Edgenuity: The Intersection of Edge and AI

Presented by Jeff White – Dell Technologies
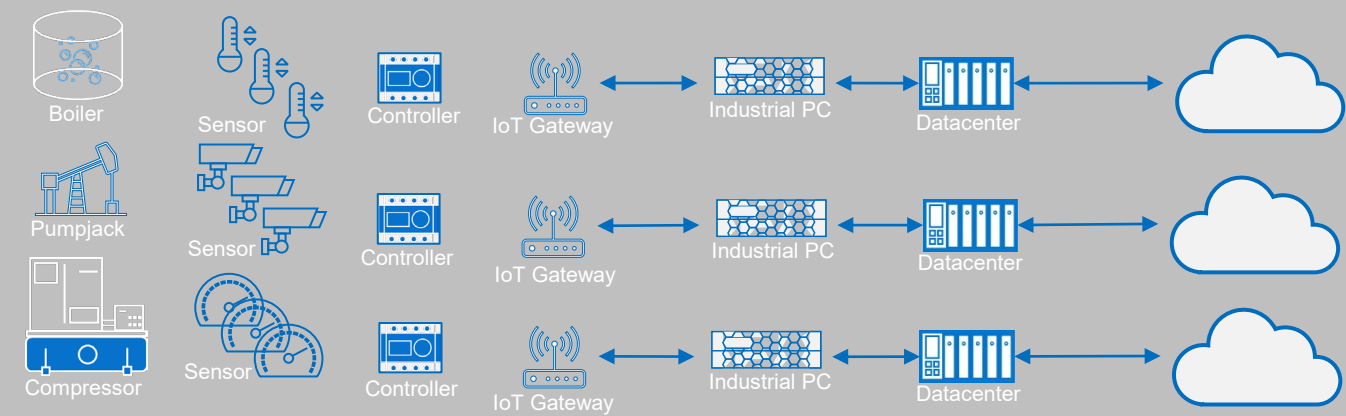
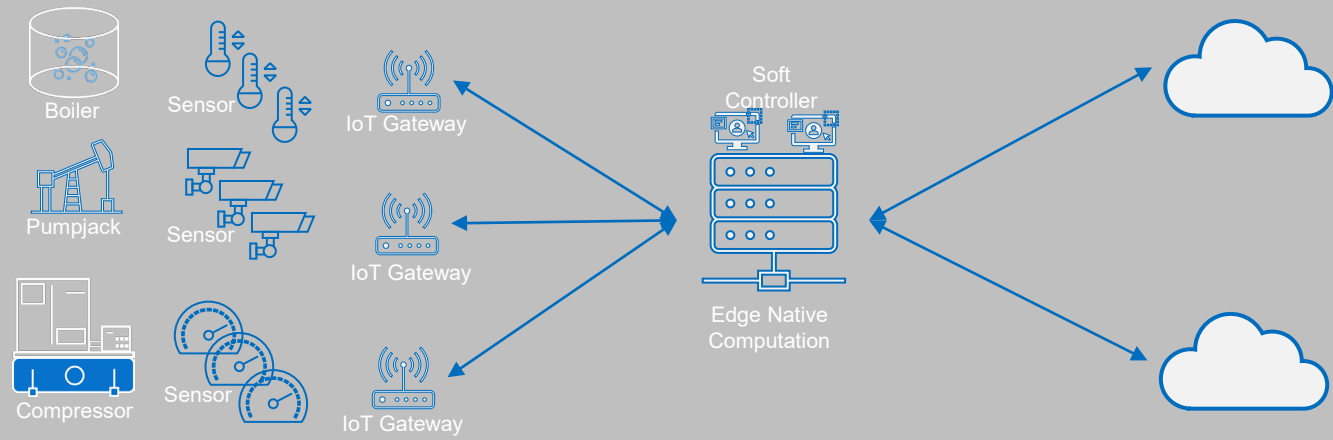# Modern Edge Computation

Present

1) Bespoke Hardware Control.
2) Fragmented Systems.
3) Limited Data Sharing.
4) Inflexible To Change.
5) Closed Systems.
6) Onsite Expert Support Required.
7) Individual Security Policy.

