CONSORTIUM OVERVIEW

CONTENT

- Vision
- Founding Members
- Deployment Models
- Motivation
- Approach
- Technical Goals
- Working Groups
- Timeline
- Example Working Group Focus - Transport
- Next steps

ULTRA ETHERNET VISION

Deliver an Ethernet based open, interoperable, high performance, full-communications stack architecture to meet the growing network demands of AI & HPC at scale

THE NEW ERA NEEDS A NEW NETWORK

Ultra Ethernet

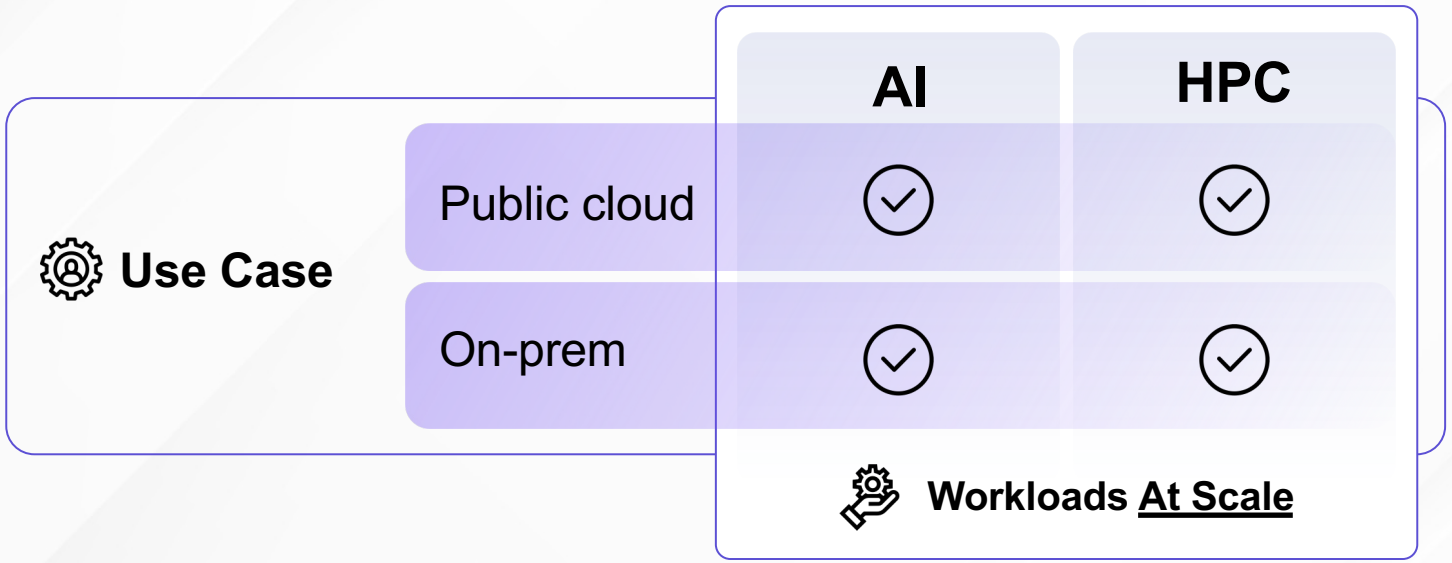
As **performant** as a supercomputing interconnect

As **ubiquitous** and **cost-effective** as Ethernet

As **scalable** as a cloud data center

STEERING MEMBERS

TARGET DEPLOYMENT MODELS / USE CASES



Profiles will be defined for AI and HPC use cases

ETHERNET IS THE WAY

Why?

- **Open / Multivendor**: Switches, NICs, cables, optics, tools, software
- **Scalable**: Addressing and routing for rack-, building-, DC- scale networks
- **Tools**: Many tools for testing, operations, measurements
- **Cost**: Economies of scale and competitive market
- **Supporting Standards**: Regular progress in IEEE, for many technologies, across layers

The largest AI and HPC networks are based on Ethernet

APPROACH

The founding companies are seeding the consortium with highly valuable contributions in four working groups: **Physical Layer, Link Layer, Transport Layer and Software Layer**.

