DSC516: Cloud Computing

Part I: Basic Concepts and Models

Modern Computing Landscape

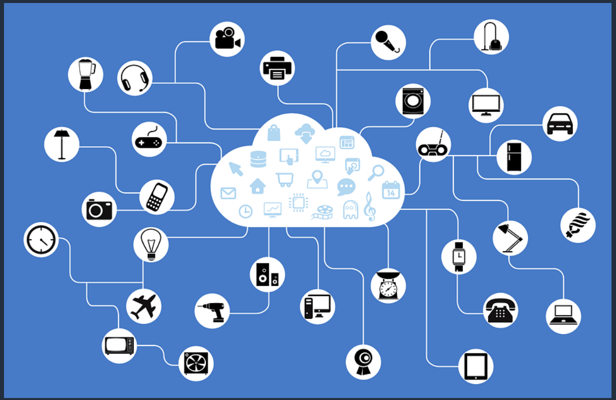
University of Cyprus
Department of Computer Science

Learning Objectives

- Understand and explain the current computing landscape.
- Understand and explain concepts, key constraints and key challenges in Edge and Fog computing.
- Memorize and use values of key properties of distributed systems' components and their evolution: CPU speed, network latencies, power consumption.

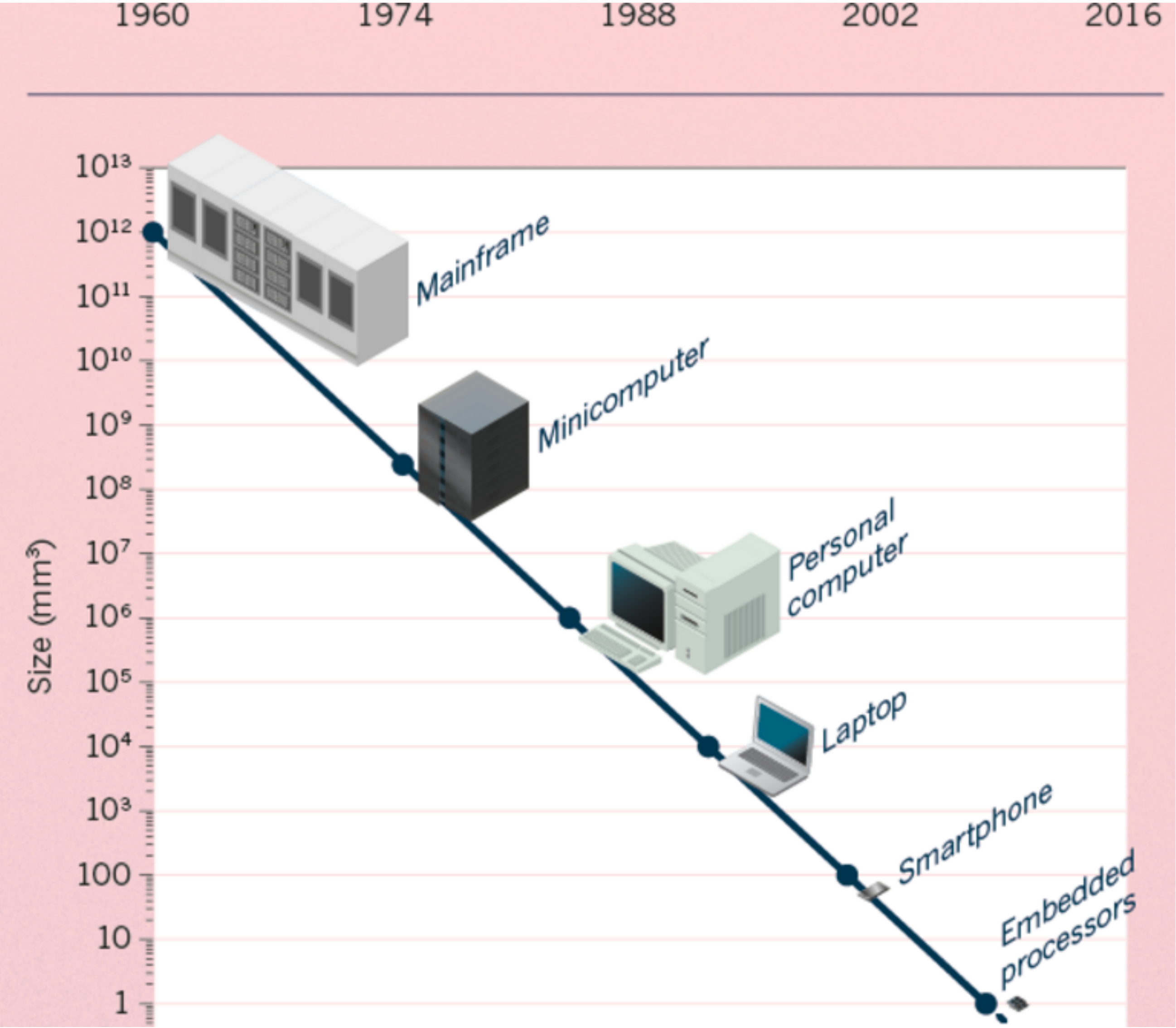
Modern Computing Landscape

From Embedded Devices to the Internet of Things (IOT)