Future

1) Common System Platform.
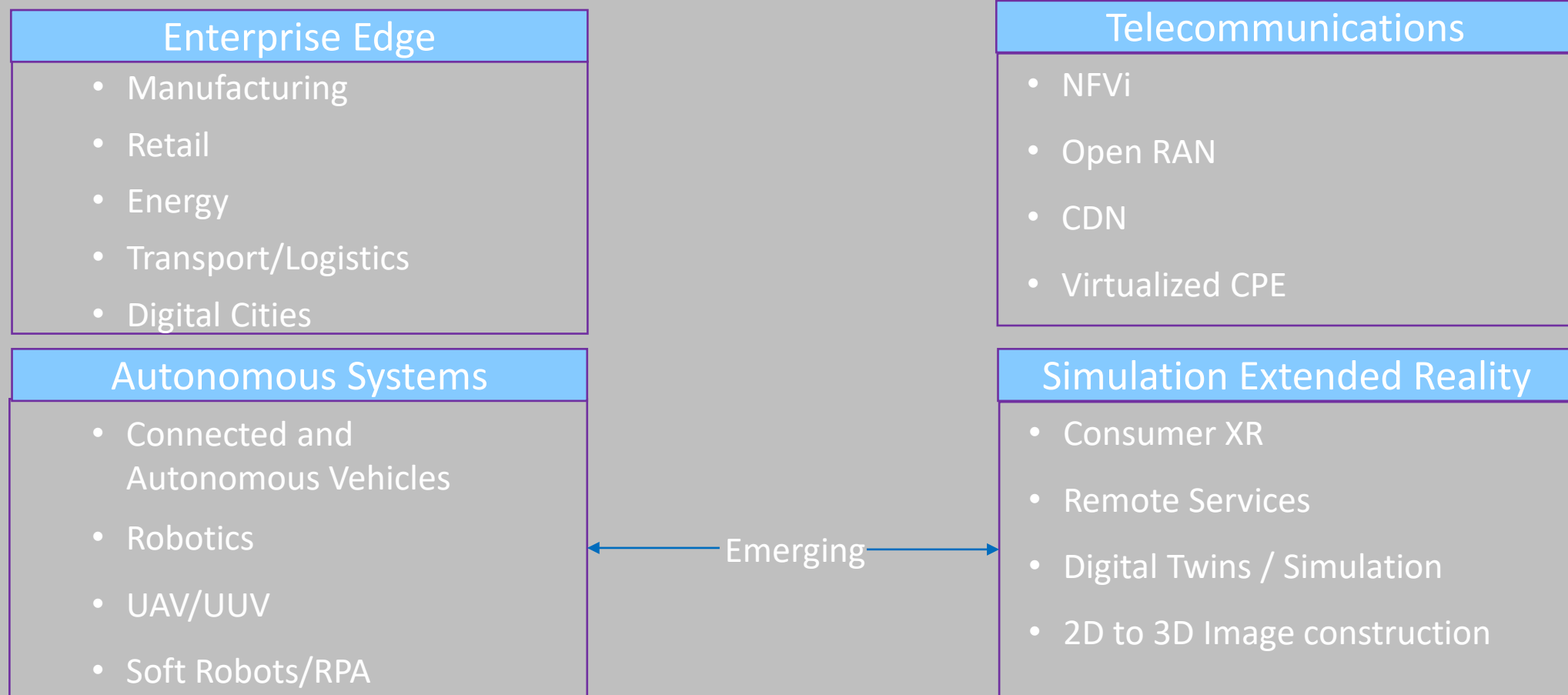2) Higher Computational Power.
3) Shared Data/Networking.
4) Inherent MultiCloud Native.
5) Open System.
6) Remote Support.
7) Unified Security Policy.
8) Software Defined Concurrency.
9) Power Efficient.
10) Dynamic Twin/AI/ML Enabled.

Modern Edges must secure, simplify and automate complex operations and technology from Edge Through Cloud

COMPUTE, MEMORY,
AND STORAGE SUMMIT

Enterprise Verticals and Telco/Emerging Use Cases

Enterprise Edge
- Manufacturing
- Retail
- Energy
- Transport/Logistics
- Digital Cities

Telecommunications
- NFVi
- Open RAN
- CDN
- Virtualized CPE

Autonomous Systems
- Connected and Autonomous Vehicles
- Robotics
- UAV/UUV
- Soft Robots/RPA

Simulation Extended Reality
- Consumer XR
- Remote Services
- Digital Twins / Simulation
- 2D to 3D Image construction

← Emerging →

Public Sector / Defense Applications Require Aspects of All Above

SNIA COMPUTE, MEMORY, AND STORAGE SUMMIT

AI Technology Frameworks

- Artificial Intelligence is an area of Computer Science that utilizes functions that can be trained to perform a variety of statistical/probabilistic tasks classification, ranking, regression, clustering graph/geometry discovery, density estimation, etc…. AI is an universal function approximator, and….

- AI can be used to automate **intelligent behavior**, but defining intelligent behavior is very difficult.

