# The Shift to Edge Data Processing

Christian Kern



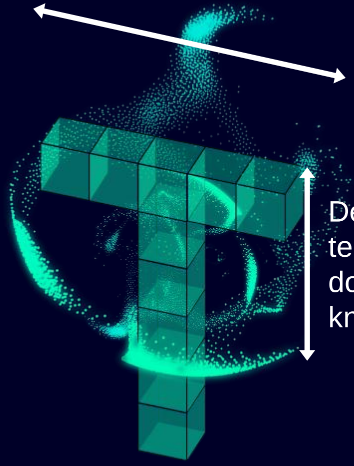
SIEMENS

# Who are we?

| Digital Industries | Smart Infrastructure | Mobility | Siemens Healthineers[1] |
|---|---|---|---|
|  |  |  |  |

Leveraged across
all Siemens businesses

Predevelopment
Research
Consulting
…



Deep
technological
domain
know how

1 Publicly listed subsidiary of Siemens; Siemens' share in Siemens Healthineers: 75%

## SIEMENS Technology

**2,100**
Employees worldwide

**1,700**
Researchers

Total@SIEMENS:
42,500 R&D employees
311,000 employees

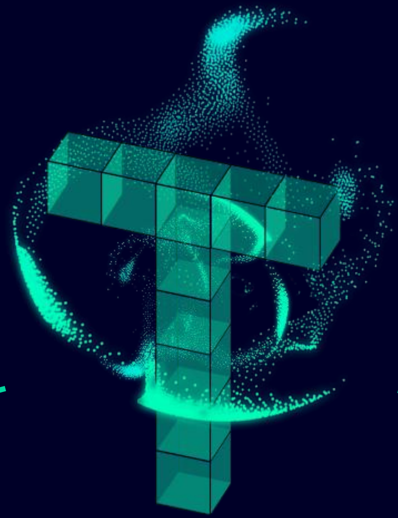
**430**
Patent experts



SIEMENS

# Siemens Technology - Internal Structure

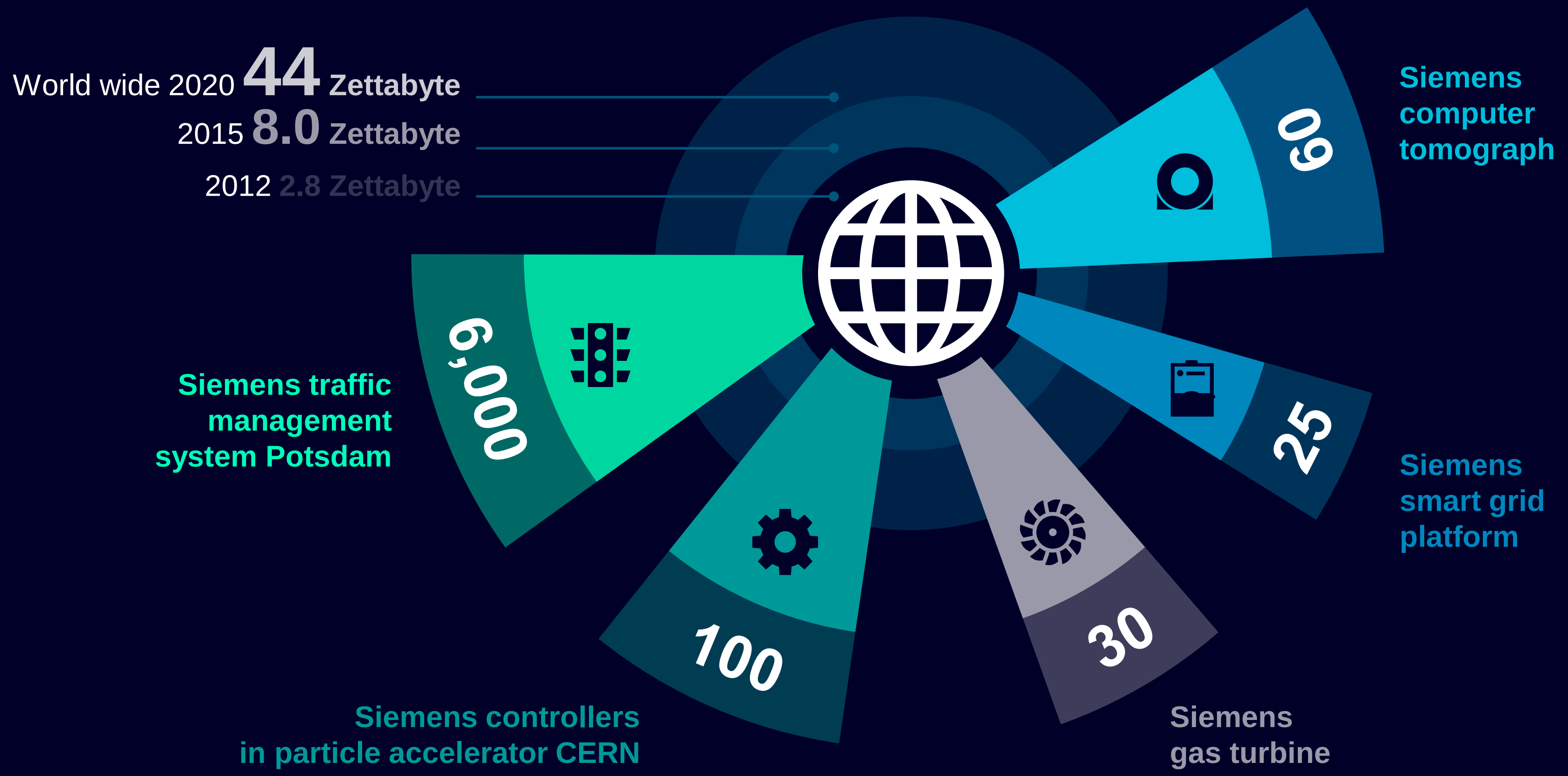
**Company Core Technologies (CCTs)**



- Software Systems & Processes
- Sustainable Energy & Infrastructure
- Connectivity & Edge
- Additive Manufacturing & Materials
- Power Electronics
- Simulation & Digital Twin
- User Experience
- Integrated Circuits & Electronics
- Data Analytics & Artificial Intelligence
- Cybersecurity & Trust
- Automation

# Installed base produces tremendous amounts of data that can be leveraged for new applications.

The amount of data produced worldwide and by Siemens products[1]




1 Gigabytes per day

| | |
|---|---|
| 100 | **high frequency variables** and more |
| >3 million | **sensor values** per hour |
| >1.8 GB | **data processed** locally per hour |

| | |
|---|---|
| ~100TB/h | **data generated** all our PLCs |



SIEMENS

# Bridging the Gap Between Field and Cloud Level

## Centrally managed local compute power close to the data sources

**Cloud**

- Management of edge and field devices, edge apps, app store
- Machine learning, data consolidation, global analytics

MindSphere | amazon web services | Alibaba Cloud | Azure | IBM Cloud | Google Cloud

**Edge**

**Local processing**
of large amounts of data using complex algorithms

**Remotely managed**
device configuration and updates via central services

**Extensible**
functionality during entire lifetime of the system (apps)

Backend/cloud connectivity

0110110
1011101
0011011
0101010
1100101

Flexible data processing and storage
for analytics, perception, anomaly detection, …