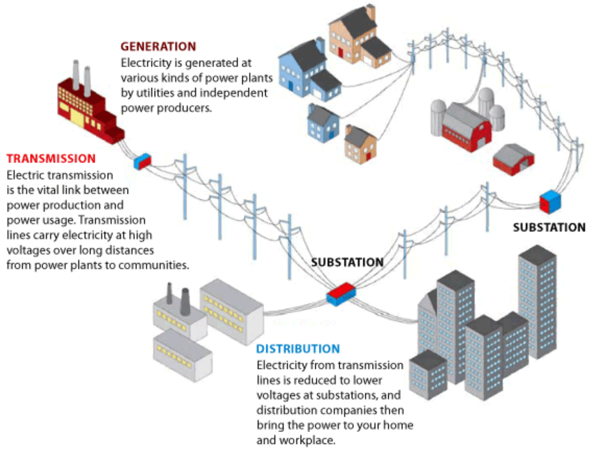
University of Cyprus
Department of Computer Science

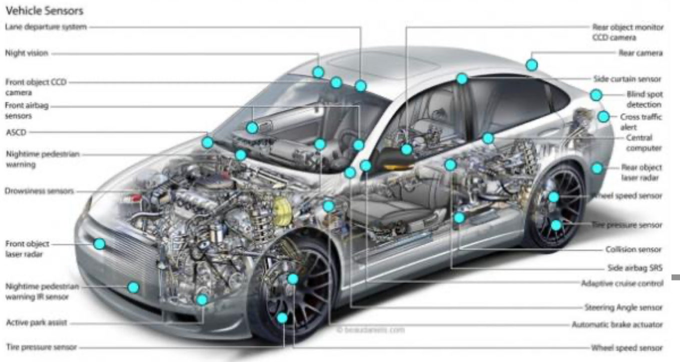
Embedded Devices

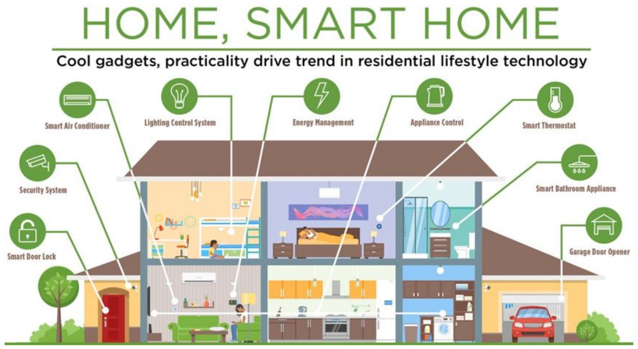


Cyber-physical Systems (CPS)

- Integration of computation and physical processes.
- Embedded computers and networks monitor and control physical processes, usually with feedback loops where:
  - physical processes affect computations and vice versa.
- Multi-scale and heterogeneous, mixing wide ranges of technologies.

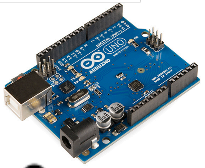

GENERATION
Electricity is generated at various kinds of power plants by utilities and independent power producers.
TRANSMISSION
Electric transmission is the vital link between power production and power usage. Transmission lines carry electricity at high voltages over long distances from power plants to communities.
SUBSTATION
SUBSTATION
DISTRIBUTION
Electricity from transmission lines is reduced to lower voltages at substations, and distribution companies then bring the power to your home and workplace.





Internet of Things (IoT)

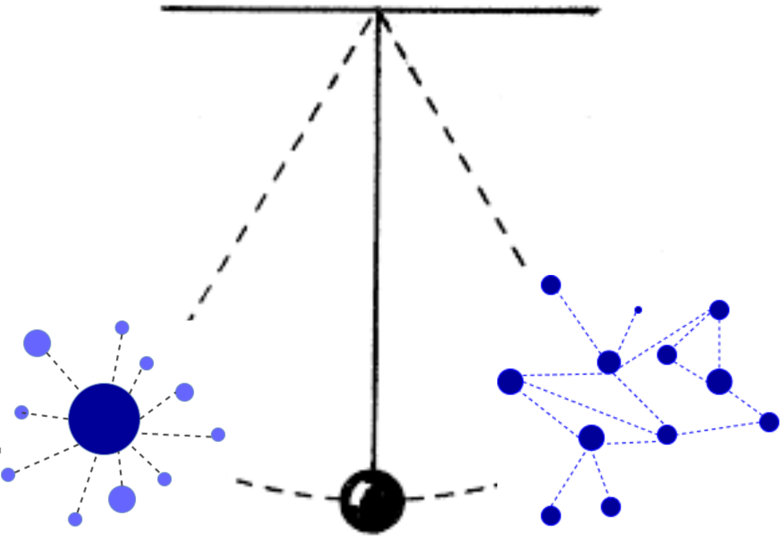
- Smaller, simpler, cheaper, Internet-enabled, (often) battery-powered, sensor-oriented embedded devices.
- Single-board computers (SBCs) & SB Micro-controllers as a platform to host and connect various sensors: Raspberry Pi, Arduino.