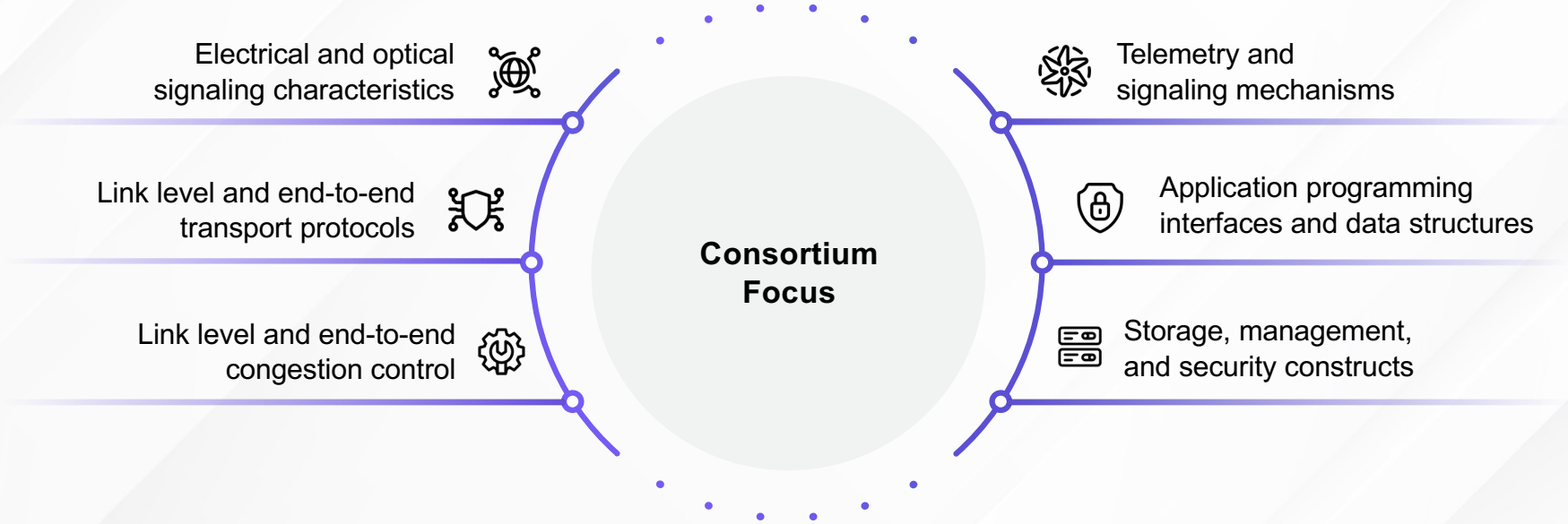
UEC will follow a **systematic approach with modular, compatible, interoperable layers** and tight integration of these layers to provide a holistic improvement for demanding workloads is paramount.

The consortium will work on **minimizing communication stack changes** while maintaining and **promoting Ethernet interoperability**.