Field connectivity

**Open interfaces**
for partners and third-party app developers / providers

**Separation**
of (critical and non-critical) apps via HW/SW isolation

**Field**

X-Ray



SIEMENS

# Benefits and Opportunities of Edge Computing

## Meet the needs and boundary conditions of industrial applications


Real-time and high-frequency data processing

Improved security and privacy of critical data

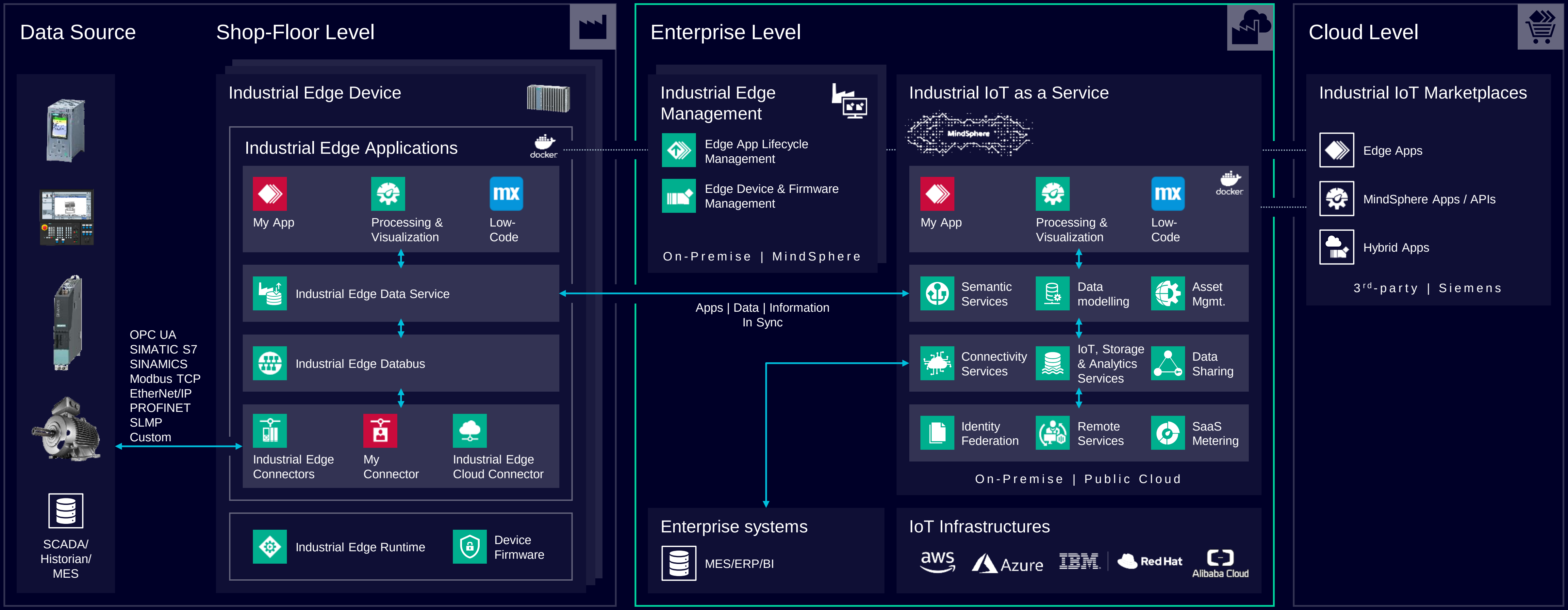
High availability independent of cloud connectivity



SIEMENS

# Siemens Industrial Edge

## Open technology stack from edge to cloud for factory automation




Platform services & apps   Individual apps



SIEMENS

# Internet of Things and Edge Computing from an Industrial Perspective
Selected examples for data-driven value generation

