分配：应支持对不同功能节点间的互连关系进行动态定义，支持网络资源的按需分配与利用率优化。

7.4 协议兼容性要求

架构应具备多协议承载能力，确保异构系统的接入，并包括以下内容。

- a) 互连协议兼容：应兼容包括以太网、InfiniBand、PCIe、NVLink 等多种主流高速互连协议的物理层传输需求，如高速互连协议的误码率要求。
- b) 业务透明性：物理层架构应与上层协议解耦，确保不同协议栈在同一全光平面内的共存与切换。

7.5 安全与稳定要求

架构应集成物理层保护与逻辑隔离机制，并包括以下内容。

- a) 故障隔离：应提供物理层故障隔离机制，在检测到链路劣化或节点失效后，宜通过光路切换实现拓扑重构，将异常组件从网络中逻辑分隔，避免故障传播；故障恢复后应支持路径重建与连接回切。
- b) 故障自愈：应建立物理层链路监测机制，在识别链路或设备异常后，应支持路径重构与故障恢复。
- c) 多租户隔离：应支持通过管控平台实现虚拟资源单元的划分。架构应提供不同租户或工作负载间的物理层或逻辑层隔离，保障系统运行的安全性。

8 关键设计流程

8.1 业务需求分析

在进行全光组网架构设计前，应开展多维度的业务需求识别，确定架构规模与性能基准，包括但不限于以下内容。

- a) 流量模式识别：应识别业务流的物理拓扑属性、流量突发特征及通信模式。
- b) 性能基准定义：应明确业务层对端到端物理时延、时延抖动及拓扑重构收敛时间的容忍边界。
- c) 空间与规模规划：应评估模块内（Intra—Pod）、模块间（Inter—Pod）及跨地域 DCI 场景下的光纤互连密度与总带宽需求。

8.2 物理拓扑设计

应根据业务规模及物理环境，开展光层拓扑规划与资源选型，包括但不限于以下内容。

- a) 部署位点选型：应确定 OCS 在多级交换架构中的逻辑层级或识别跨地域传输路径中的转发点。
- b) 端口映射编制：应编制 OCS 物理端口与各功能节点及广域网接口的逻辑映射表，并建立统一的端口标识规范。

8.3 仿真验证与环境评估

在系统交付前，应开展全路径的功能性仿真与性能预估，包括但不限于以下内容。

- a) 光链路预算核算：应计算全链路插入损耗。针对跨地域长距路径，应计入光放大、色散补偿及环境损耗余量，确保物理链路质量满足传输要求；对于不大于 2 km 的短距数据中心内部链路，可不计入色散补偿。
- b) 协议一致性仿真：应模拟 OCS 拓扑重构过程对网络路由协议、拥塞控制机制及上层应用协议的影响，验证业务抗抖动能力。
- c) 可靠性模拟：应开展物理层故障模拟，评估跨地域链路在中断情况下的自动保护切换时延与自愈路径的有效性。

8.4 部署实施与交付规程

架构实施阶段应遵循标准化的工程操作流程，包括但不限于以下内容。

- a) 物理链路稽核：应进行光纤端面检测、清洁及损耗实测，确保每一路物理连接符合设计规范。
- b) 控制平面对接：应完成 OCS 与管控平台的南向接口联调，验证数据模型封装与指令下发的准确性。
- c) 光电协同配置：应建立 OCS 拓扑切换与电交换层路由发布的联动机制。在跨地域场景下，应实现 OCS 与传输系统的保护倒换协同。
- d) 高密度布线设计：应定义高密度光缆组件的路由走向。针对跨地域场景，应明确光传输系统的物理连接逻辑。

8.5 运行优化与闭环管理

部署实施与交付后，应建立长效的性能监控与优化机制，包括但不限于以下内容。

- a) 实时状态监测：应利用监控技术对光物理层参数及链路利用率进行实时采集与可视化管理。

- b) 调度算法调优：应根据实际业务负载与任务执行效率，动态优化 OCS 拓扑重构的触发策略与路由策略；应定期对比光重构前后的任务完成时间（JCT）变化，用于调度策略评估。