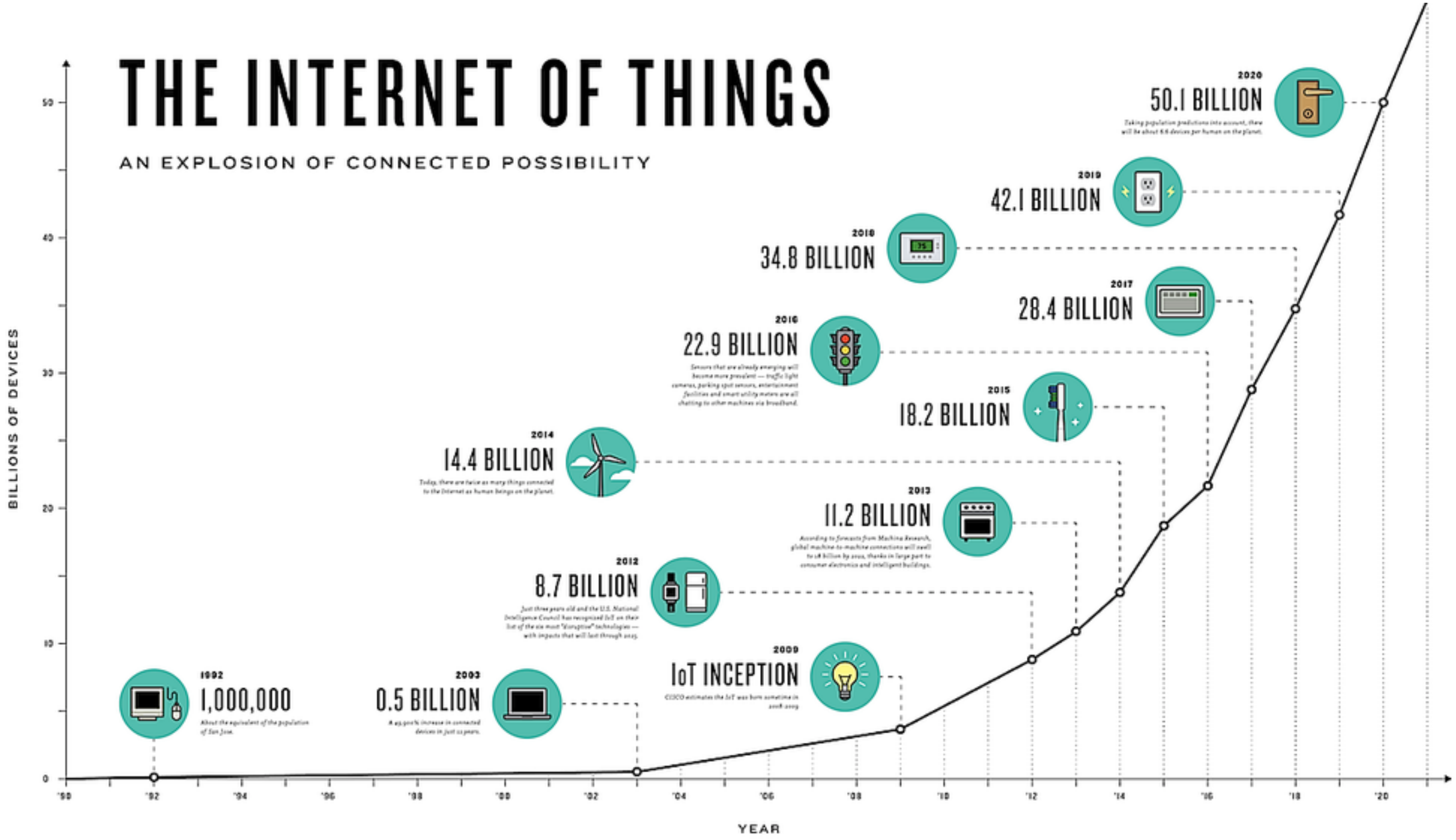
- Industrial equipment and robots
- Devices with various levels of mobility:
    - Smartphones with 3/4/5G & WiFi, battery-powered, with embedded or externally connected sensors (accelerometer, GPS, microphone, etc.).
    - Wearables with some wireless connectivity, with embedded sensors for activity monitoring and behavioral biometrics.
    - Battery-powered drones with cameras and wireless video link.
    - Autonomous vehicles continuously sensing their own operation and their surroundings.
- Tiny microelectronic devices, which run by harvesting incident electromagnetic energy (visible light or RF), and which sense intermittently their environment: RFID tags, swallowable capsules, "smart dust."