## Quality Assurance

Factory Automation

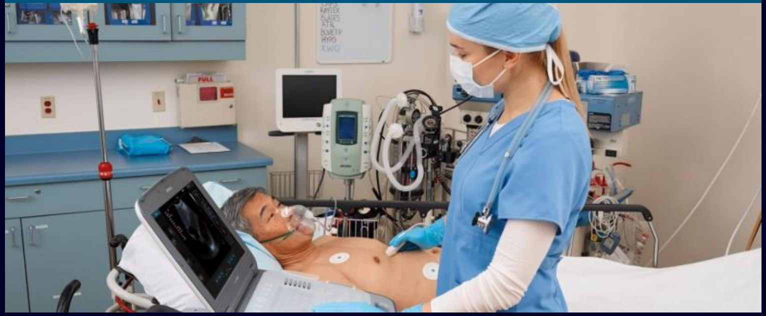
## Diagnostics and Failure Detection

Energy Grids

## Predictive Maintenance

Mobility & Transportation

## Healthcare



Hospitals and Medical Centers

## Condition Monitoring

Power Plants

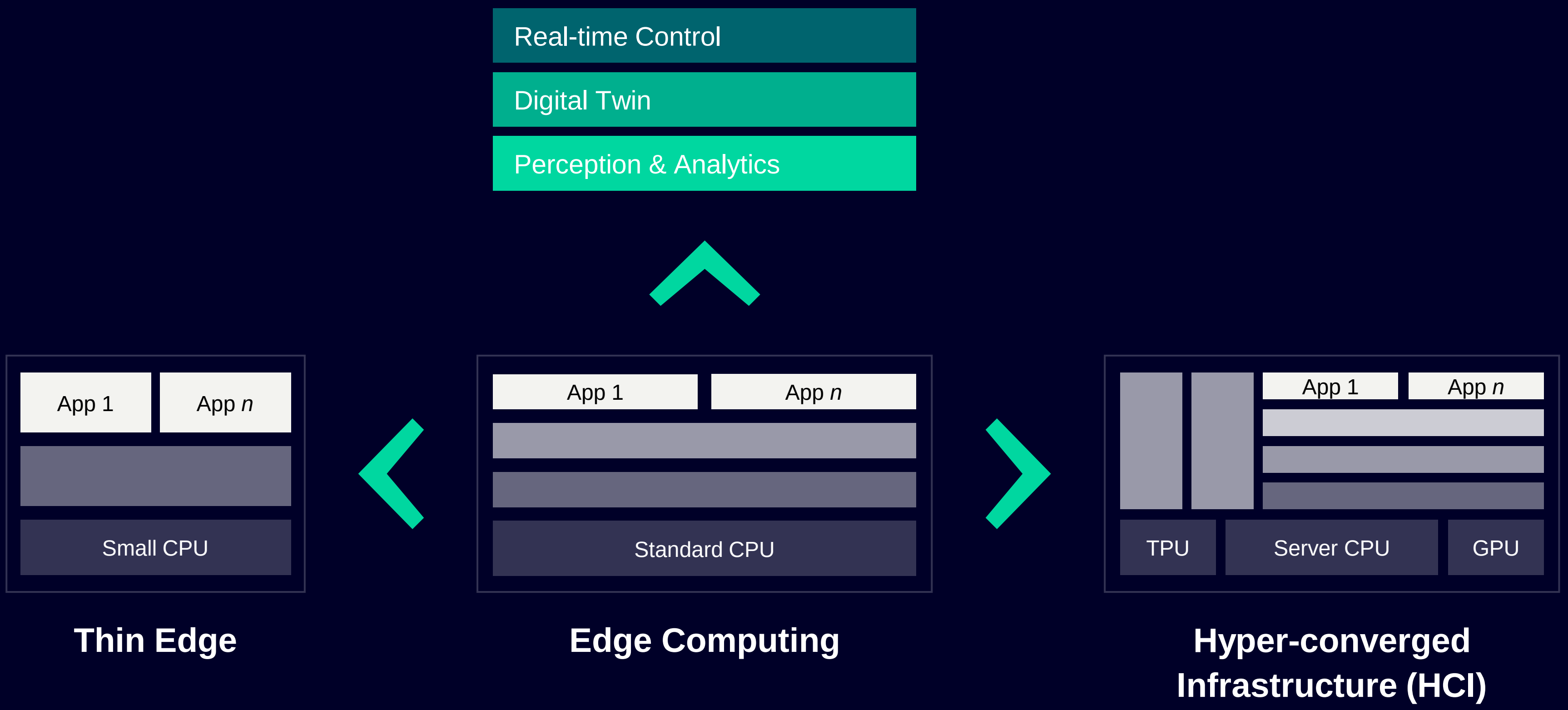
## Asset Management

Smart Cities



SIEMENS

# Edge Computing Evolution
## Extended functionalities on a wide range of device classes




“Through 2028, Gartner expects a steady increase in the **embedding of sensor, storage, compute and advanced AI capabilities in edge devices**.”

