LEADERSHIP. ACTION. RESULTS.℠

2035 - THE EDGE AI REVOLUTION: WHEN 80% OF AI MIGRATES TO THE EDGE

From centralized AI to distributed intelligence: Technologies, Markets and Strategic Issues 2020-2035

Target Sectors

Industry 4.0, Healthcare, Smart Cities, Automotive, IT & Telecom, Retail, Agriculture, Energy & Utilities

Author

Yannick Gablin

Director – MEA / Digital Services

ALVAREZ & MARSAL

CONFIDENTIAL – NOT FOR DISTRIBUTION

SECTION 6 –
OPERATING MODEL

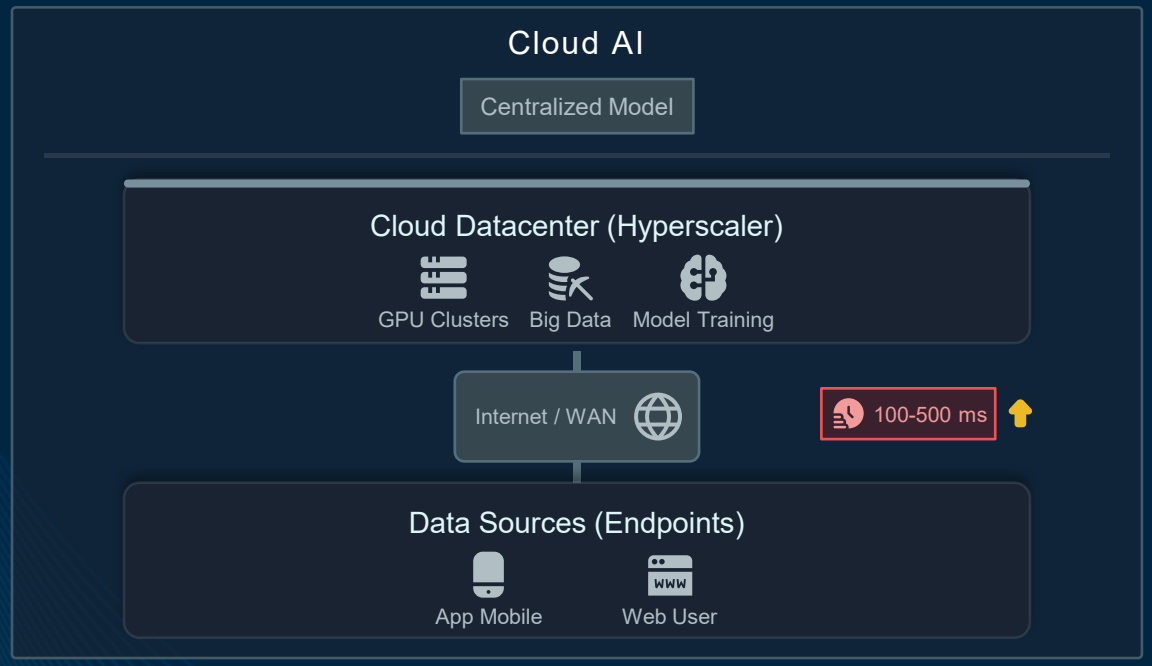
CONFIDENTIAL: NOT FOR DISTRIBUTION

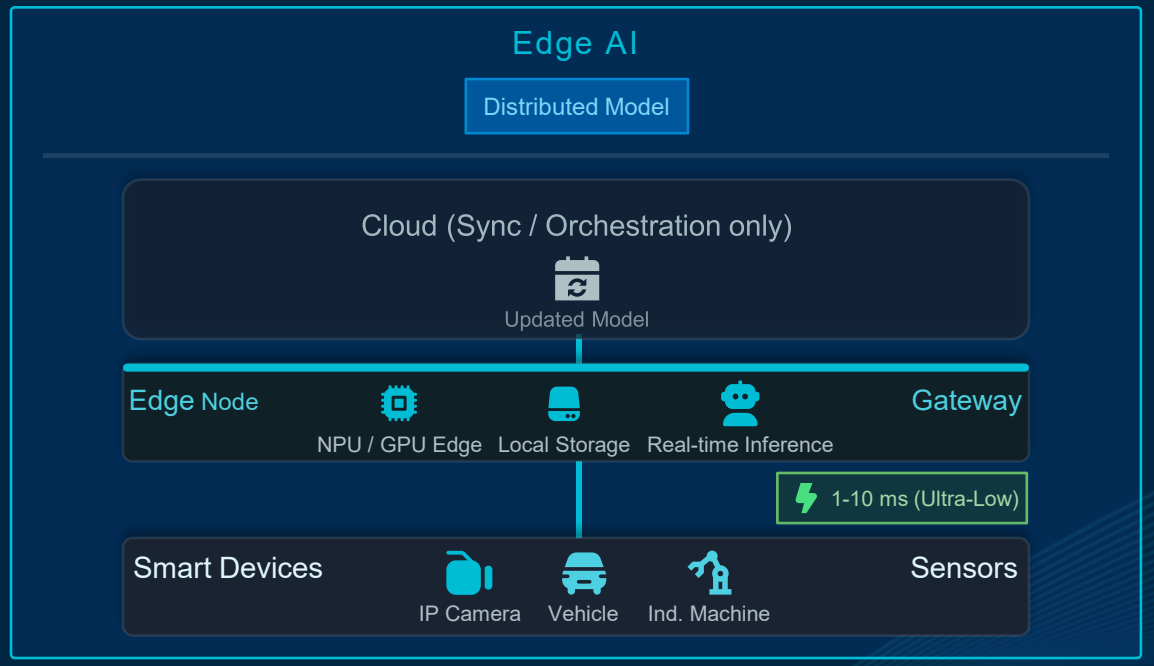
ALVAREZ & MARSAL

TECH INFOGRAPHIC

Cloud vs. Edge : 8 fundamental differences showing the shift in paradigm

Structural divergence: Massive centralization vs. real-time distributed processing





Feature	Cloud AI (Training & Massive Inference)	Edge AI (Real-Time Inference & Privacy)
Latency	Variable, depends on the network (100ms+)	Deterministic, Real-Time (< 10ms)
Bandwidth	High consumption (Upload Data + Download Result)	Efficient (Local processing, only insights come out)
Security	Transmission of sensitive data off-site	Data Residency Guaranteed (Data Remains Local)
Costs (TCO)	High OPEX (Cost per API call + Storage)	Initial CAPEX, Reduced OPEX (No egress fees)
Maintenance	Centralized, instant update for everyone	Distributed, requires fleet management (MDM/OTA)
Scalability	Near-infinite horizontal scalability (elasticity)	Scalability limited by local hardware (addition of nodes)
Energetic Constraint	Massive consumption (energy-hungry data centers)	Optimized for low consumption (battery, IoT)
Offline Operation	Complete dependence on network connectivity	Autonomous offline operation (resilience)

ALVAREZ & MARSAL

HORIZON 2035

2035 : 4 Players, 4 Revolutions - The New Edge AI Value Chain

Structural transformation of the 4 key players in the Edge AI / Cloud AI value chain

Industries

Current 2025

Cloud-first, OT/IT silos. High latency, long ROI.

Target 2035