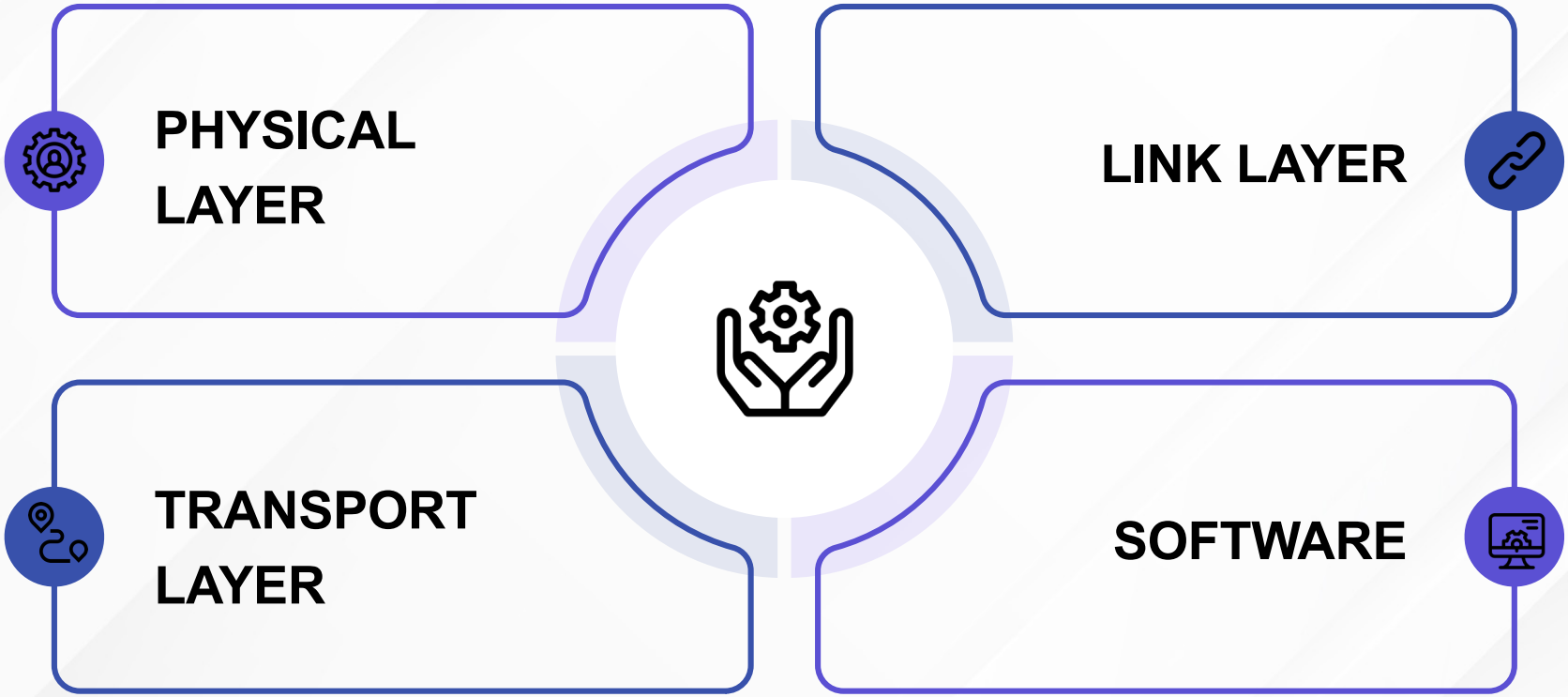
Project under the Joint Development Foundation (JDF) of the Linux Foundation

TECHNICAL GOALS

