PON in the Datacenter: Hyperscale for Management and Console

NANOG 94

MJ (Mike Joseph)
Network Engineering Manager

Meta

Agenda

01 Introduction

Infrastructure Network Engineering (INE)

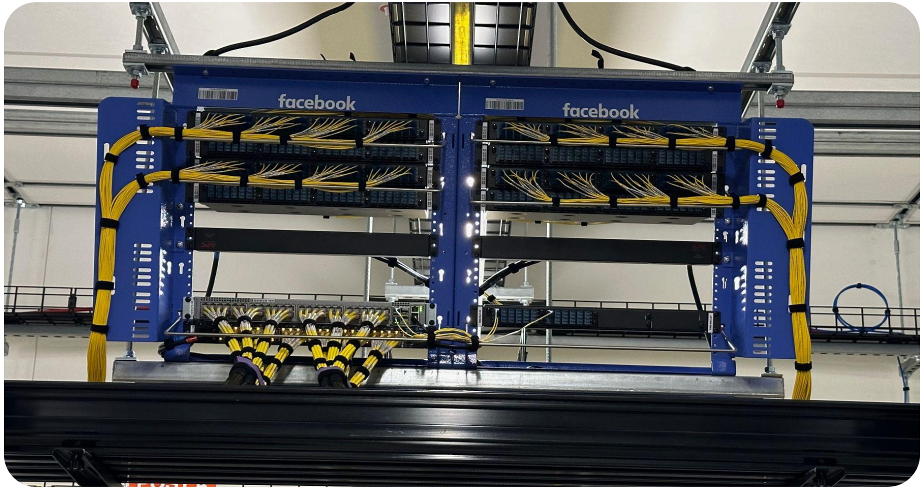
We are the team responsible for network element management connectivity (ethernet & serial) across Meta's footprint. Additionally, we provide out-of-band and disaster-recovery solutions. Our services are furnished to every rack in every datacenter and every PoP, worldwide.

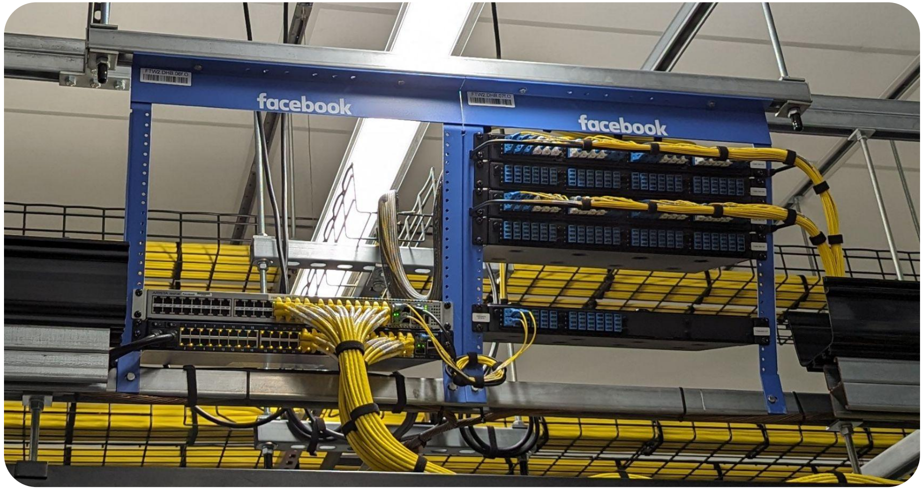
To these ends, we operate the largest of Meta's various networks by racks served. If it's production and <= 1GE or RS-232, it's provided by INE.

Datacenter Infrastructure

Meta datacenters utilize elevated “goalpost” racks for fiber and management infrastructure.

Management connectivity is currently handled by fixed ethernet switches and terminal servers located in the goalpost at the end of each row.





Limitations

- Fixed (and limited) footprint for ethernet & serial connections at each rack position.
  - Alternative rack types require row-level coordination.
  - High variability in each row.

- Large number of row-level devices to manage.

- Eventual maximums in overall management rack capacity (whether goalpost or otherwise).

- Significant copper cabling requirements between goalposts and rack positions.

02 PON Rationale

Using PON in the Datacenter

Passive Optical Network (PON) is a typically a residential/SMB ISP access technology. Why use it in a datacenter environment?