- AI comes in different technology frameworks all of which are used in the Edge:

Machine Learning

$$f(x) = \sum_{i=1}^{n} \alpha_i * K(x_i, x) + b$$

Discriminative

$$P(Y|X)$$

Symbolic

$$\forall x \forall y (Rx \odot Cy < Qxy)$$

Reinforcement Learning

$$q\pi(s, a) = E\pi[R_t+1 + \gamma q\pi(S_{t+1}, A_{t+1})|S_t = s, A_t = a]$$

Generative (≠ LLM)

$$P(X,Y)$$

COMPUTE, MEMORY, AND STORAGE SUMMIT

AI Technology for Edge Use Cases

Edge Use Cases with AI Relevance (Not an exhaustive list)									
Quality Inspection	Worker Safety	Overall EE	Robotics Control / RPA	Predictive Mtce	Loss Prevention	Security & Safety	Inventory Analytics	XR/VR	Digital Twin

AI Capabilities

Machine Vision   **Near Real Time Analytics**   **Autonomous System Support**   **Data Processing / Analysis**

AI Technology

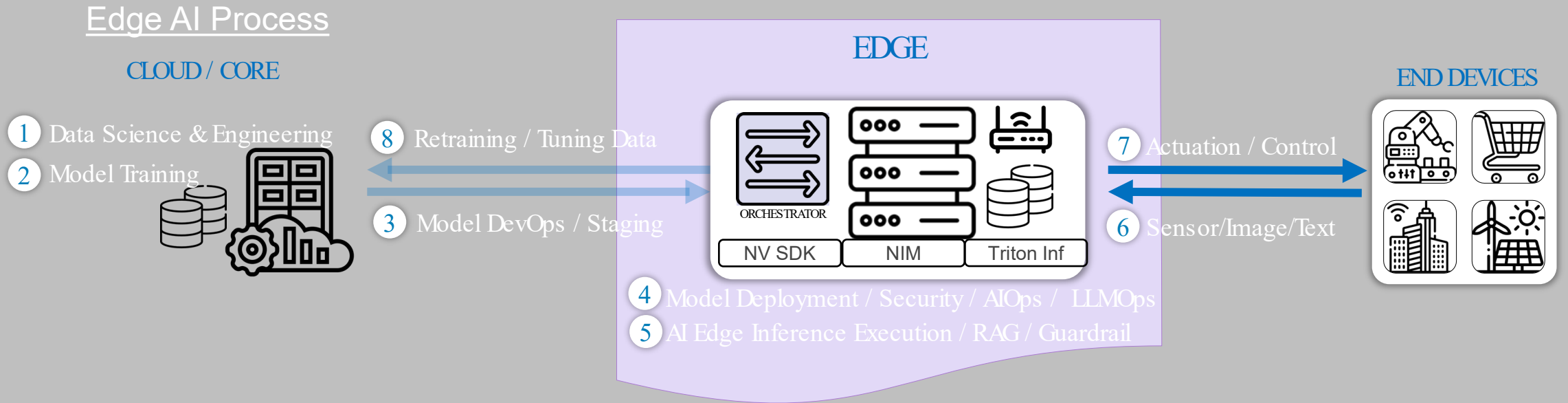CNN / ViTransformer / ResNet Multi-sensory Inference

KNN / SVM / Bayes / Spectral / RF / Gaussian

RL / Transformer / Causal Multimodal

Autoencoder / Transformer / SVD / RF / PCA

Edge AI supports business critical, safety & online RT/NRT operations
AI IN-ON-FOR

COMPUTE, MEMORY, AND STORAGE SUMMIT

Edge AI Operational Support Functions

Edge AI Process

CLOUD / CORE

1 Data Science & Engineering
2 Model Training

8 Retraining / Tuning Data

3 Model DevOps / Staging

EDGE

ORCHESTRATOR

NV SDK | NIM | Triton Inf

4 Model Deployment / Security / AIOps / LLMOps
5 AI Edge Inference Execution / RAG / Guardrail

END DEVICES

7 Actuation / Control

6 Sensor/Image/Text

Edge AI Operations Functions

A AI Edge Accelerator Management

B AI Edge Model Management
C AI Edge Inference Operations
D AI Model Security Operations
E AI Distributed Training / Tuning / Augmentation

F Edge Data Management

SNIA COMPUTE, MEMORY, AND STORAGE SUMMIT

Future of Edge AI

- Multimodal Generative and Discriminative techniques for Autonomous Systems and Edge Platform Process control.

- Communicating Agents based systems with Machine Reasoning ($\neq LLM$) generative techniques and system models.

- Interactive distributed training / fine tuning / model augmentation.

- AI security to protect the models and protect the systems from the models.

COMPUTE, MEMORY, AND STORAGE SUMMIT

Please take a moment to rate this session.

Your feedback is important to us.

Jeff White

jeff.white@dell.com