Why Do We Need Edge Computing? Motivation & Evolution

Chapter 1

Outline

- Background
  - Big data & IoT
  - Centralized cloud
  - Latency, bandwidth and privacy driver
- Evolutionary history
  - Distributed databases -> P2P
  - CDNs -> MEC -> Fog
  - Cloud-Sea computing
- What is edge computing
  - Definitions
  - Edge-edge intelligence link
  - Edge growth
- Summary & Practice
  - Summary
  - Practice

Background: Internet of Things

IoT connects physical objects to the Internet — enabling identification, positioning, tracking, monitoring and management without human intervention.

Sensor Technology

Acquires information from the physical world and converts it into output signals — the “senses” of IoT (eyes, nose, skin).

RFID Technology

Radio-frequency + embedded tech to automatically identify objects and read data — no human intervention, even in harsh settings.

Embedded Systems

Hardware + software + circuits in one — from smartwatches to satellites. The “brain” that classifies and processes information.

Sensors are the senses, the network is the nervous system, and the embedded system is the brain that processes what comes in.

Background: Internet of Things to Internet of Everything (IoE)

Cisco introduced the IoE concept in December 2012 — the evolution of IoT for future connectivity.

IoT

Things-to-thing

- Connects physical objects to the Internet
- Focus on identification & data collection
- Objects exchange information