### R&D TOPICS

- Zero-trust security models and mass provisioning
- Legacy applications and edge-native workloads
- Normalization of IoT data using web-based formats
- Real-time operating systems and networking
- Mixed-criticality workloads on common infrastructure
- Hardware accelerators for AI and federated learning



SIEMENS

# Thank you!

**Dr. Christian Kern**
Siemens AG
T CED SES-DE
Otto-Hahn-Ring 6
81739 Munich
Germany

+49 173 3142018

christian.kern@siemens.com



SIEMENS

# MLOps on the Edge

Thomas Kaufmann



SIEMENS

# About

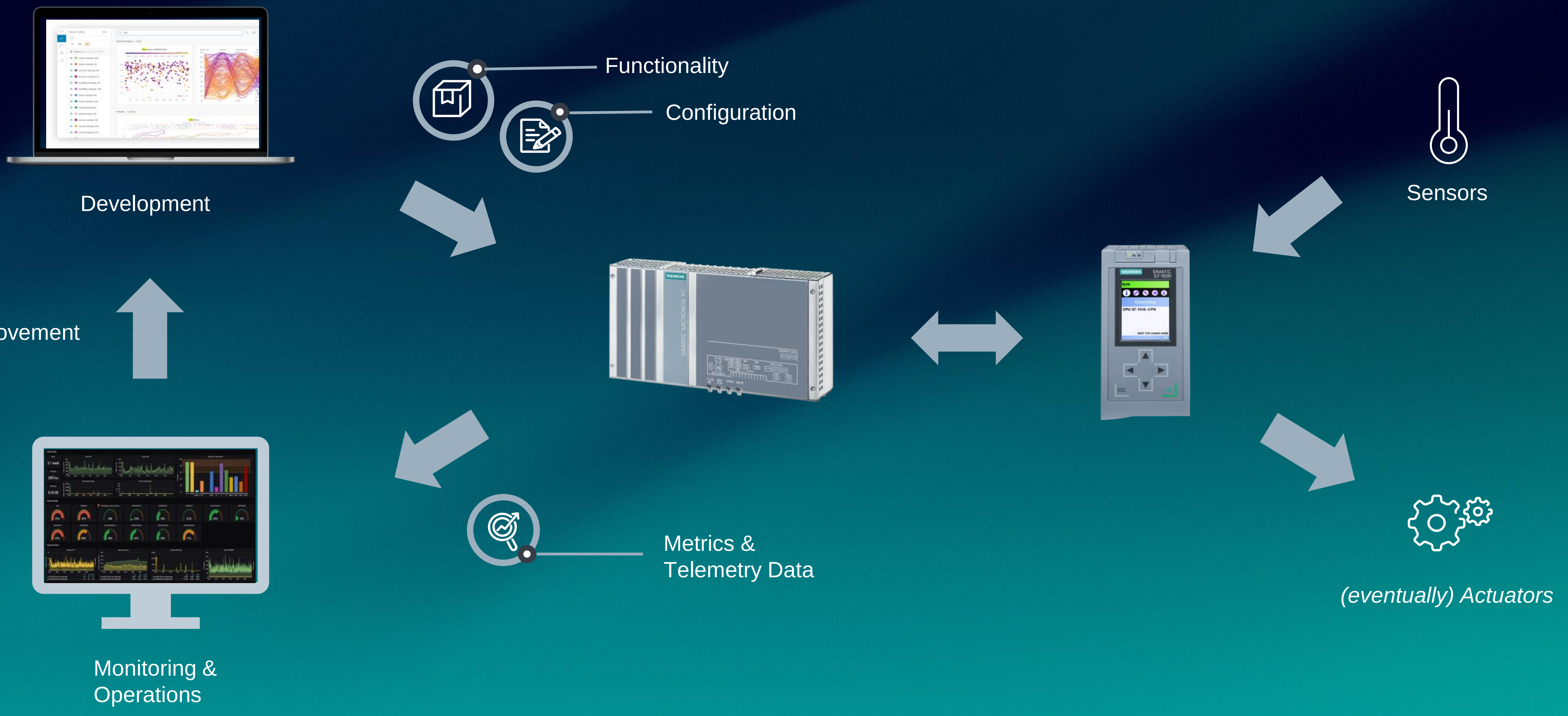
- **Siemens Technology Vienna**
- **Group: Distributed AI Systems (Data Analytics & AI)**
- „Try to explore the interplay between ***data-driven applications*** and **edge-computing“**
    - Applications of Computer Vision in Manufacturing
    - Federated Learning in Industrial Environments
    - Edge computing in closed loop systems, control systems etc.