For us, it solves specific flexibility, scale, and efficiency requirements while reducing overall cost and power. Moreover, by standardizing the handoff to a rack, PON allows Meta to decouple the rack design from the datacenter topology and evolve each independently.

Flexibility

Our current PON-based solution provides for up to 20 Ethernet connections and 20 console connections, or up to 32 Ethernet connections and 8 console connections, per rack. The exact configuration of each rack can be different without impacting the connectivity options of other racks in the row. Defining a new rack type is as simple as specifying the combination of PON ONUs and console cards for that type of rack.

We're actively developing even larger rack-level products. Future rack-level products will remain compatible with current datacenter PON infrastructure and handoffs.

Scale

We are able to provide for up to 1280 Ethernet connections and up to 800 console connections (with some tradeoff between them) **per row**, in our current generation of datacenters, using only the rack-level PON products available today.

We can manage up to 12 rows of equipment by just 2 centralized routers serving as OLT aggregation devices. The use of OMCI-based management also provides for uniform management of multi-vendor ONU & console server devices. By utilizing purpose-built console devices, Meta is further able to standardize our console access mechanisms.

Workflow Improvements

We are able to deliver the required scale and flexibility while reducing operational impact due to fully-automated configurations. Because ONUs are scanned in the field as racks are deployed, we can dynamically update our asset management databases in real-time.

The configuration for our OLTs (and thus also for all ONUs in a datacenter) is generated directly from our network modeling service and our asset management databases. Because ONUs are stateless devices, replacing failed hardware is as simple as scanning a new one.

Efficiency

Our PON-based solution drives efficiency by:

- Reducing wasted ports by deploying in increments of 4 UNIs
- Reducing effective cost and power per-port of both management Ethernet and RS-232 console
- Improving space utilization through the creation of the new canopy design on the top of the rack

Cable Plant Simplification

Adopting PON has allowed Meta to entirely eliminate all copper cable trays and copper distribution from the data halls inside our datacenters. Copper for management ethernet and serial are now localized to within each rack.

The PON fiber handoffs to racks are also standardized. We provide for 2 or 4 fiber drops to each rack position, allowing for full fungibility of rack types within the data hall. All complexity is contained to the rack-level, decoupling the datacenter PON topology from the rack requirements.

New Product Development

One of the major benefits of a PON-based solution is the ability to onboard new ONU products quickly, regardless of chipset or embedded OS, due to the uniform nature of OMCI management. There's no need to adopt tooling any particular ONU's network operating system as you would need to do with a switch or router.

This also creates opportunities for supply-chain diversification.

Staffing

Because it is an industry standard technology, adopting PON also provides a pathway for Meta to be able to hire network engineers from more traditional wireline ISP backgrounds. Exposure to evolving access technologies like PON similarly benefits existing Meta network engineers by providing unique growth opportunities.