→

IoE

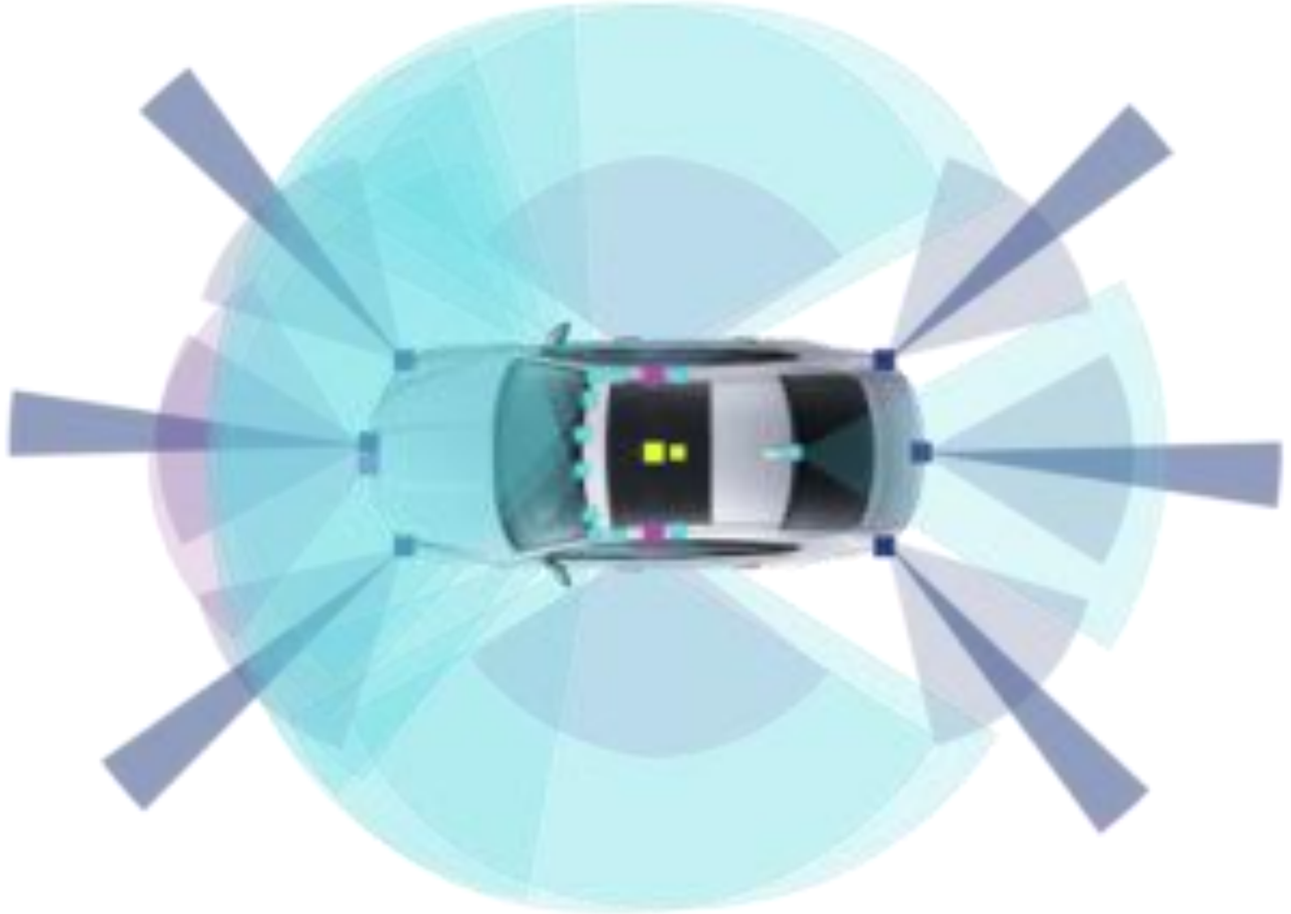
Human-to-thing + context

- Distributed architecture: apps, computing & storage
- Every “thing” gains context awareness & sensing
- Billions–trillions of connected nodes
- Driven by IP, higher bandwidth & IPv6

Background: Data Never Sleeps

IoE applications demand short response times and generate enormous, privacy-sensitive data.

https://www.domo.com/learn/infographic/data-never-sleeps-11



- A single forward facing **radar** operating at 2800 MBits / s generates more than **1.26 TB of data per hour**, and a typical data collection **car** generates in excess of **30TB a day**.
- A two megapixel camera (24 bits per pixel) operating at 30 frames per second generates 1440 Mbits of data every second, so a **five camera setup** can generate in excess of **24 TB per day**.

Background: Data Explodes in Three Dimensions

Velocity, Variety, Volume — the cloud paradigm works when data is confined in speed, size and format. IoE removes those limits.

Velocity

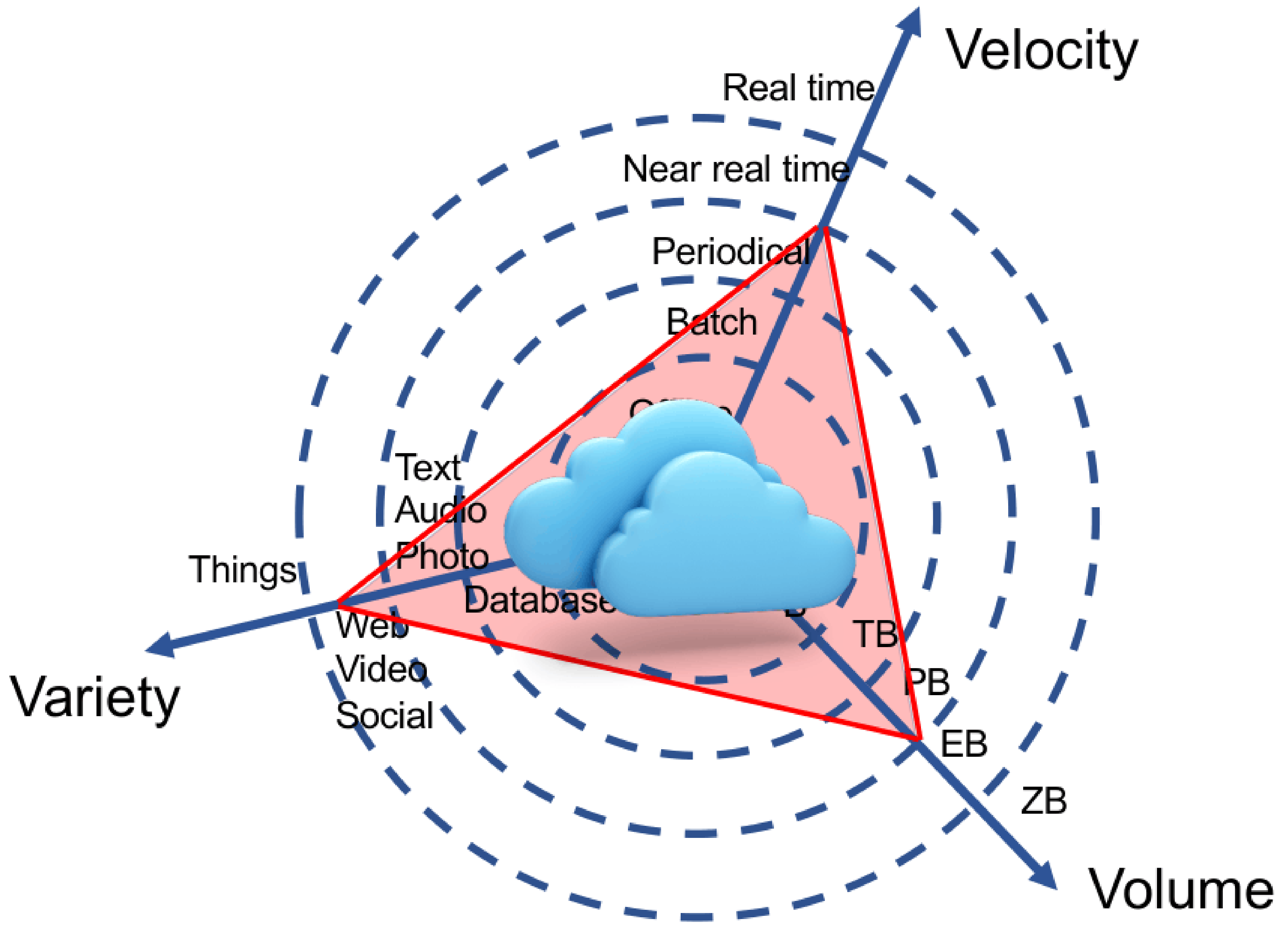
How fast data arrives — real-time · periodic · batch