SIEMENS

# OpenLab: Edge Computing Use-Case

*Model Predictive Control of Air Handling Units* [1]

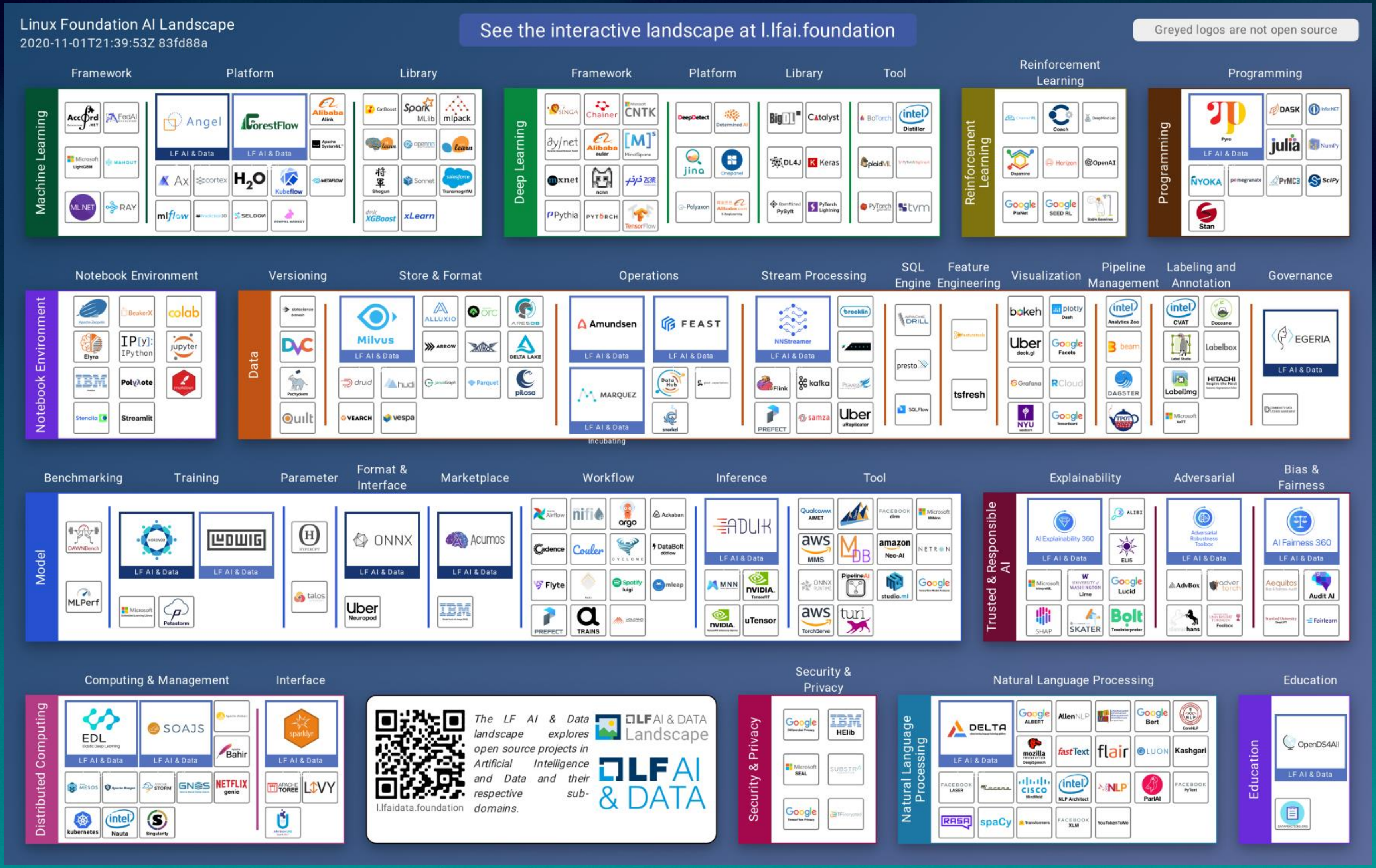



[1] Ghawash, Faiq & Hovd, Morten & Schofield, Brad & Monteiro, Diogo. (2022). Model Predictive Control of Air Handling Unit for a Single Zone Setup.




SIEMENS

# MLOps Landscape




https://ml-ops.org/content/state-of-mlops




SIEMENS

# How this has been addressed by Siemens recently…

**1) GPU-enabled Hardware**

- IPCs
- Edge Devices (Boxes)

**2) MLOps Stack for the Industrial Edge Platform**

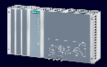
 

SIEMENS

# SIMATIC IPC

### 847E with 850W PS and Nvidia RTX4000

- Visual inspection, all in one (GPU, frame grabber),
- Predictive maintenance



### 427E with Google Coral accelerator

- Enabling our established customer base, e.g. machine builders to add AI applications (predictive maintenance, visual inspection) without changing their footprint
- Approaching new customers, e.g. retail who previously did not work with Siemens

### IPC1047E with 1 or 2 Nvidia RTX4000 or Nvidia RTX5000

### IPC547J with 850W PS and Nvidia RTX4000

- Visual inspection, all in one (GPU, frame grabber)
- Predictive maintenance

### IPC520A TENSORBOX

- Visual inspection, predictive maintenance
- Based on Nvidia Jetson NX module

**be curious, there is more to come………**

# Industrial AI Portfolio Overview
## General approach for creating AI Applications

### Cloud level

**Industrial Edge Hub (IEH)**

aws, MindSphere, Azure, Google Cloud

**AI Software Development Kit**

*Train, validate and package models into AI standard format*

### Factory level

**Industrial Edge Management (IEM)**

On-Premise

Openness API

**AI Model Manager**

*Enables IEM by an AI Fleet Management capability and standard AI Models deployment*

**AI Model Monitor[1]**

Keep track of inference performance, evaluate data quality metrics and potential data drifts

**AI Model Deployer**

*Model conversion and optimization for HW accelerated automation devices (e.g. TM NPU, Coral-TPU)*