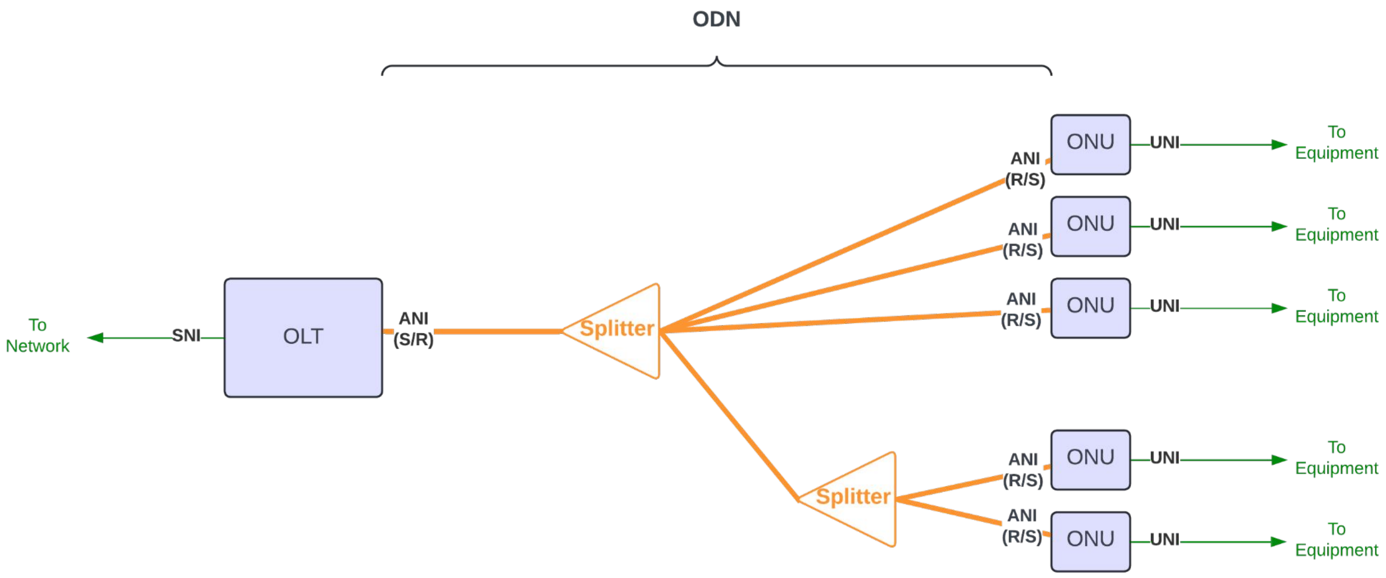
The gains in overall team efficiency from managing fewer devices in a more centralized way frees up operational network engineering resources, further allowing team members to focus on higher-level design work.

03 PON Fundamentals

PON Overview

Passive Optical Network (PON) is designed as an ISP access technology and is typically used for residential/SMB service.

It is a shared-medium technology involving an Optical Line Terminal (OLT) and many Optical Network Units/Terminals (ONUs/ONTs) on a fiber path. In addition to the OLT directly controlling timing and bandwidth on the PON, the ONUs/ONTs are also fully-managed by the OLT headend, using a pair of management protocols: PLOAM & OMCI.



PON uses a point-to-multipoint (tree) topology. The OLT controls all aspects of the PON and is at the root of the tree.

Between the OLT and ONUs is the Optical Distribution Network (ODN), which is made up of fiber trunks and optical (passive) splitters.

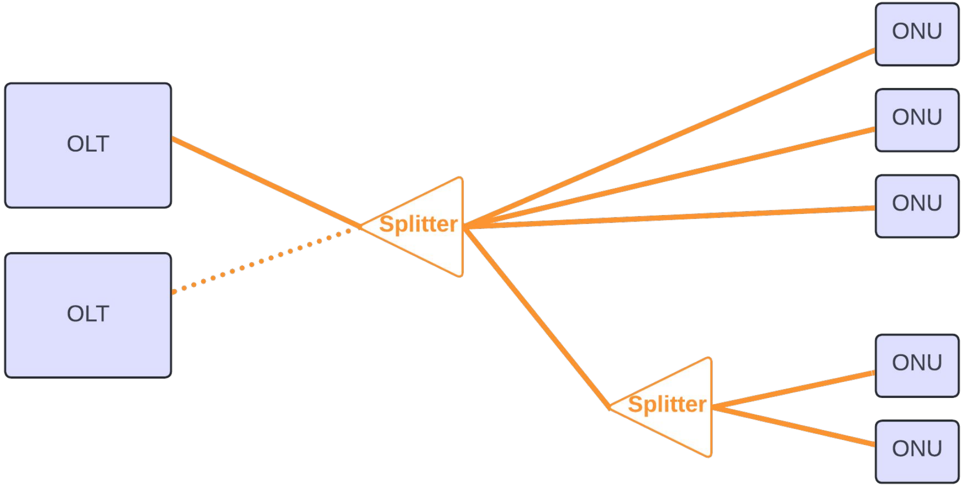
The endpoints of the PON are the ONUs. They provide various types of User Network Interfaces (UNIs) to client devices.