Variety

How many forms it takes — text · audio · photo · video · social

Volume

How much is produced — MB & TB are now the everyday units



The increase of data pushes the evolution of edge computing

Background: Three Drivers

Edge computing's distinct advantages over cloud for mobile & IoT applications.

Alleviates Bandwidth & Data-Center Pressure

Most IoE data is temporary and never needs long-term storage — devices generate ~2 orders of magnitude more data than is stored. Process it at distributed edges instead of shipping it all upstream.

Enhances Service Responsiveness

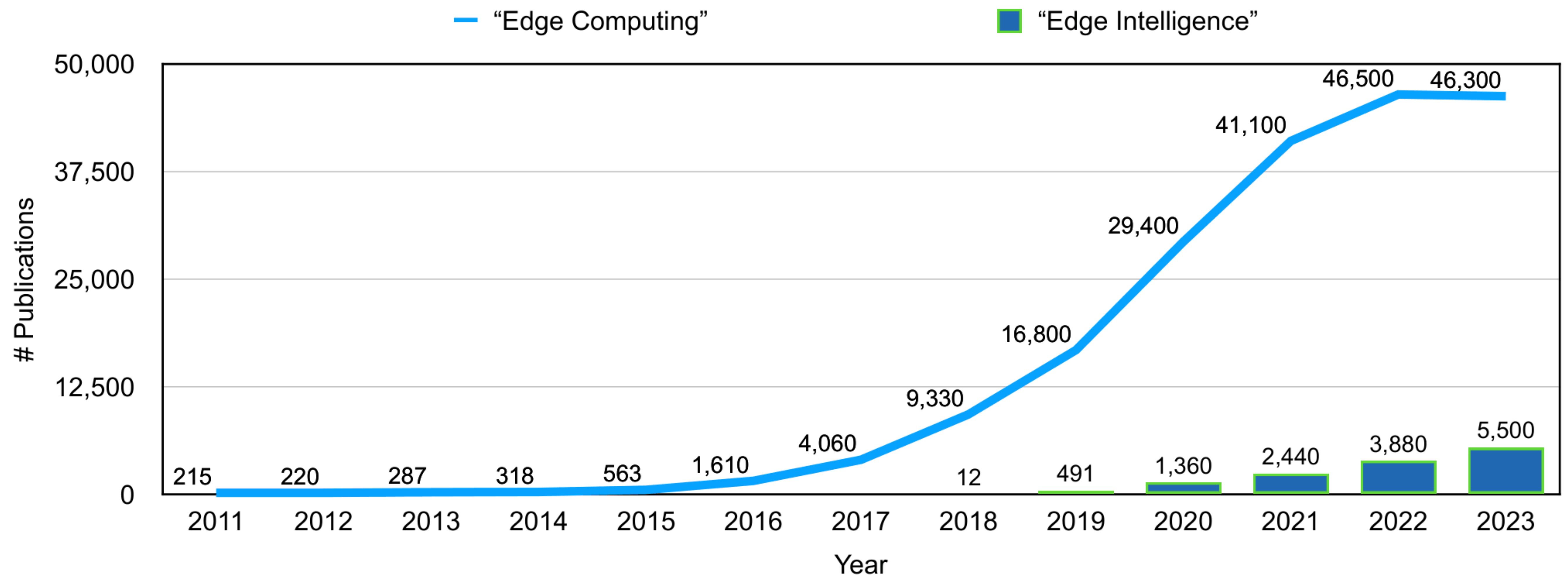
Serving computation near the user means low latency through proximity and less jitter via simpler routing. Edge and 5G/6G reinforce each other on throughput, latency and reliability.

Protects Privacy & Enhances Security

~86% of the U.S. public is concerned about data privacy. Keeping sensitive data on edge devices behind local firewalls gives fine-grained control the cloud model can't.