### Field level

**Industrial Edge Device (IED)**

Edge Runtime

Openness API

**AI Inference Server**

*Siemens Standard Inference Server for Running AI @ Edge*

**Connectors**

*S7, PROFINET, OPC/UA, Vision*

Data

**Automation Equipment**

[1] Descriptive titles, no portfolio or product names

SIEMENS

# Scalable Use-Cases with *Templates*

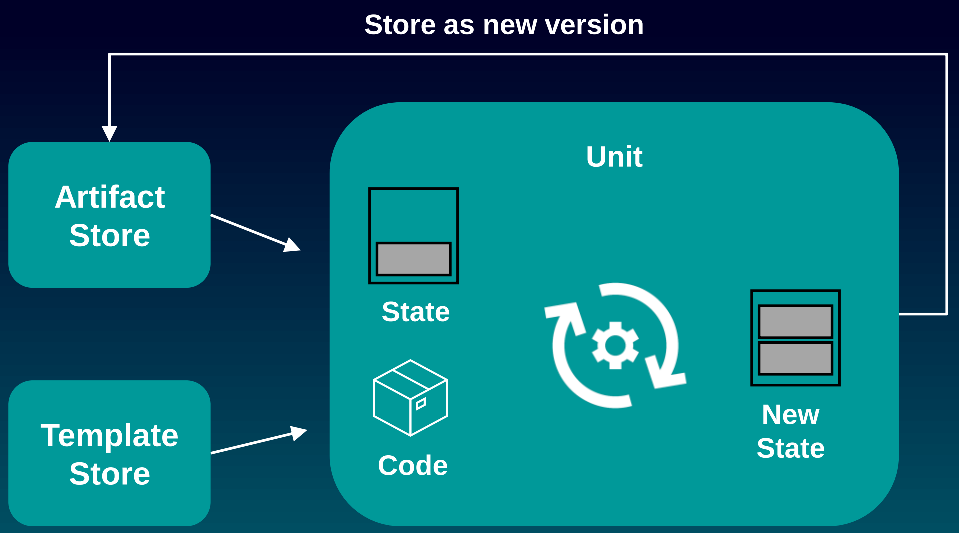
# Objectives

- Solve *repetitive*, *moderately complex* use-cases (compared to use-cases in CERN)
- Maintainable „large-scale“ rollout models
- Support heterogeneous platforms
  - Native cloud, k8s, fog, edge, …
- Widely distributed (edge) systems
  - Isolated devices
  - Not necessarily always connected → refinement on the edge
- Managed execution not necessarily involving a „user“
- „*MLOps*“ Aspects
  - Configurations, Parameters, Versioning, Immutability & Locking
- Broad applicability, e.g. generic computations on the edge, like MPC


SIEMENS

# Concepts

- *Service Orchestration Framework* where *ML-use-cases* are modelled as stateful services
- Self-contained, FaaS-like packages containing pipelines as atomic unit of execution
- High-level abstractions for executing pipelines in different environments & with different configurations
- Components for
  - Lifecycle Management
  - Managing artifacts & configurations
- SDK for developing pipeliens & APIs for consuming and producing immutable *state*






SIEMENS

# SDK

- Unintrusive definition of processing pipelines via decorators
- Define structure and interfaces of data-driven applications
- Stage separation and different Execution Modes (Batch, Continuous)
- *Statefullness*: APIs to produce and consume *state*
- *Expose* and *consume* configurations
- Agnostic towards data science frameworks

```
@template.apply(stage=Stage.TRAINING)
@template.from_config("parameter")
@template.inject_context()
def training(data: Dataset, parameter, context):
    model = train(data)
    with context.open(“model.p5”) as fp:
      write(model, fp)
    return SUCCESS
...

@template.apply()
def shared(data):
    # ...
    return normalize(data)
...

@template.apply(stage=Stage.PREDICTION)
@template.inject_context()
def inference(data : Image, context) -> int:
    with context.open(“model.p5”) as fp:
        model = load_model(fp)
    return infer(model, data)
```

# Projects and Applications

**Visual Quality Inspection**

- Scaled rollout of CV-based quality inspection models
- Heterogeneous training environments, continuous refinements of models

**Federated Learning**[1]

- *Templates* are deployed in a distributed manner and build cohorts for FL

**Predictive Maintenance in Cooling Facilities**

Partially in the CERN Openlab




1. Hiessl, T., Lakani, S. R., Kemnitz, J., Schall, D., & Schulte, S. (2022). Cohort-based federated learning services for industrial collaboration on the edge. *Journal of Parallel and Distributed Computing*, *167*, 64-76.

# Thank you!

**Thomas Kaufmann**
T DAI DAS
Siemensstraße 90
1211 Wien
Austria

**thomaskaufmann@siemens.com**

SIEMENS