PON Technologies

- PON uses bidirectional (BiDi) optics and simplex fiber, requiring separate upstream and downstream wavelengths
- PON can share the ODN with certain other technologies, including other PON versions
- Faster speeds (25G, 40G, 50G, 100G, 200G) are also defined, and IEEE (non-ITU) versions (EPON family) also exist

G-PON	XG-PON	XGS-PON
ITU-T G.984	ITU-T G.987	ITU-T G.9807.1
Up: 1.25 Gbps 1310 nm	**Up:** 2.5 Gbps 1270 nm	**Up:** 10 Gbps 1270 nm
Down: 2.5 Gbps 1490 nm	**Down:** 10 Gbps 1577 nm	**Down:** 10 Gbps 1577 nm
125µs clock	125µs clock	125µs clock

Redundancy (Type B & “Type F”)

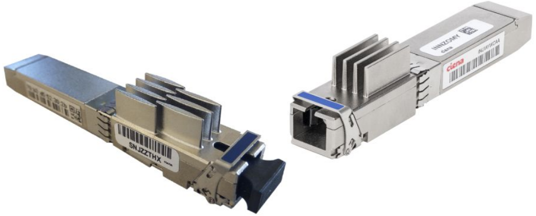


04 Meta’s Design

Development

Meta worked closely with our initial PON equipment vendor to bring our visions for PON in the datacenter to reality. We drove requirements and helped to iterate design concepts with them, resulting in a publicly-available product line, for which we are their launch customer.

All active equipment and most passive equipment shown in this talk are general catalog items. A few limited items (e.g. specialized splitters or cables) are proprietary to Meta. Where possible, we have worked to contribute back to public standards (e.g. ORv3) to allow for wider adoption of this technology.



Headend

Centrally-located in each Meta datacenter.

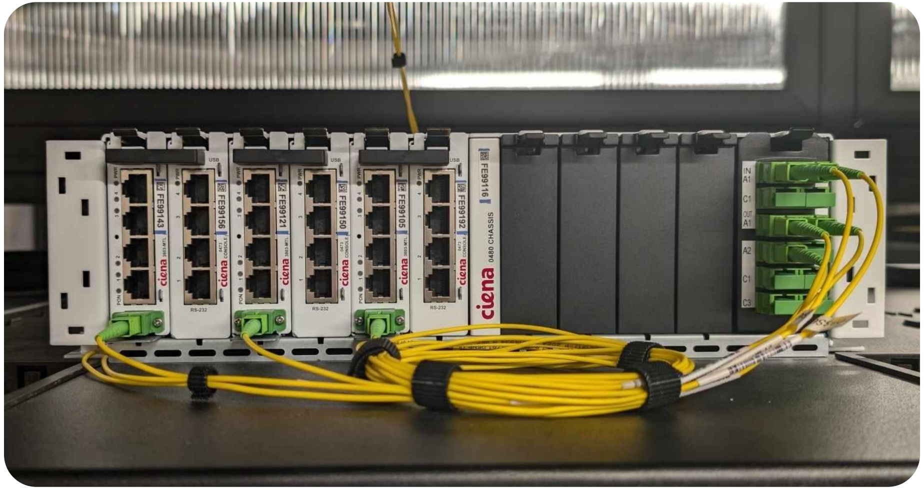
Rack(s) of OLTs to serve all rows of the datacenter building(s). In our topology, each of these 1 RU devices provides for 36 PONs across ~500 racks.

Upper-right image source: Ciena

Rack Canopy

New PON canopy chassis on top of each Meta rack. Attachment features published in ORv3 standards.

Each rack is fed a uniform 2-fiber or 4-fiber drop. All variability is self-contained at the rack.