Edge-first, local agentic AI, ROI <6 months.

- **Organization**
  OT/IT Silos → CoE Edge AI Convergence
- **Processus**
  Reactive maintenance → CI/CD MLOps Edge
- **Tools**
  SCADA/Cloud → Unified Platforms

- Decentralized MLOps "at the foot of the machine"
- Data Sovereignty & Zero Trust

Connectivity

Current 2025

Centralized NSA 5G. Limited MEC, best-effort.

Target 2035

5G SA/6G distributed. On-demand slicing, guaranteed SLAs.

- **Organization**
  Net/IT Silos → DevNetOps Teams
- **Processus**
  Manual provisioning → Zero-touch auto
- **Tools**
  Legacy OSS → Cloud-native Orchestration

- Scaled MEC & Compute-Network Conv.
- B2B2X Revenue-Share Models

Cloud AI

Current 2025

Centralized. High egress costs, variable latency.

Target 2035

Hybrid Cloud-Edge. Federated Learning, orchestration dist.

- **Organization**
  Centralized → Dist. Edge Teams
- **Processus**
  Monolith deployment → Federated Learning
- **Tools**
  IaaS/PaaS → Edge Extensions (Arc/Outposts)

- Sovereign "Cloud-to-Edge" Extension
- AI/Agent Model Marketplace

Edge AI

Current 2025

Silicon-centric. Fragmented SDKs, pilots, one-off revenues.

Target 2035

Full-Stack Platform. Silicon + Runtime + OTA. Recurring Rev.

- **Organization**
  Silicon focus → Full-stack teams
- **Processus**
  Custom development → Unified SDKs & OTA
- **Tools**
  Fragmented → Unified Toolchains (ONNX)

- Standardization & Portability Models
- Hardware/software "as-a-Service" model

Cross-Functional KPIs Targets 2035

- 5 – 10ms
  Target Latency
- ↓ 40 – 70%
  Cost per Inference
- > 70%
  Part Edge Inference
- 100%
  Zero Trust Native

ALVAREZ & MARSAL

END USERS

Industries : ROI < 6 Months or Failure - Edge AI Becomes Non-Negotiable

Transforming the model : From cloud centralization to distributed, edge-first intelligence

2025 Cloud-Centric Model

Architecture & Infra

Cloud-first approach. Raw data sent to the cloud for analysis. Rigid silos between OT (Operations) and IT (Information).