Evolutionary History: Three Stages of Development

Using "edge computing" as a Google Scholar keyword, publications reveal three clear phases.



Publications for "edge computing" and "edge intelligence."

Technology Preparation *before 2015*

Dormancy → presentation → definition. Data-oriented models mature.

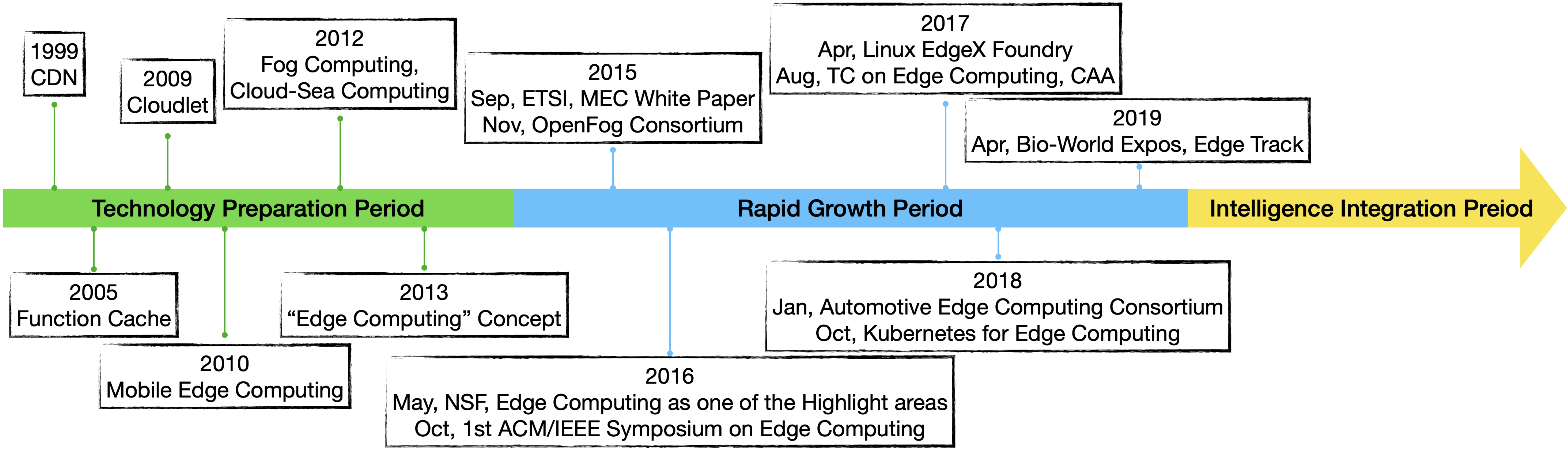
Rapid Growth *since 2015*

Papers grow 10×. Government, academia & industry pile in.

Intelligence Integration *since 2020*

Academia + industry converge; edge meets AI/ML.

Evolutionary History: Road to the Edge



The evolution of edge computing and key milestones.

Distributed DBs → P2P → CDN → cache/cloudlet → MEC → fog → Cloud-Sea, each shifting computation toward the network edge.

Evolutionary History: Distributed Databases & P2P

Distributed Database Models

- Database tech + network tech; data spread across many machines
- Homogeneous vs heterogeneous; SQL, NoSQL, XML, NewSQL
- NoSQL & NewSQL most widely used (Spanner, VoltDB…)
- vs edge: less attention to heterogeneous compute; needs lots of space; weaker privacy; consistency is hard

Peer-to-Peer (P2P) Models

- Term coined in 2000 for file-sharing — early “compute at the edge”
- Key ideas: decentralization, scalability, churn tolerance
- Gave us DHTs, gossip protocols, multimedia streaming
- Edge computing extends P2P concepts to network-edge devices — a fusion of P2P and cloud

Evolutionary History: CDNs, Function Cache & Cloudlet

CDN (Akamai, 1998)

Internet-based caching network. Routes users to the nearest cache server via load balancing — cutting congestion, boosting response speed. The early "edge" = CDN cache servers.

Active CDN (ACDN)

Improves on CDN: apps deploy on any server, with algorithms to replicate & migrate them as needed — no need to predict resource locations in advance.

Function Cache & Cloudlet

Function cache (Ravi et al., 2005) caches computation, not just data. Cloudlet (Satyanarayanan et al.) = a trusted, resource-rich "small cloud" at the edge for mobile devices.

Today's "edge" goes far beyond CDN cache servers — it is any compute/storage/network resource from data source to cloud.