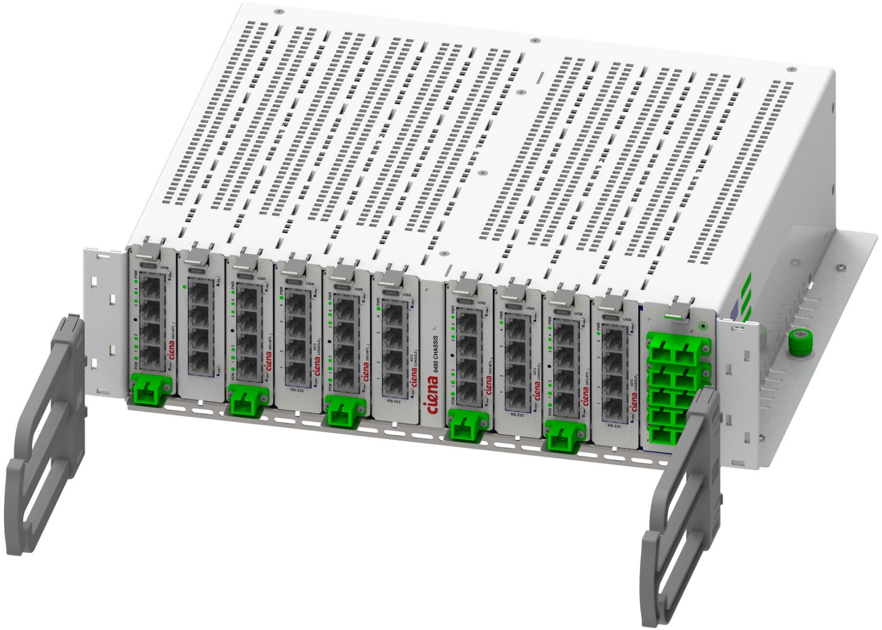


PON Chassis

Each chassis houses a flexible combination of splitter, ONUs, and console cards.

The chassis provides only power to the ONU. All data connections are on the front panel.

Additional form factors under development.



Lower-right image source: Ciena

ONU & Console

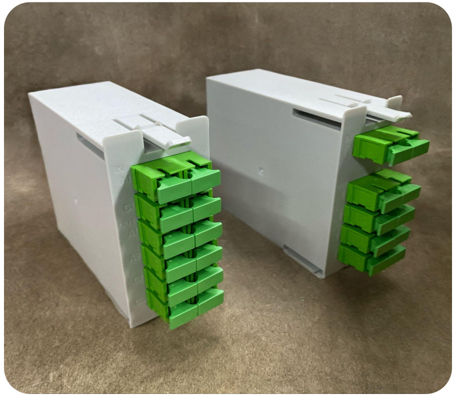
Current hardware uses a 4-port GE ONU with an optional 4-port RS-232 daughter card, connected via USB-C on the front.

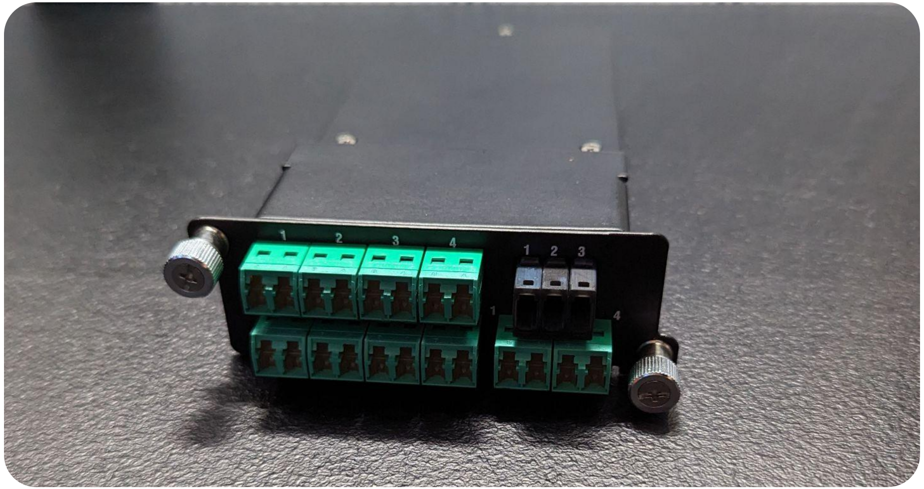
PLOAM/OMCI are exclusively used for management. The ONU is not individually-managed (no shell or management services running). Dedicated SSH ports are used for console access.

Splitters

Our topology utilizes splitters in the headend, distribution (goalposts), and in the rack itself as part of the PON canopy chassis. They are a mix of standard and proprietary products.

We also locally breakout an extra branch from the splitters in the goalpost for the provision of monitoring ONUs at the row-level.