Internet of Things (IoT)

- IoT is an ecosystem of inter-connected devices with sensing and some processing capabilities which
- … exchange continuous data streams with other network-enabled devices, systems and humans...

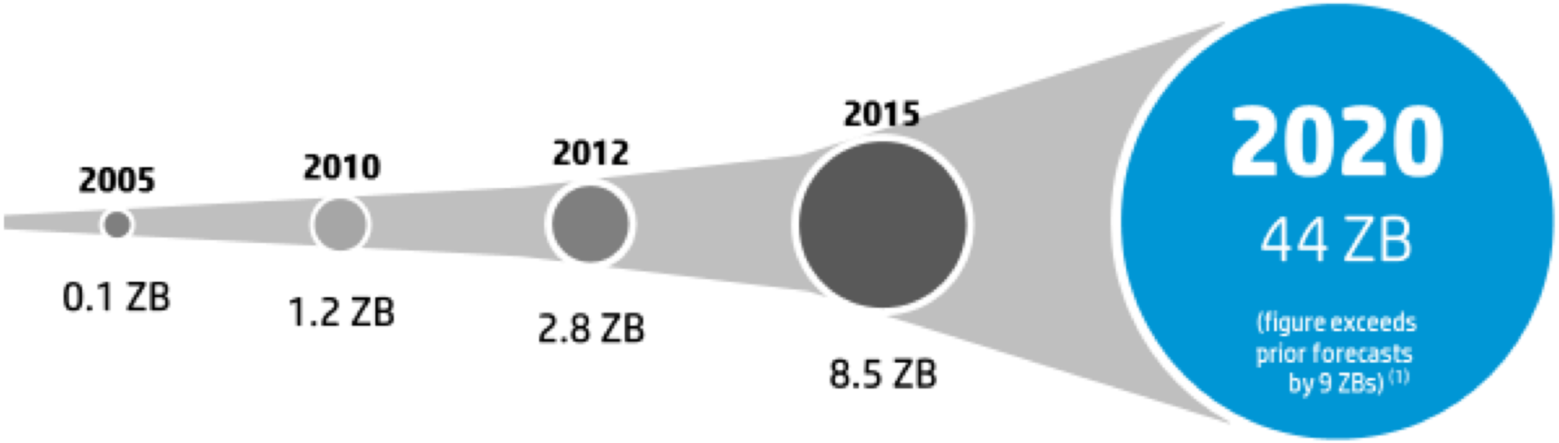


A Deluge of Devices



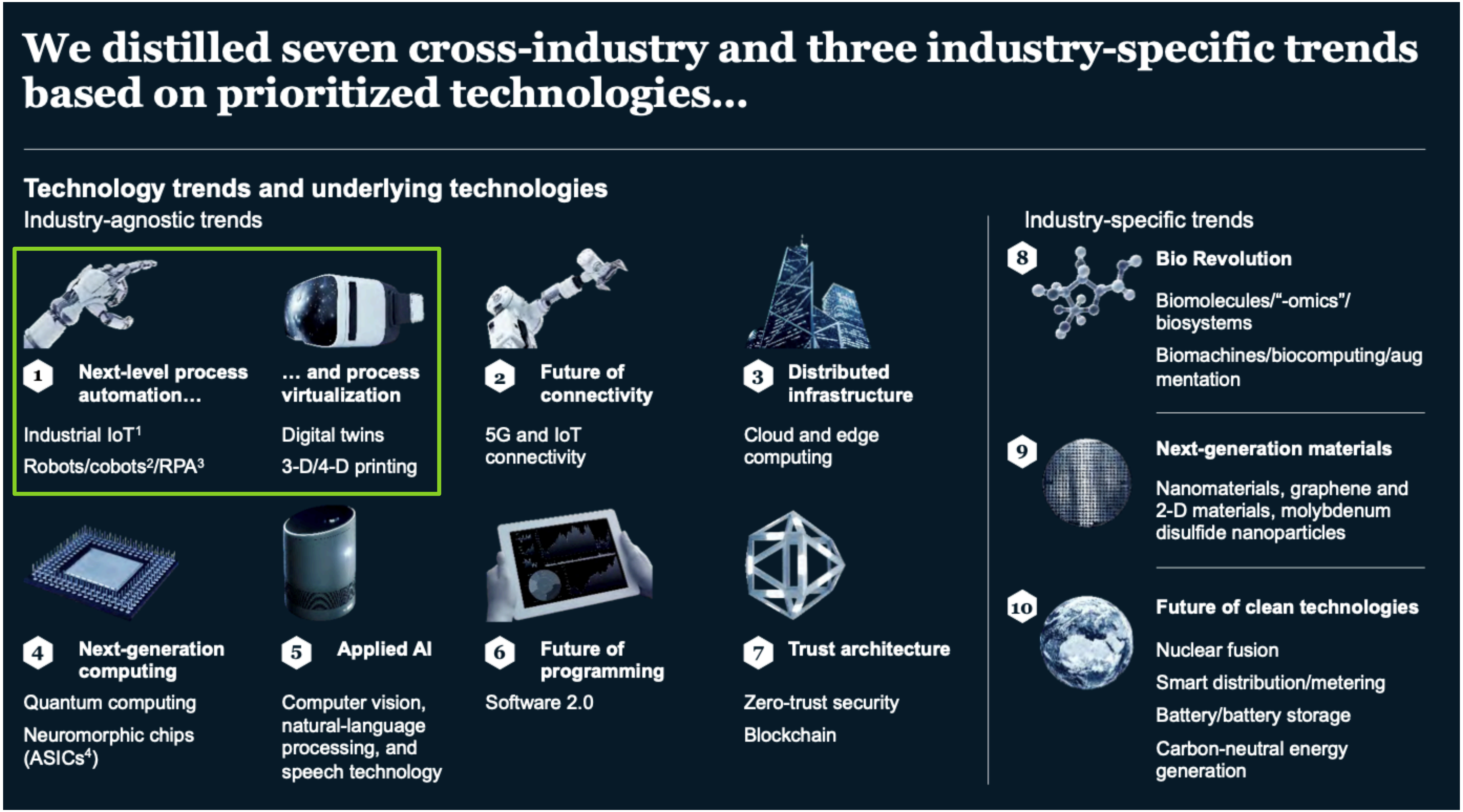
Number of IoT devices is expected to grow to as many as 20 billion connected devices by 2025 [Gartner 2017]

A Deluge of Data



- By 2025 the global amount of data will reach 175ZB and 30% of these data will be gathered in real-time [IDC 2018]
- IoT "things" will be generating 79.4ZB of data in 2025 [IDC, Nov. 2019]
- IoT data accounted for 2% of digital universe in 2012, with projections that by 2022 it will rise above 12% [Cisco]

Top 10 Trends Shaping Next Decade



We distilled seven cross-industry and three industry-specific trends based on prioritized technologies…

Technology trends and underlying technologies

Industry-agnostic trends

1 **Next-level process automation…**

- Industrial IoT[1]
- Robots/cobots[2]/RPA[3]

… and process virtualization

- Digital twins
- 3-D/4-D printing

2 **Future of connectivity**

- 5G and IoT connectivity

3 **Distributed infrastructure**

- Cloud and edge computing

4 **Next-generation computing**

- Quantum computing
- Neuromorphic chips (ASICs[4])