Value Creation

Occasional efficiency gains. Basic preventive maintenance. Few monetized data-driven services.

Organization / Processes / Tools

Organization : Separate OT/IT silos. Small, centralized IT data science teams.

Processus : Reactive maintenance. Long deployment cycles (months). Dispersed POCs.

Tools : Legacy SCADA/MES. Cloud ML (SageMaker, Azure ML). Manual integration.

Costs & CAPEX

High cloud OPEX (storage, egress). CAPEX limited to pilot projects (POC). Hidden integration costs.

Ecosystem

Dependence on Hyper scalers (AWS, Azure) and traditional IT Integrators.

As Is KPIS

80 – 200 ms	Silos	12 – 18 months	< 10%
Average Latency	OT Security	Average ROI	Local Processing

IT/OT Convergence

MLOps ‘At the Edge’

Zero-Trust Native

Data Governance

2035 Edge-First Model

Architecture & Infra

Edge-first approach. Local NPUs/IPCs, Embedded Digital Twins. Decentralized local MLOps.

Value Creation

X-as-a-Service models (Quality, Safety). Monetization of industrial data. Autonomous operations.

Organization / Processes / Tools

Organization : OT/IT convergence. Integrated MLOps teams. Edge AI Center of Exc.

Processus : Predictive maintenance. CI/CD templates. Continuous OTA deployment.

Tools : Unified Edge Platforms (Jetson/OpenVINO). Kubeflow Edge. Private 5G.

Costs & CAPEX

Overall TCO ↓ 50% (less cloud/network). CAPEX modernization of lines & Edge Nodes.

Ecosystem

OEM Edge (NVIDIA, Qualcomm), Telcos (Private 5G), Hyperscalers (Hybrid).

Target KPIS

< 10 ms	Unified	< 6 months	> 80%
Target Latency	OT Security	Average ROI	Local Processing

ALVAREZ & MARSAL

Telcos 2035 : From Silent Pipes to Edge AI Orchestrators (+30% ARPU)

Transforming the model : From pure connectivity to monetized Edge & API services

2025 "Dumb Pipe" model

Architecture & Infra

5G NSA (Non-Standalone), Cloud-centric Backhaul. Very limited MEC (Multi-Access Edge). Network slicing exists but is poorly monetized.

Value Creation

B2B ARPU under pressure. Price competition on data. Basic IoT offerings with low margins.

Organization / Processes / Tools

- **Organization :** Silos Network/IT, OSS/BSS legacy
- **Processus :** Manual provisioning, best-effort SLA
- **Tools :** RAN/Core trad. (Ericsson/Nokia)

Costs & CAPEX

High network OPEX. Heavy 5G investments. Complex slicing management. Expensive energy/data transport.

Ecosystem

Hyperscalers (limited area partnerships), traditional RAN equip. manufacturers.

Current KPIS

20 – 50 ms	< 10%
5G Latency	MEC Attachment

- Marketplace APIs
- Edge Co-investments
- Revenue-Share Models
- Network-as-a-Service

2035 Edge-Native Service Model

Architecture & Infra

5G SA / 6G. Massively distributed MEC, local UPF. Nationwide Edge POPs. Exposed and standardized network APIs.

Value Creation

API monetization (on-demand QoS). Vertical slicing (NPN factories/automotive). Edge PaaS. B2B ARPU +20-30%.

Organization / Processes / Tools

- **Organization :** Network-Cloud convergence, API products.
- **Processus :** Zero-touch provisioning, real-time SLA
- **Tools :** O-RAN, K8s cloud-native, NWDAF

Costs & CAPEX

Automation (AIOps). Energy efficiency. CAPEX targeted at local MEC/Cores rather than transportation.

Ecosystem

Hyperscalers (Wavelength, Outposts), Vertical ISVs, OT Integrators.

Target KPIS

5 – 10 ms	30 – 40%
Target Latency	MEC Attachment

ALVAREZ & MARSAL

AWS, AZURE, GCP

Hyperscalers 2035 : Why AWS, Azure and Google are investing heavily in the Edge

Transforming the model : From massive centralization to distributed Cloud-Edge hybridization

2025 Centralized Model

Architecture & Infra

Centralized training and inference. Massive GPU clusters. Latency >50ms. Sovereignty and lock-in challenges.