9 场景类型

9.1 数据中心内互连

9.1.1 超大规模横向扩张（Scale-out）场景

9.1.1.1 场景描述：互连该场景面向万亿级及以上参数大模型训练对算力集群规模持续扩展的需求，特别是在数据并行和流水线并行任务中的通信效率。旨在利用光交换节点的透明传输特性，代替架构层级中的电交换节点或利用大规模光交换节点减少交换层级，解决多级网络高复杂度问题和大量光—电—光转换的高成本高功耗问题。

9.1.1.2 架构设计：在该场景下，引入 OCS 作为数据中心网络互连（DCNI）的关键网络设备，在 Spine/Super-Spine 层作为电交换机的补充或局部替代，形成光电混合调度平面。

9.1.1.3 典型架构：应采用基于多级交换矩阵的光电混合架构（如 Spine-Leaf 或 Clos 拓扑等）作为基础物理架构，将 OCS 部署于核心调度层（Spine/Super-Spine），以实现 Pod 间的物理层直达互连。在考虑典型实现架构宜参考 FlexFly、ReFly 等具备动态重构能力的扁平化全光拓扑方案，其中，FlexFly 架构通过 OCS 实现高基数、低直径的网络结构，适用于大规模算力集群的横向扩展；ReFly 架构通过基于动态环分解的 OCS 动态配置构建环状逻辑拓扑，优化集体通信模式，提升有效总线带宽，为各类 AI 模型训练（包括一般密集模型及 MoE 等稀疏模型）提供了高效的网络基础互连能力。

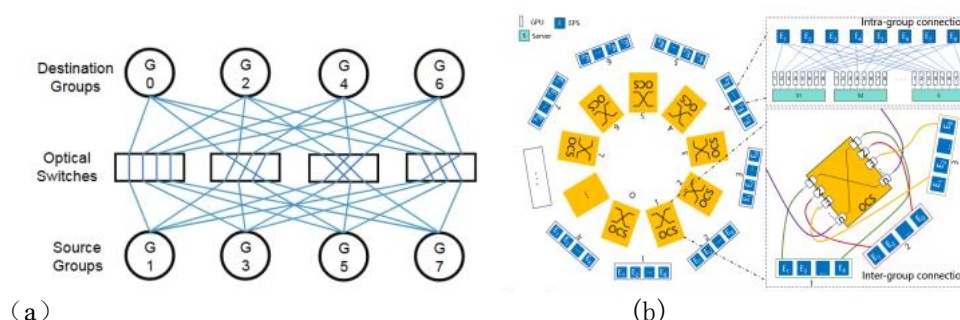


图1 超大规模横向扩张（Scale-out）场景典型架构（a）FlexFly (b) ReFly

9.1.2 高性能纵向扩展（Scale-up）场景

9.1.2.1 场景描述：该场景面向超级节点（Superpod）内 GPU/NPU 的高频通信需求，特别是在张量并行和专家并行任务中的通信效率。旨在利用光交换节点的可重构和低损耗传输的物理特点，代替 Pod 内的传统电交换机，突破跨机柜电信号在功耗、距离、规模等方面的物理局限问题。

9.1.2.2 架构设计：在该场景下，互连引入 OCS 作为数据中心网络互连（DCNI）的关键网络设备，在 Leaf/Access 层作为电交换机的补充或替代，形成光电混合调度平面。

9.1.2.3 典型架构：应采用多级交换矩阵，3D Torus 等架构作为基础架构，面向多级交换矩阵架构可将 OCS 部署于接入层（Leaf/Access），面向 3D Torus 架构，可将 OCS 部署于物理机柜之间。典型架构宜参考如 mFabric 或 Ironwood Superpod 架构等。其中，mFabric 利用光交换节点可重构特性提升 MoE 模型的训练成本效率，Ironwood Superpod 利用光交换节点突破电交换的端口密度限制，提升 TPU Cube 的互连规模。