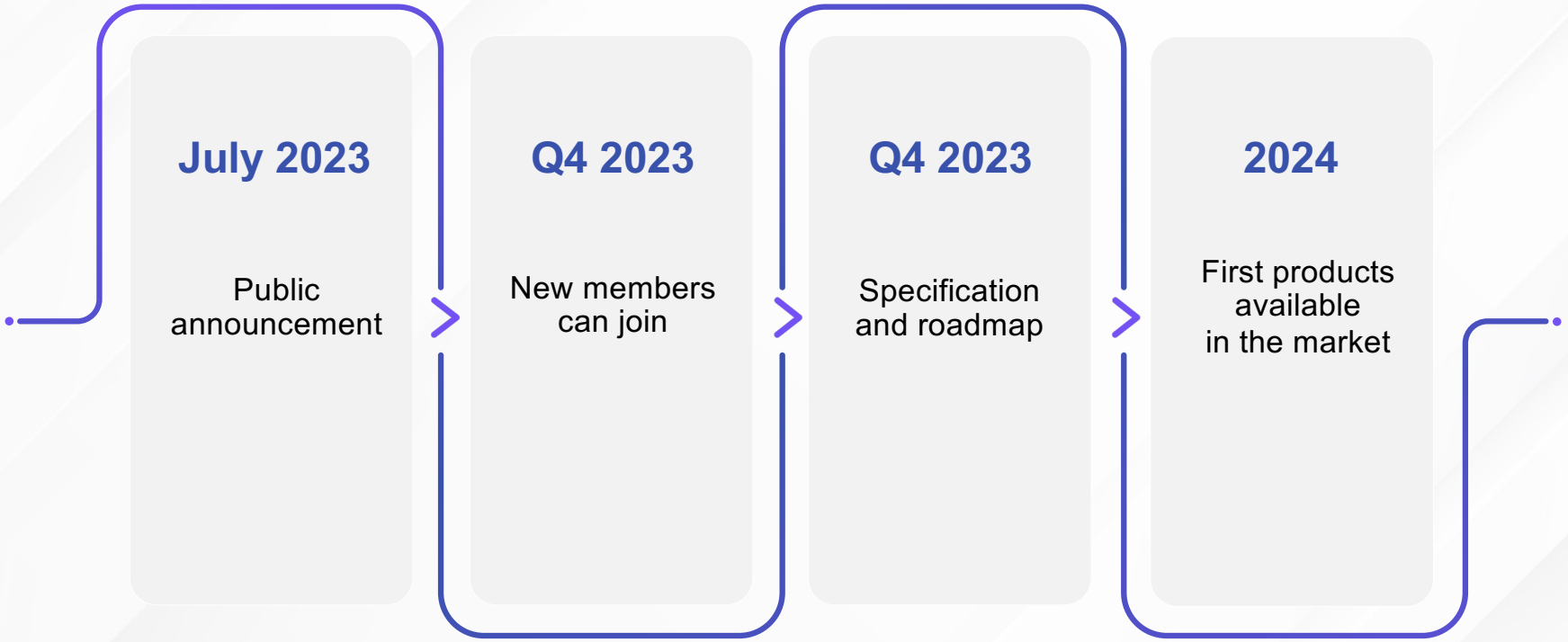
Open specifications, APIs, source code for optimal performance of AI and HPC workloads at scale.