5 **Applied AI**

- Computer vision, natural-language processing, and speech technology

6 **Future of programming**

- Software 2.0

7 **Trust architecture**

- Zero-trust security
- Blockchain

Industry-specific trends

8 **Bio Revolution**

- Biomolecules/"-omics"/ biosystems
- Biomachines/biocomputing/augmentation

9 **Next-generation materials**

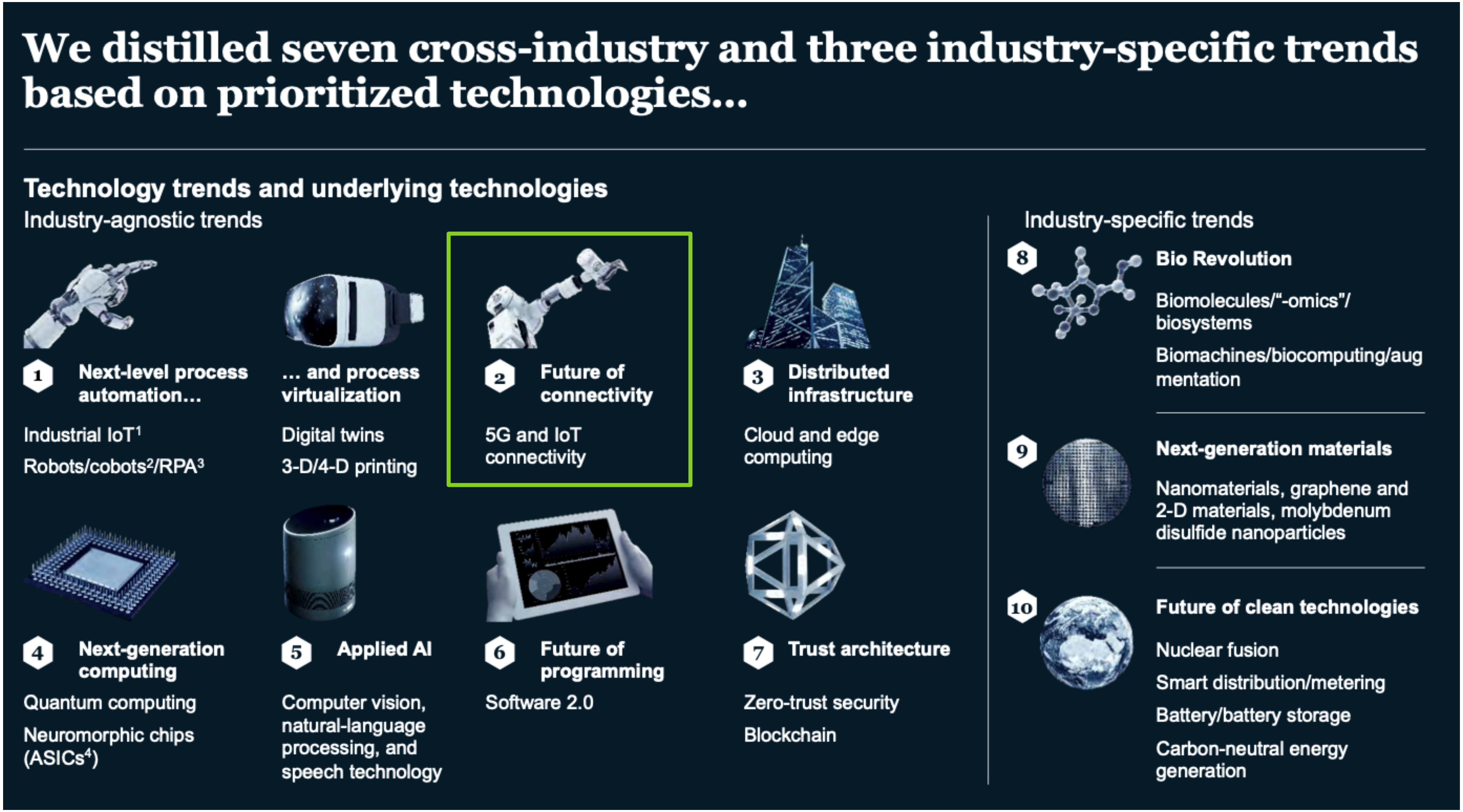
- Nanomaterials, graphene and 2-D materials, molybdenum disulfide nanoparticles

10 **Future of clean technologies**

- Nuclear fusion
- Smart distribution/metering
- Battery/battery storage
- Carbon-neutral energy generation

"Top Trends in Tech," McKinsey, 2021

Top 10 Trends Shaping Next Decade



We distilled seven cross-industry and three industry-specific trends based on prioritized technologies…

Technology trends and underlying technologies

Industry-agnostic trends

1 **Next-level process automation…**

Industrial IoT[1]

Robots/cobots[2]/RPA[3]

… and process virtualization

Digital twins

3-D/4-D printing

2 **Future of connectivity**

5G and IoT connectivity

3 **Distributed infrastructure**

Cloud and edge computing

4 **Next-generation computing**

Quantum computing

Neuromorphic chips (ASICs[4])