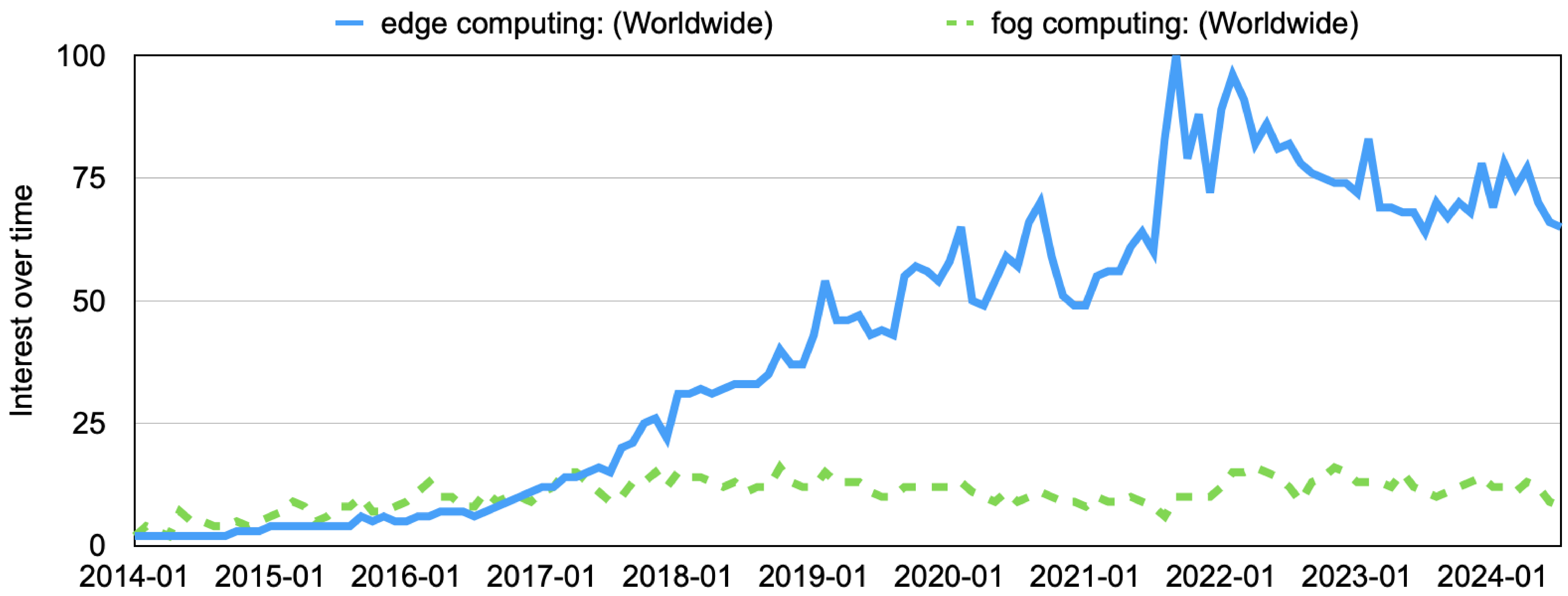
Evolutionary History: Mobile Edge Computing (MEC) & Fog Computing

MEC — standardized by ETSI (2014)

- IT services + cloud capability inside the wireless access network
- Offload intensive mobile tasks to nearby edge servers → lower latency, higher bandwidth

Fog — introduced by Cisco (2012)

- A virtualized platform migrating cloud tasks to edge devices
- Adds an intermediate “fog layer” of servers between cloud & devices



The evolution of edge computing and key milestones.

Evolutionary History: Cloud-Sea Computing & Rapid-Growth Surge

Cloud-Sea Computing

- “Sea” = augmented client-side of human- & world-facing devices
- Integrates the cloud system with the sea system
- Focuses on the two ends; edge focuses on the path between them

Rapid Growth — Key Milestones (2015→2019)

2015	ETSI MEC white paper; OpenFog Consortium founded
2016	NSF names edge a highlighted area; 1st ACM/IEEE SEC; ECC founded in Beijing
2017	ETSI → “multi-access edge computing”; Linux EdgeX Foundry launched
2018	Automotive ECC formed; Kubernetes enters the edge (→ KubeEdge)
2019	Bio-IT World adds an edge track — edge reaches healthcare

Edge Computing: No Single Definition

IBM

A distributed computing framework that brings enterprise applications closer to data sources — IoT devices or local edge servers.

Cisco

A model that shifts computing resources from central data centers or public clouds closer to devices — embedded at the edge of provider networks.

Satyanarayanan

A paradigm where substantial compute & storage — cloudlets, micro data centers, fog nodes — sit at the Internet's edge near devices/sensors.