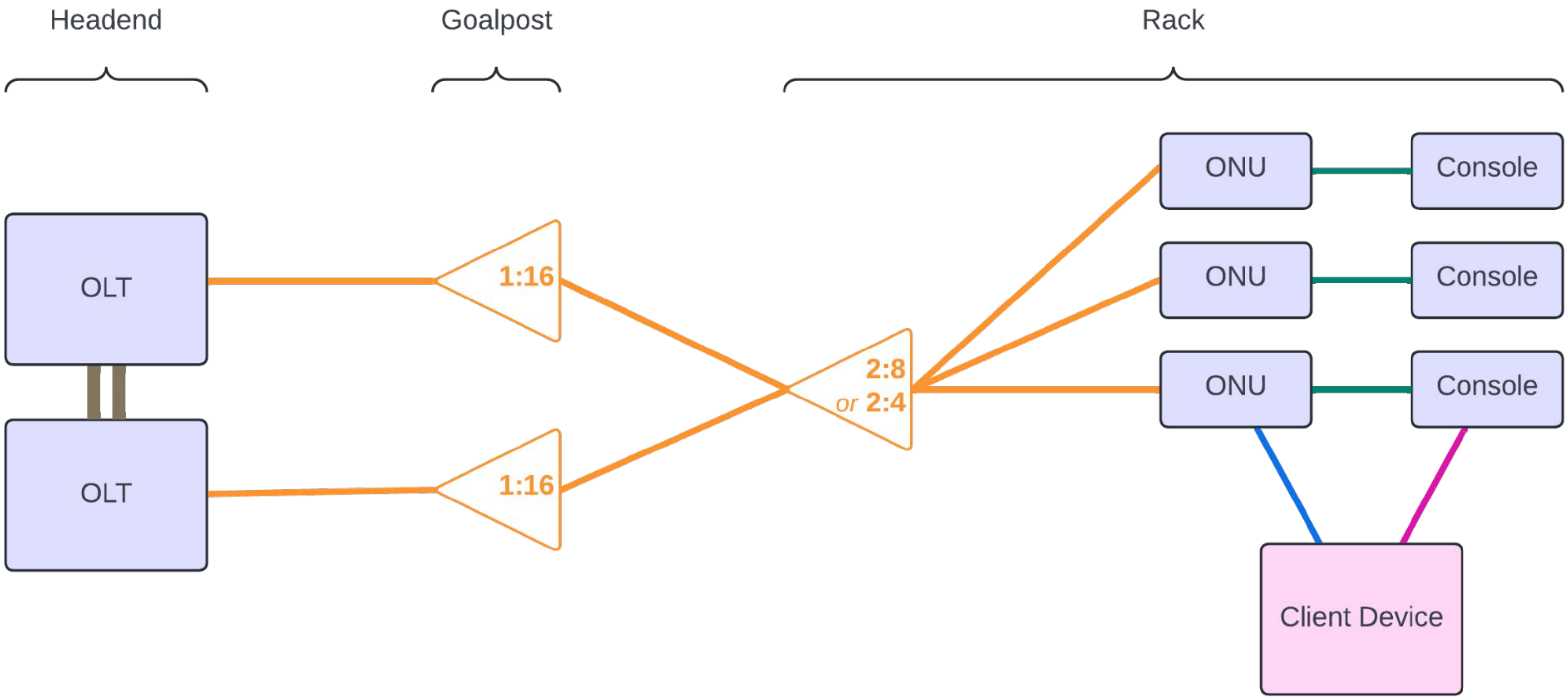




Normal Availability (NA) PON Network

Most Meta datacenter equipment uses a single management ethernet and single serial port. This equipment is served by our Normal Availability (NA) PON network, our primary solution.