5 **Applied AI**

Computer vision, natural-language processing, and speech technology

6 **Future of programming**

Software 2.0

7 **Trust architecture**

Zero-trust security

Blockchain

Industry-specific trends

8 **Bio Revolution**

Biomolecules/"-omics"/ biosystems

Biomachines/biocomputing/augmentation

9 **Next-generation materials**

Nanomaterials, graphene and 2-D materials, molybdenum disulfide nanoparticles

10 **Future of clean technologies**

Nuclear fusion

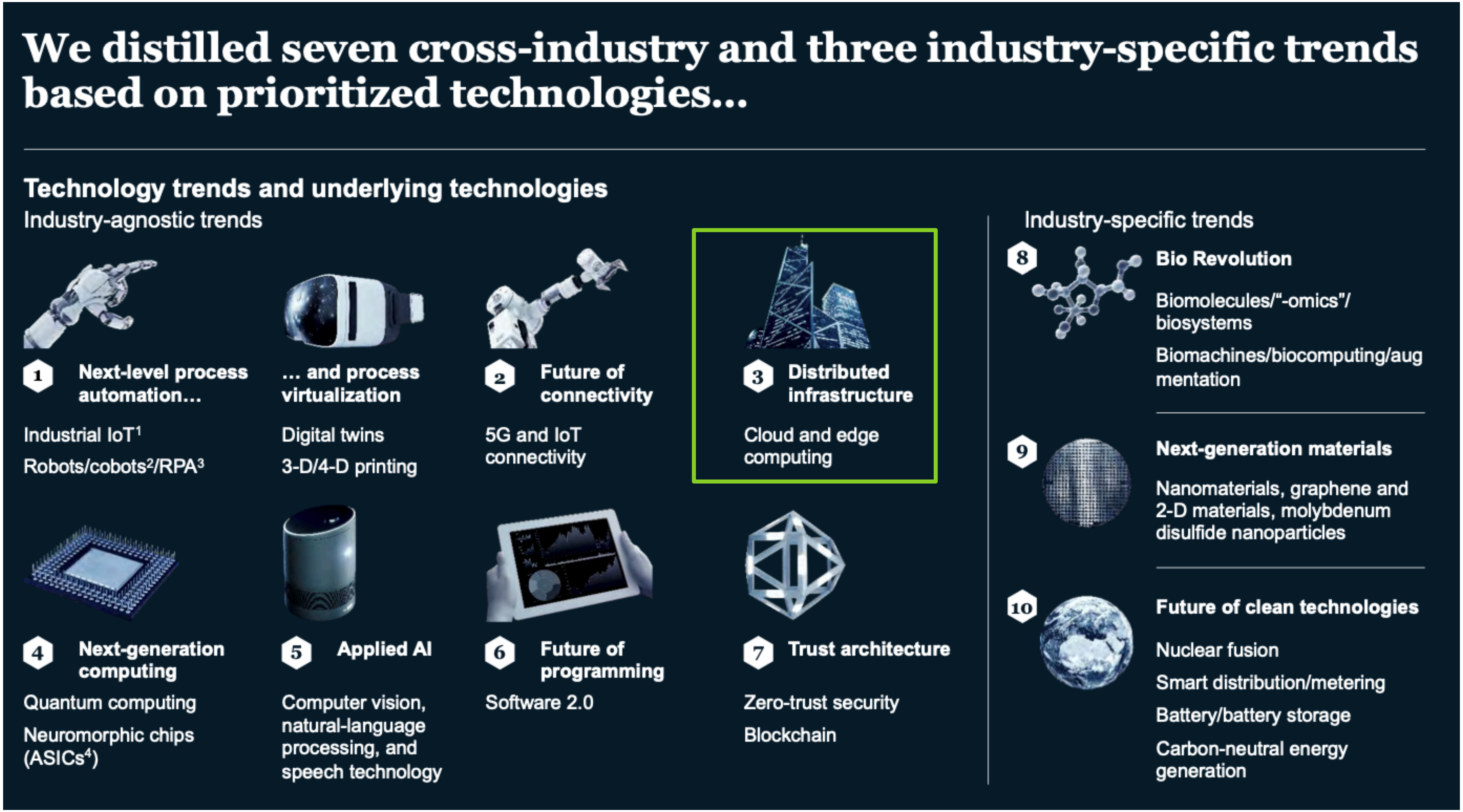
Smart distribution/metering

Battery/battery storage

Carbon-neutral energy generation

"Top Trends in Tech," McKinsey, 2021

Top 10 Trends Shaping Next Decade



We distilled seven cross-industry and three industry-specific trends based on prioritized technologies…

Technology trends and underlying technologies

Industry-agnostic trends

1 **Next-level process automation…**

Industrial IoT[1]

Robots/cobots[2]/RPA[3]

… and process virtualization

Digital twins

3-D/4-D printing

2 **Future of connectivity**

5G and IoT connectivity

3 **Distributed infrastructure**

Cloud and edge computing

4 **Next-generation computing**

Quantum computing

Neuromorphic chips (ASICs[4])