Value Creation

High IaaS/PaaS/MLaaS consumption. High egress fees. Margins under pressure (GPU/energy cost).

Organization / Processes / Tools

Organization : Centralized teams, IaaS/PaaS focus, regional silos.

Processus : Centralized deployment, egress fees, monolithic models.

Tools : TensorFlow, PyTorch, SageMaker, Azure ML, Vertex AI.

Costs & CAPEX

CAPEX for massive data centers. OPEX for energy and cooling that are constantly increasing.

Ecosystem

Nvidia (GPU), Intel (CPU), Databricks (ML Ops). Vendor lock-in.

As Is KPIS

> 50 ms	IaaS / PaaS	High	Complex
Average Latency	Billing	Egress Fees	Sovereignty

- Sovereign Offers
- Federated Learning
- Co-managed SLA
- Marketplace Agents

2035 Distributed Hybrid Model

Architecture & Infra

Hybrid cloud-edge. Low-latency gateways, regional caches. Serving on-premises/sovereign.

Value Creation

Inference token-based billing. Vertical platforms (Industrial / Automotive / Healthcare). Model governance services.

Organization / Processes / Tools

Organization : Distributed edge teams, focus on vertical platforms, geo-distributed pods.

Processus : Hybrid deployment, federated learning, distributed models, multi-site orchestration.

Tools : Edge extensions (AWS Wavelength, Azure Arc, GCP Distributed Cloud), multi-cluster Kubeflow, ONNX Runtime.

Costs & CAPEX

Cost/inference ↓ 40%. Optimized egress (↓ 30-50%). Energy efficiency (liquid cooling).

Ecosystem

Open and co-managed ecosystem : Telcos (MEC), Edge silicon vendors

Target KPIS

< 20 ms	Token	Native	↓ 30-50%
Edge Bridge SLO	Billing	Sovereignty	Egress Fees

ALVAREZ & MARSAL

SILICON & PLATFORM

From Silicon to Software : How Edge AI is Reinventing Chipmakers

Transforming the model: From Silicon-Centric to Full-Stack Edge Platform

2025 Silicon-Centric Model

Architecture & Infra

Focused on **hardware (silicon)**. Fragmented SDKs and compilers. Focus on demos and technical drivers.

Value Creation

Transactional hardware sales (chips/SoCs). Limited SDK licenses. Ad-hoc professional services for integration.

Organization / Processes / Tools

Organization : Silicon-centric teams, vendor-fragmented SDKs, responsive support.

Processus : Custom development, manual integrations, long release cycles.

Tools : CUDA, OpenVINO, TensorRT (siloed), proprietary compilers.

Costs & CAPEX

CAPEX : Silicon R&D 15-20% of revenue, High BOM (fabs, tapeouts). OPEX: Technical support, certifications (ISO 26262, ASIL).

Ecosystem

OEM device partners, niche integrators, open-source developer communities.

As Is KPIS	20 – 40 ms	< 15%	Variable	Complex
	Average Latency	Soft. Part	Security	Portability

Toolchain Standard.

OEM Co-design

Secure Supply Chain

Compliance (ISO26262)

2035 Full-Stack Platform Model

Architecture & Infra

Full-Stack approach (Silicon + Runtime + Unified SDK + MLOps + OTA). Certified vertical references (Safety/Medical).

Value Creation

Recurring Software Mix >40% (Licenses, Support, Marketplaces). Managed Edge Services.

Organization / Processes / Tools

Organization : Full-stack teams (HW+SW), unified platform, proactive Customer Success.

Processus : Standardized SDKs (ONNX), OTA deployment, native CI/CD, marketplaces

Tools : ONNX Unified Runtime, standardized toolchains, native MLOps, TEE/SE.

Costs & CAPEX

CAPEX: Full-Stack R&D 12-18% of revenue. Optimized BOM ↓15-20% (adv. nodes, co-design). OPEX: Cloud platforms, Cust. Success, exp.certifications.

Ecosystem

Hyperscalers (Compilers/Backends), Telcos (MEC), Vertical ISVs, Indust. OEMs.

Target KPIS	< 10 ms	> 30 – 40%	TEE / SE	> 80%
	Target Latency	Soft. Part	Security	Portability

ALVAREZ & MARSAL

LEADERSHIP. ACTION. RESULTS.

ALVAREZ & MARSAL

CONFIDENTIAL – NOT FOR DISTRIBUTION

Alvarez & Marsal Holdings, LLC. All rights reserved. ALVAREZ & MARSAL®, ® and A&M® are trademarks of Alvarez & Marsal Holdings, LLC.

© Copyright 2026