Yousefpour et al.

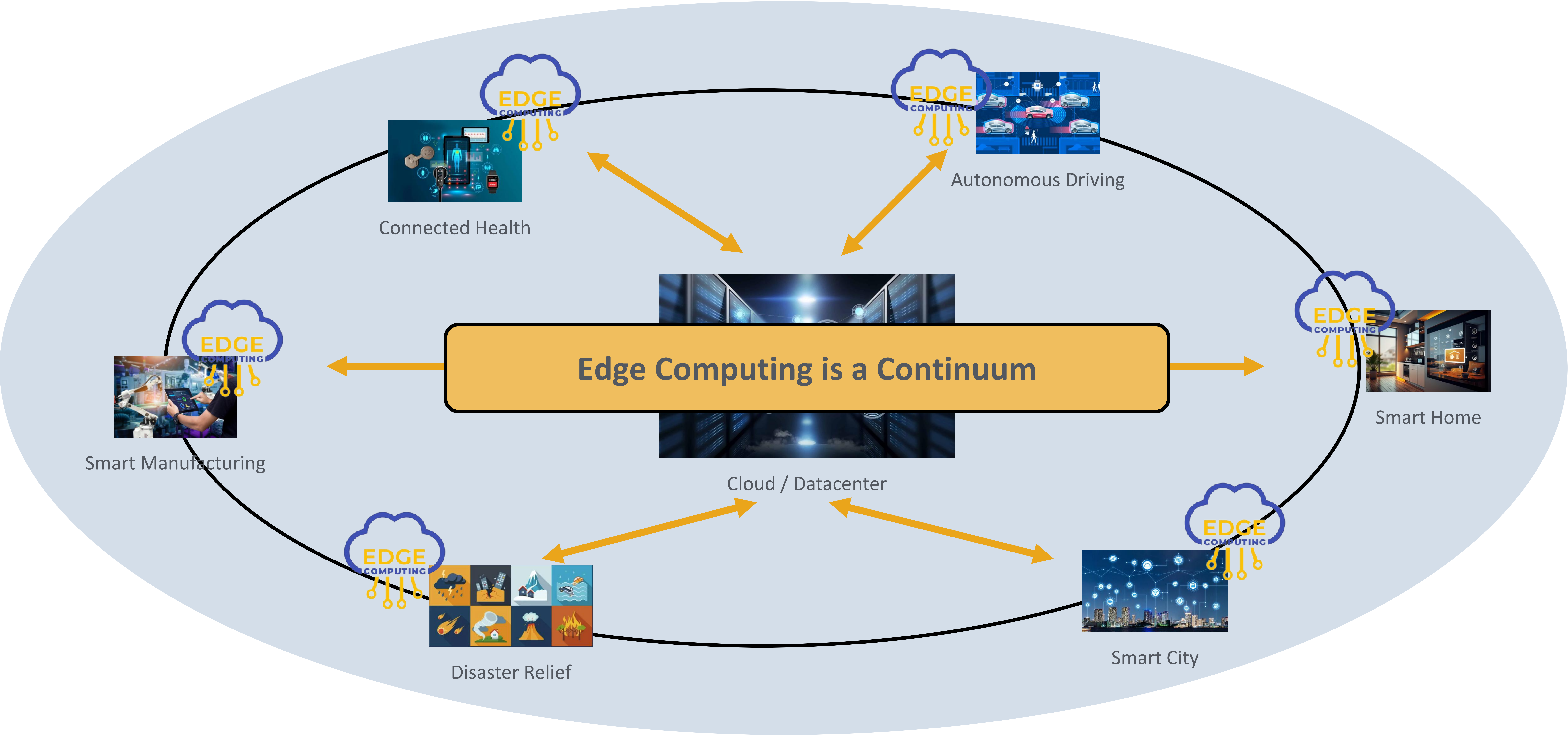
Located at the network edge near IoT devices; the edge can be more than one hop away within the local IoT network.

Shi et al., "Edge Computing: Vision and Challenges," 2016

> *"Edge computing refers to the enabling technologies allowing computation to be performed at the edge of the network — on downstream data on behalf of cloud services, and upstream data on behalf of IoT services."*

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE internet of things journal, 3(5), 637-646.