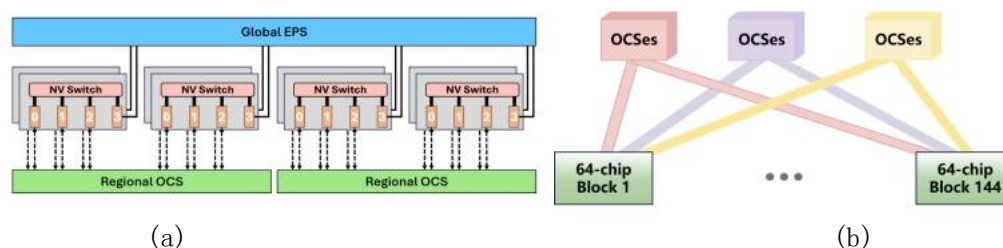


图2 超大规模横向扩张（Scale—out）场景典型架构（a）mFabric（b）Ironwood Superpod

9.2 数据中心间互连

9.2.1 跨地域光路直连算力调度场景

9.2.1.1 场景描述：该场景面向智算资源动态调度需求，支撑算力资源在不同数据中心、不同资源池之间的快速匹配与协同使用。利用 OCS 的光路径快速重构能力，确保算力流转在需求时间内完成路径建立，支撑构建自主可控的算力网络生态。

9.2.1.2 架构设计：通过 WDM、OXC 及 OCS 网元构建全光底座。建立智算中心间的高速直接连接，实现跨地域算力资源的实时感知与按需调度。

9.2.1.3 典型架构：应采用基于全光调度的广域互连架构作为基础架构，将 OCS 部署于数据中心出口网关或跨地域传输节点，实现算力资源在物理层面的跨池化匹配。典型实现架构宜参考如波长级动态重构拓扑或基于全光交换的算力承载网方案。

9.2.2 跨地域链路灵活复用场景

9.2.2.1 场景描述：该场景面向多业务融合弹性调度需求，支撑 AI 大模型训练、推理请求及通用云业务等在跨地域环境下的灵活复用与按需分发。利用 IP 层的 L3 统计复用与精细化流控能力和光层的高吞吐传输能力，确保多业务并行按需传输。

9.2.2.2 架构设计：通过路由器（IP 层）与 WDM/OTN 等（光层）的深度协同构建承载底座，实现不同粒度算力任务的逻辑寻址与择优转发，在保障高吞吐流量的同时，最大化跨地域链路的带宽利用率。

9.2.2.3 典型架构：应参考 IETF RFC 8453 定义的流量工程网络抽象与控制框架，通过在光传送层部署基于 OCS 的全光交叉调度节点，构建 IP 路由平面与全光物理平面的协同编排体系。



图3 跨地域链路灵活复用场景

9.3 数据中心调度

9.3.1 故障自动规避与容灾场景

9.3.1.1 场景描述：该场景面向长周期 AI 训练任务和高性能计算任务的稳定运行需求，提升网络在故障条件下的可靠性和业务连续性。

9.3.1.2 架构设计：在该场景下，OCS 可利用其物理层可重构能力，通过预配置或快速重构光连接，在节点或链路发生故障时实现互连路径的快速切换，并与电交换网络协同完成故障规避和业务恢复。该机制有助于缩短故障恢复时间，降低故障对计算任务的影响，提高网络系统的容错能力。

9.3.1.3 典型架构：应采用基于全光物理层快速自愈的冗余互连架构作为基础物理架构，将 OCS 部署于核心交换节点或跨地域网关层，作为物理层面的保护切换中心。该架构通过感知物理链路实时损耗，实现对算力集群关键路径的毫秒级故障规避。

9.3.2 资源池化场景

9.3.2.1 场景描述：该场景面向数据中心资源统一管理和按需使用需求，通过网络层面的灵活互连支撑计算、存储等资源的解耦与动态组合。

9.3.2.2 架构设计：在该场景下，可利用 OCS 构建透明的光交换矩阵，实现计算资源（如 GPU）与存储资源之间的物理动态挂载。根据任务负载特征和资源需求情况，对互连拓扑进行按需调整，实现资源的灵活组合与弹性拆分，从而提升整体资源利用效率和系统运行灵活性。

9.3.2.3 典型架构：应采用基于资源池化的弹性组网架构作为基础物理架构，将 OCS 部署于计算、存储与加速资源池的公共交换平面，实现各异构算力组件间的直接物理映射。该架构通过动态定义逻辑链路，支持硬件资源按需编排与池化共享。