UEC WORKING GROUPS



TARGET TIMELINES

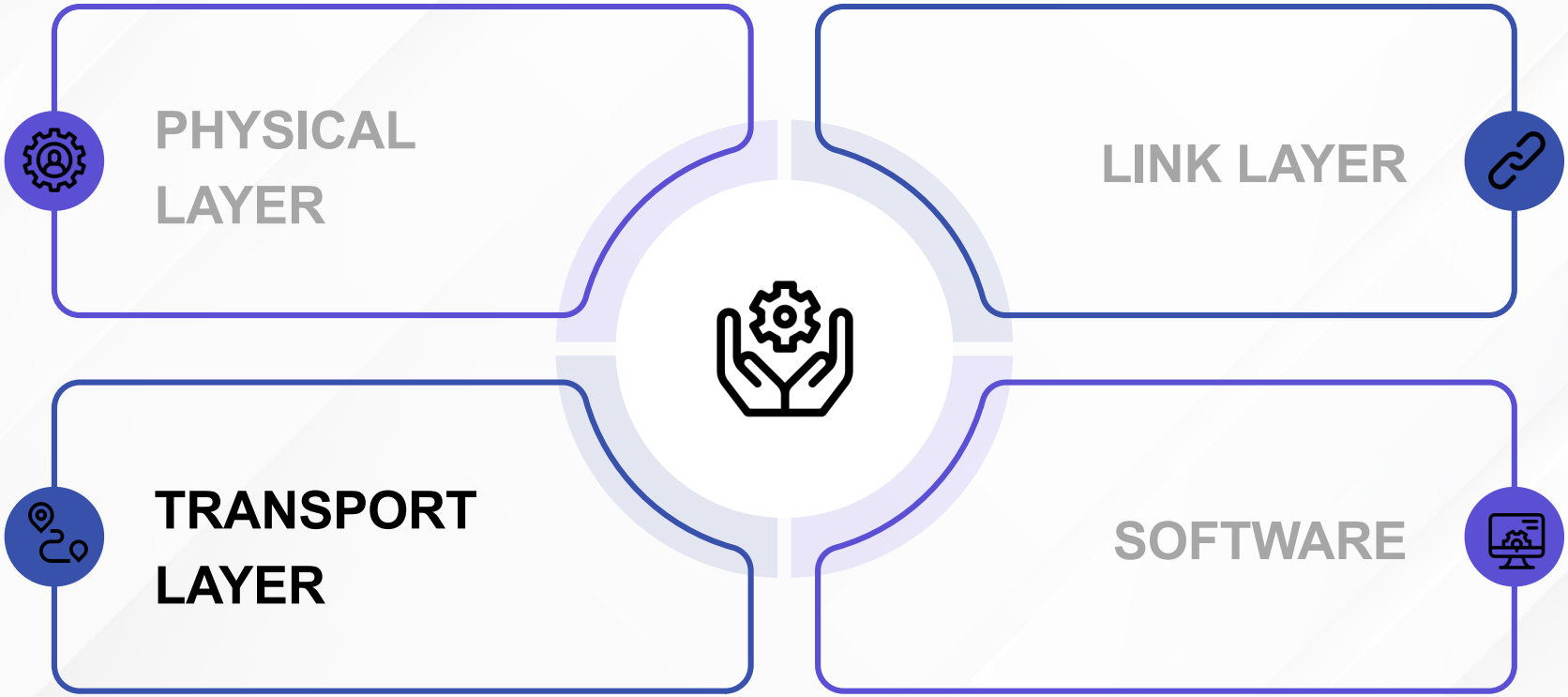


UEC SUMMARY

- Evolves Ethernet to meet the growing network demands of AI and HPC **at scale**
- UEC members create and operate some of the **largest AI and HPC networks**
- UEC specifications will be **open** and broadly available
- UEC will provide a **lasting benefit to AI and HPC** apps of the future

Ultra Ethernet Consortium

UEC WORKING GROUPS



UEC TRANSPORT – MOTIVATION

RDMA HAS BEEN GOOD

- The key benefit is **direct transfer** from applications to/from network

  Zero-copy, no OS intervention

- And this is indeed *very* important

but…

RDMA (AS IMPLEMENTED TODAY) IS DEFICIENT

- **Lack of multipathing** makes load balancing difficult and solutions brittle
- Requires **in-order packet delivery**
- **Go-back-n:** massively inefficient for dropped packets
  necessitates a “lossless” network
- DCQCN congestion control is **brittle and hard to tune**
  - Specific to workload and network details

➔ **Challenges with scale**

It’s time to modernize RDMA

UEC TRANSPORT – KEY PROPERTIES

- Scales to **1,000,000 nodes**
- **Packet-level multipathing** for very high network utilization
- **AI-optimized, configuration-free congestion control**
  - **Incast management** to address fan-in at the last hop
  - **Rate control** to ramp quickly to wire rate without impacting existing flows
- Support for **out-of-order packet delivery** with in-order message completion
- **Low tail latency**

Highest infrastructure utilization** at **ultra-high scale, without tuning

OTHER UEC EFFORTS

- **UEC Security**
  - Scalable, first-class citizen
  - Leverages the best of IPSec and PSP
  - Encrypts all traffic within a job
  - Adds efficient key management for jobs, with small session state
- **Streamlined APIs**
  - Simplified RDMA
  - APIs for AI and HPC: *CCL, MPI, PGAS, OpenShmem
- **AI- and HPC-optimized congestion notification and control**
- **APIs and header formats for in-network computation**

Modern Transport and RDMA Services Needs for AI and HPC

Requirement	UEC Transport	Legacy RDMA	UEC Advantage
Multi-Pathing	Packet spraying	Flow-level multi-pathing	Higher network utilization
Flexible Ordering	Out-of-order packet delivery with in-order message delivery	N/A	Matches application requirements, lower tail latency
AI and HPC Congestion Control	Workload-optimized, configuration free, lower latency, programmable	DCQCN: configuration required, brittle, signaling requires additional round trip	Incast reduction, faster response, future-proofing
E2E Telemetry	Sender or Receiver	ECN	Faster congestion resolution, better visibility
Simplified RDMA	Streamlined API, native workload interaction, minimal endpoint state	Based on IBTA Verbs	App-level performance, lower cost implementation
Security	Scalable, 1st class citizen	Not addressed, external to spec	High scale, modern security
Large Scale with Stability and Reliability	Targeting 1M endpoints	Typically, a few thousand simultaneous end points	Current and future-proof scale

LEARN MORE AND JOIN THE MOVEMENT AT

www.ultraethernet.org


Ultra Ethernet
Consortium