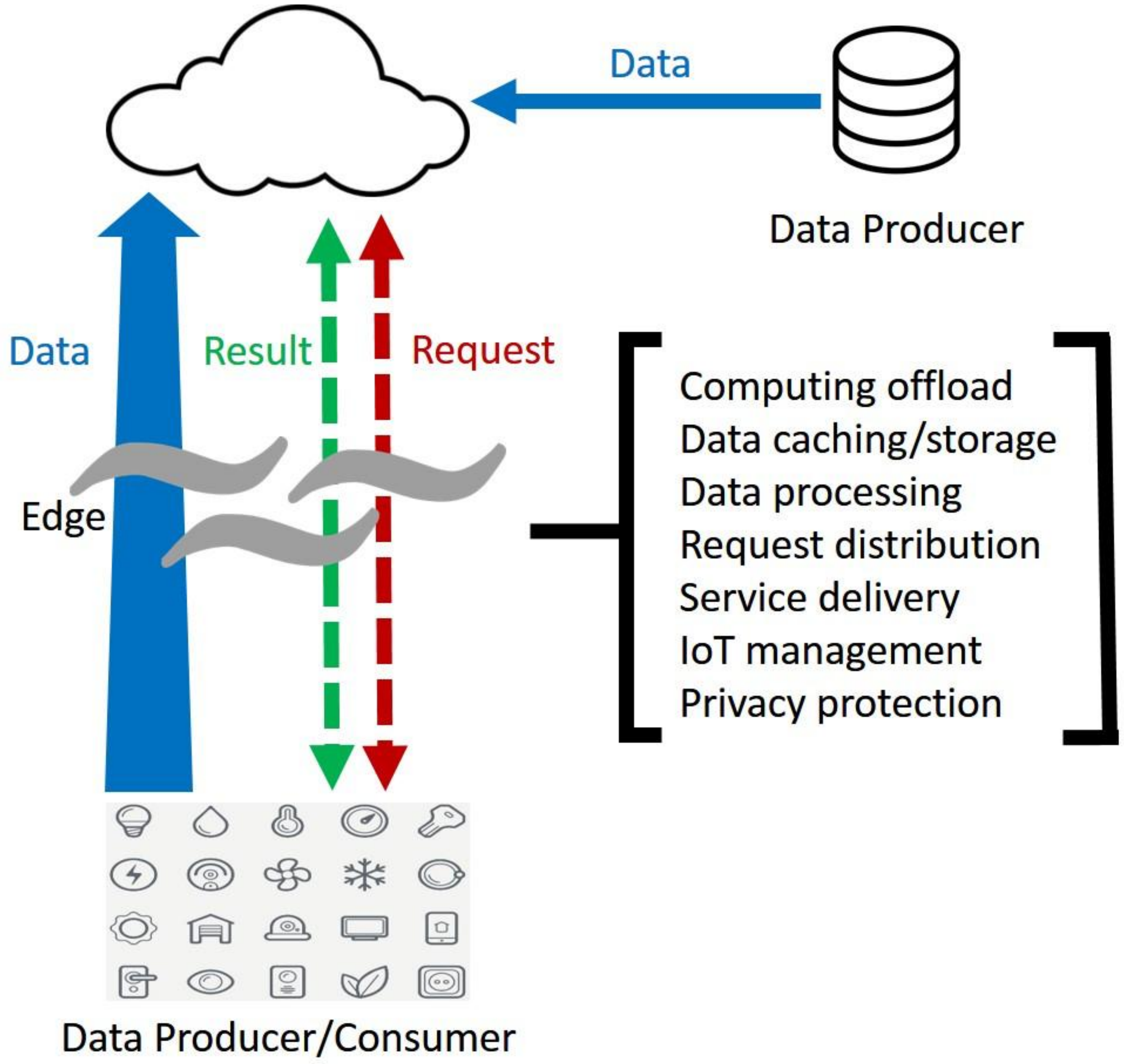
Edge Computing: Along the Path from the Data Source to the Cloud



Edge Computing: What the Edge Actually Does

Two-way computing streams

- Things are not only **data consumers** but also **data producers**.
- Computation can be performed at the edge of the network.
- On **downstream data** on behalf of cloud service & **upstream data** on behalf of IoT services.



Edge Computing Paradigm

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE internet of things journal, 3(5), 637-646.