10 实现要求

10.1 一般要求

数据中心运行维护管理除符合GB/T 28827.4—2019的要求外还宜满足下列要求：

- a) 全光数据中心 OCS 组网应与现有电交换网络协同部署，支持混合组网模式，并具备与主流数据中心网络架构的兼容能力；
- b) OCS 组网应支持多种光互连形态，包括但不限于点到点互连和多节点互连，满足不同规模和不同业务场景下的网络需求；
- c) OCS 设备应支持标准化的光接口和传输技术，具备对不同速率和不同代际设备的透明传输能力，支持网络的平滑演进和扩展；
- d) 网络重构操作应按序执行，并设置相应机制避免因物理路径变更导致上层链路震荡或路由收敛异常。

10.2 可视化管控

可视化管控应满足下列要求。

- a) 控制平面设计：应具备独立或集成的控制平面，负责光路计算、路由选择、连接建立与拆除、拓扑重构等核心控制功能，并应提供标准化的北向接口（如 PCE、SDN 控制器等），支持与上层编排系统或网络管理系统的集成，实现光连接的自动化调度与策略下发。
- b) 拓扑发现与呈现：应能自动发现并可视化 OCS 设备间、OCS 与终端节点间的光纤连接关系及物理拓扑。
- c) 状态监测与告警：应能实时监控各 OCS 端口的光功率、链路状态及交换矩阵运行状态，并支持异常状态（如光功率劣化等）的主动告警，可基于状态信息进行故障定位。

10.3 安全

架构安全应满足下列要求。

- a) 资源统计与调度：应能呈现 OCS 端口的占用与空闲情况，支持链路资源使用率的统计与分析，可基于资源视图实施光路的建立、拆除与调整。
- b) 安全访问控制：应对 OCS 的管理平面和控制平面实施严格的访问控制策略，防止未经授权的拓扑重构指令引发算力中断。
- c) 路由安全：OCS 组网在与电交换网络协同运行时，应避免因光路重构引入网络环路、地址冲突或异常转发行为。
- d) 日志管理：应支持对关键控制操作和网络状态变化的日志记录和审计，满足安全管理和运行追溯的要求。
- e) 漏洞防护：建立完善的漏洞扫描与防护机制，对 OCS 系统及其控制软件进行定期的安全审查。

10.4 可靠性

可靠性应满足下列要求。

- a) 快速切换能力：在检测到光链路或相关设备发生故障时，OCS 应具备快速切换或重构能力，其硬件层切换时间宜控制在毫秒级以内，以支持业务的无感恢复或快速收敛。
- b) 冗余设计：OCS 组网在设计时应充分考虑设备冗余、链路冗余以及控制系统可靠性，避免因单一控制节点失效导致网络不可用。对设备冗余，支持 OCS 核心部件 1 加 1 冗余；对链路冗余，支持 OCS 冗余端口数量按工作端口的一定比例规划；对于系统冗余，支持部署一主一备两台控制器，并预先在控制平面中规划好至少一套完整的备用光路表。

- c) 自愈机制：全光数据中心 OCS 组网应具备链路状态的实时监测能力，并能够与上层电交换网络协同工作，在发生链路或节点异常时，支持自动触发故障规避和网络重构，实现网络的自愈恢复。

10.5 安装过程

安装过程应满足下列要求。

- a) 环境核查：在设备安装前，应对机房环境条件进行核查，包括温度、湿度以及光纤配线架（ODF）资源情况，确保符合 OCS 设备的物理部署和运行要求。
- b) 光纤布线与测试要求：光纤布线和光模块安装应遵循统一的标识、编号和管理规范，确保光链路的可追溯性和可维护性。安装完成后，应对每一路光连接进行插入损耗和回波损耗测试，确保其性能指标满足验收要求。例如：单根光纤插入损耗宜不大于 0.5 dB（含连接器），回波损耗宜不小于 45 dB（APC 连接器）。
- c) 自动化部署要求：OCS 组网的初始配置宜采用标准化配置文件和自动化部署方式，支持通过脚本或集中管理手段完成设备上线和参数配置，以缩短集群交付周期并降低人工配置风险。
- d) 性能验证要求：OCS 组网在上线运行前，应完成必要的功能验证和连通性测试，确保网络运行符合设计预期。

参 考 文 献

- [1] GB/T 32910.1—2017 数据中心 资源利用 第1部分：术语
 - [2] IETF RFC 8453 Framework for Abstraction and Control of TE Networks (ACTN)
-