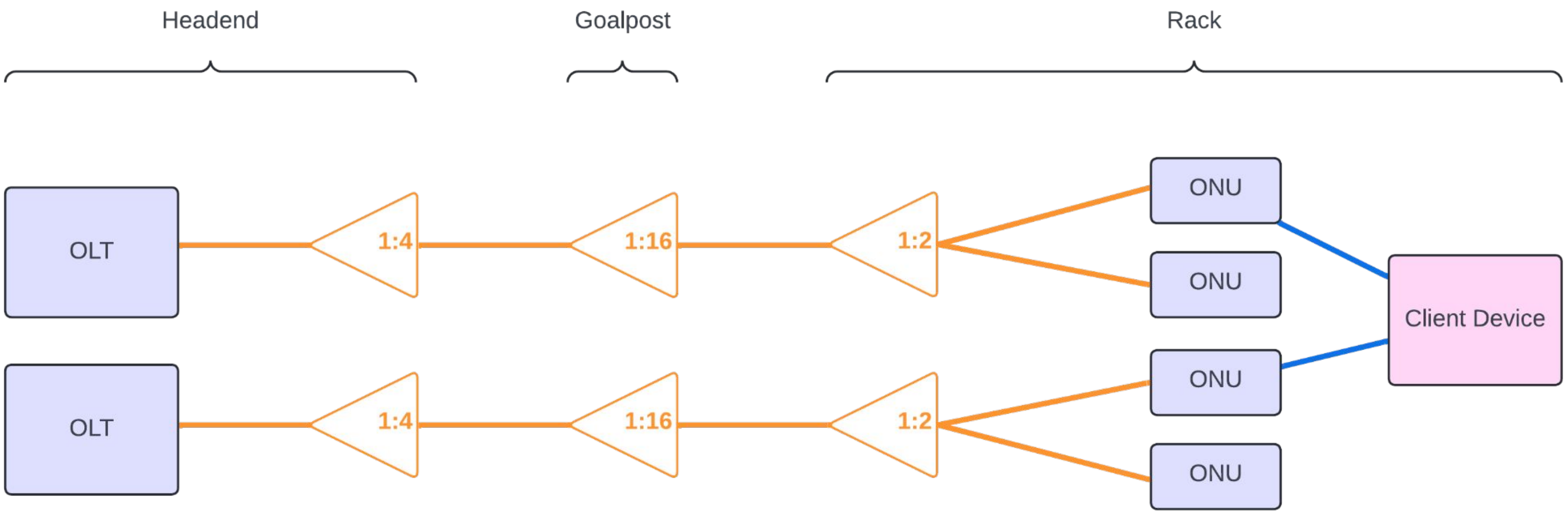
The NA network utilizes a pair of OLTs in diverse headend locations for every PON operating in Type F redundancy (warm-standby) mode. The dual fiber paths for NA converge at the rack splitter. All racks in the data hall are served by the NA network, with fibers pre-run to all data hall rack positions using multi-drop fibers running in both directions above the row.


Headend
Goalpost
Rack
OLT
OLT
1:16
1:16
2:8
or 2:4
ONU
ONU
ONU
Console
Console
Console
Client Device

High Availability (HA) PON Network

Select Meta equipment requires dual management ethernet connections via fully independent paths. We solve for this by operating an additional pair of PON networks in the datacenters.

The High Availability (HA) networks utilizes a pair of OLTs for every PON operating independently (active-active). The dual fiber paths for HA serve separate ONUs. Because most racks do not require HA connectivity, we do not pre-provision HA fiber drops to all rack positions. They are run directly from the goalpost splitters as needed.


Headend
Goalpost
Rack
OLT
1:4
1:16
1:2
ONU
ONU
Client Device
OLT
1:4
1:16
1:2
ONU
ONU

Software & Configuration

Due to the reliability requirements of an OOB network, we eschew off-box controllers. Instead, we worked with our initial PON vendor to ensure that the software stack for operation of each PON OLT router is fully-integrated into the router and all external dependencies are eliminated. This dramatically improves survivability in case of disaster-recovery scenarios and simplifying our operations.

Automation-Driven Workflows

The configuration of the OLT router — including portions for each OLT and ONU, as well as all firmware images — are deployed continuously using network device automation systems.

We exclusively utilize Netconf & gNMI to manage our OLT router fleet.

05 Next Steps

Current Status

PON is the plan-of-record for all Meta datacenter and rack builds going forward. We will be providing PON-based service to every rack in every new datacenter, worldwide.

We have PON headends up and running in a number of our next-generation full-scale datacenters currently under construction, with many additional datacenters utilizing PON starting construction in the next several quarters. We anticipate our first PON production rack activation next month.

Future Growth

We are working with our vendors to develop additional PON datacenter rack products, including a higher-density ONU, a high-capacity in-rack chassis, and a 1 RU form-factor device for non-ORv3 rack types, among other rack product design concepts.

We are examining the use of our PON solutions for other applications in the INE space at Meta, including for our datacenter facilities networks as well as in our PoPs worldwide. Unifying our management access layer is a major goal for us and PON provides a compelling path achieve this.

Meta