Edge Computing: Edge-Intelligence Link

Edge Computing

On-demand compute at the data source

→

+ AI / ML at the Edge

Analyze & decide locally, autonomously

→

Edge Intelligence

Smarter, faster, more efficient systems

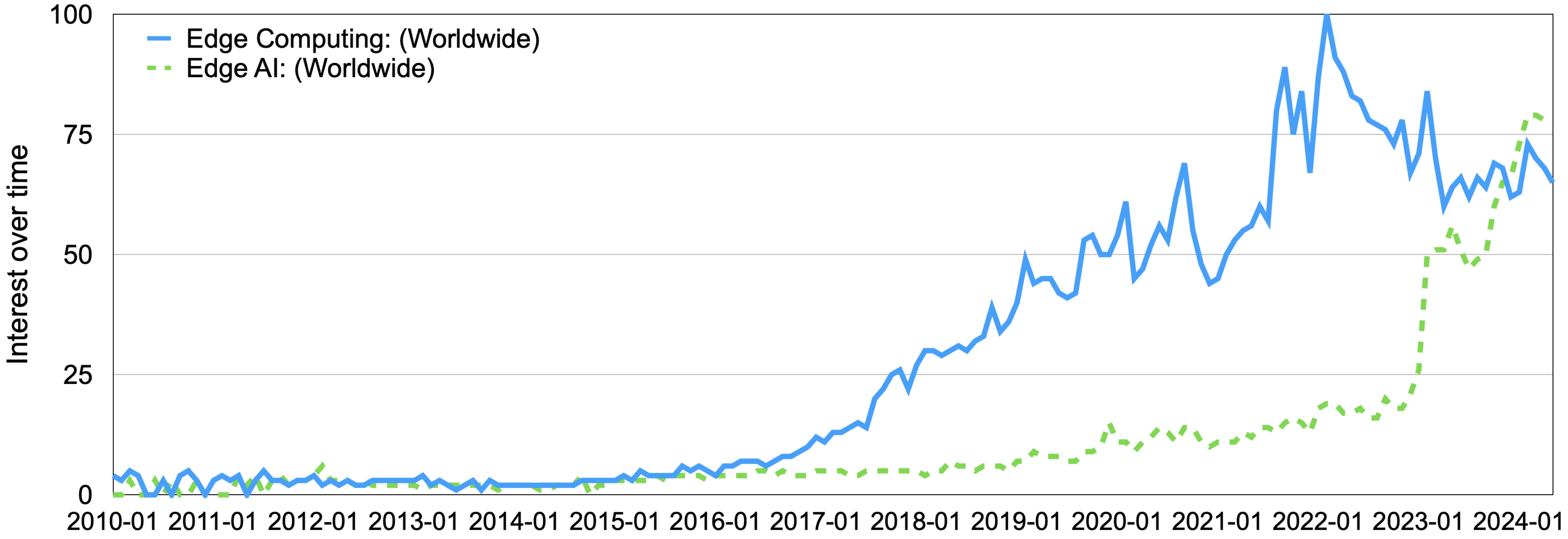
➢ **Optimization techniques**

Pruning · Quantization · Knowledge distillation — shrink AI models to run on edge devices

➢ **Application domains**

Smart surveillance · autonomous vehicles · health monitoring · industrial IoT · smart agriculture · retail

Edge Computing: Edge Growth & the Rise of “Edge AI”



Before LLMs

Focus was the paradigm itself — architectures, components, hardware/software, challenges.

After LLMs

Powerful models meet tight compute/storage limits at the edge.

The shift

Research pivots to enabling edge-based AI by tackling resource constraints.

Summary

- **Edge computing is a continuum** from data sources to cloud data centers.
- The **edge includes computing, storage, and networking resources** along this path.
- Edge computing emerged to address challenges in **big data processing and the Internet of Everything**.
- It evolved from earlier models such as **P2P, CDN, MEC, fog computing, and Cloud-Sea computing**.
- Edge computing is closely connected to the development of **edge intelligence**.

Practice

1. What is the “edge”?
2. What distinguishes edge computing from traditional cloud computing?
3. What challenges does cloud face with the data volumes from IoE devices?
4. Why is edge computing necessary in the era of the Internet of Everything?
5. Give real-world examples for each use case in the continuum (connected health, autonomous driving, smart home/city, smart manufacturing, disaster relief).
6. From the background & history, what key challenges could arise as edge computing develops?

References

- Edge Computing Consortium. Introduction of Edge Computing Consortium. http://en.ecconsort ium.net/Uploads/file/20180328/1522232376480704.pdf. Accessed: 2024-07-31.
- IBM. What is edge computing? https://www.ibm.com/topics/edge-computing. Accessed: 2024-05-11.
- OpenFog Consortium. OpenFog Reference Architecture for Fog Computing. https://www.iicon sortium.org/pdf/OpenFog_Reference:Architecture_2_09_17.pdf. Accessed: 2024-07-31.
- Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE internet of things journal, 3(5), 637-646.
- Mahadev Satyanarayanan et al. “The case for VM-based cloudlets in mobile computing”. In: IEEE Pervasive Computing 8. 4 (2009), pp. 14–23.
- Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making ai less" thirsty": Uncovering and addressing the secret water footprint of ai models. arXiv preprint arXiv:2304.03271.