5 **Applied AI**

Computer vision, natural-language processing, and speech technology

6 **Future of programming**

Software 2.0

7 **Trust architecture**

Zero-trust security

Blockchain

Industry-specific trends

8 **Bio Revolution**

Biomolecules/"-omics"/ biosystems

Biomachines/biocomputing/augmentation

9 **Next-generation materials**

Nanomaterials, graphene and 2-D materials, molybdenum disulfide nanoparticles

10 **Future of clean technologies**

Nuclear fusion

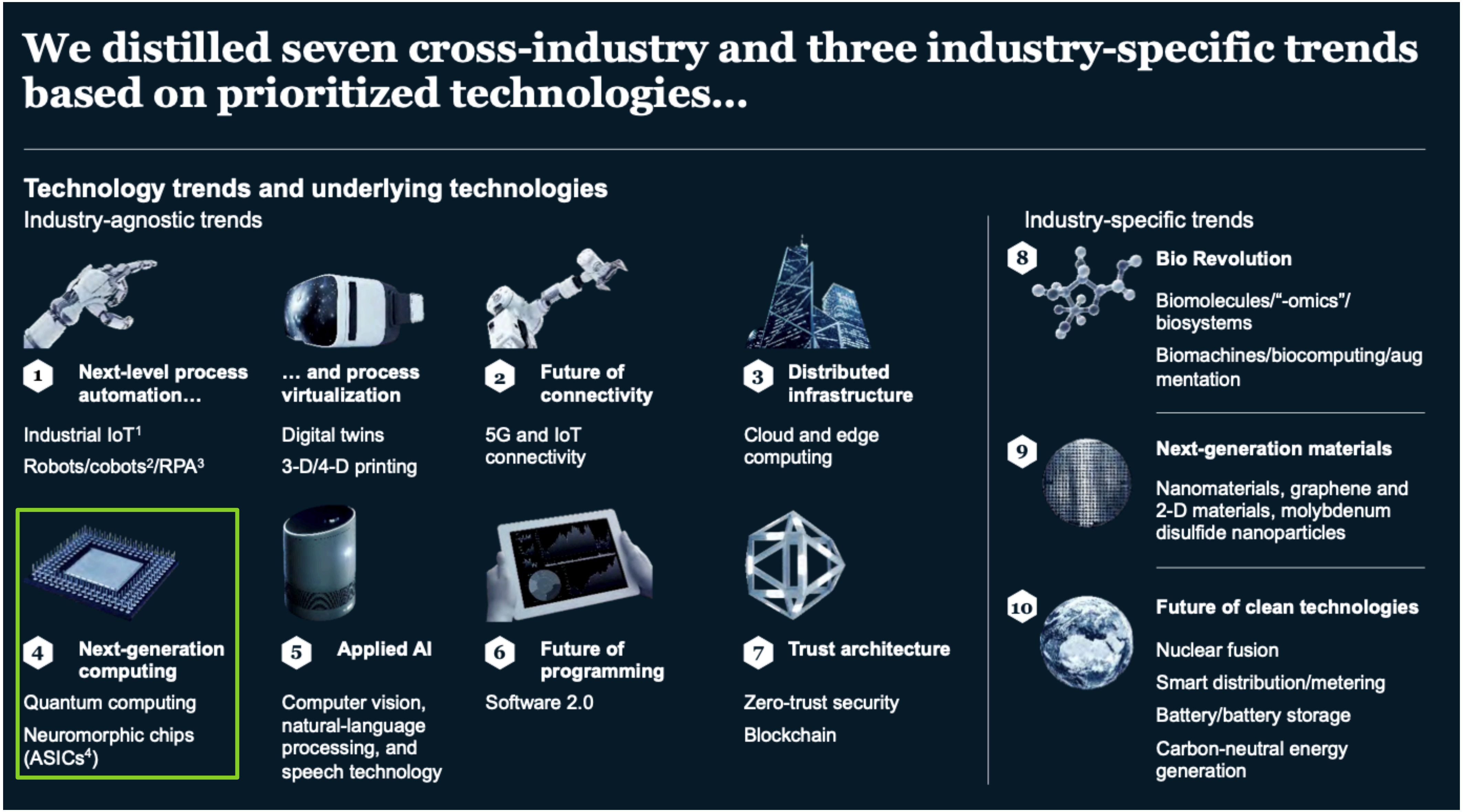
Smart distribution/metering

Battery/battery storage

Carbon-neutral energy generation

"Top Trends in Tech," McKinsey, 2021

Top 10 Trends Shaping Next Decade



We distilled seven cross-industry and three industry-specific trends based on prioritized technologies…

Technology trends and underlying technologies

Industry-agnostic trends

1 **Next-level process automation…**

Industrial IoT[1]

Robots/cobots[2]/RPA[3]

… and process virtualization

Digital twins

3-D/4-D printing

2 **Future of connectivity**

5G and IoT connectivity

3 **Distributed infrastructure**

Cloud and edge computing

4 **Next-generation computing**

Quantum computing

Neuromorphic chips (ASICs[4])

5 **Applied AI**

Computer vision, natural-language processing, and speech technology

6 **Future of programming**

Software 2.0

7 **Trust architecture**

Zero-trust security

Blockchain

Industry-specific trends

8 **Bio Revolution**

Biomolecules/"-omics"/ biosystems

Biomachines/biocomputing/augmentation

9 **Next-generation materials**

Nanomaterials, graphene and 2-D materials, molybdenum disulfide nanoparticles

10 **Future of clean technologies**

Nuclear fusion

Smart distribution/metering

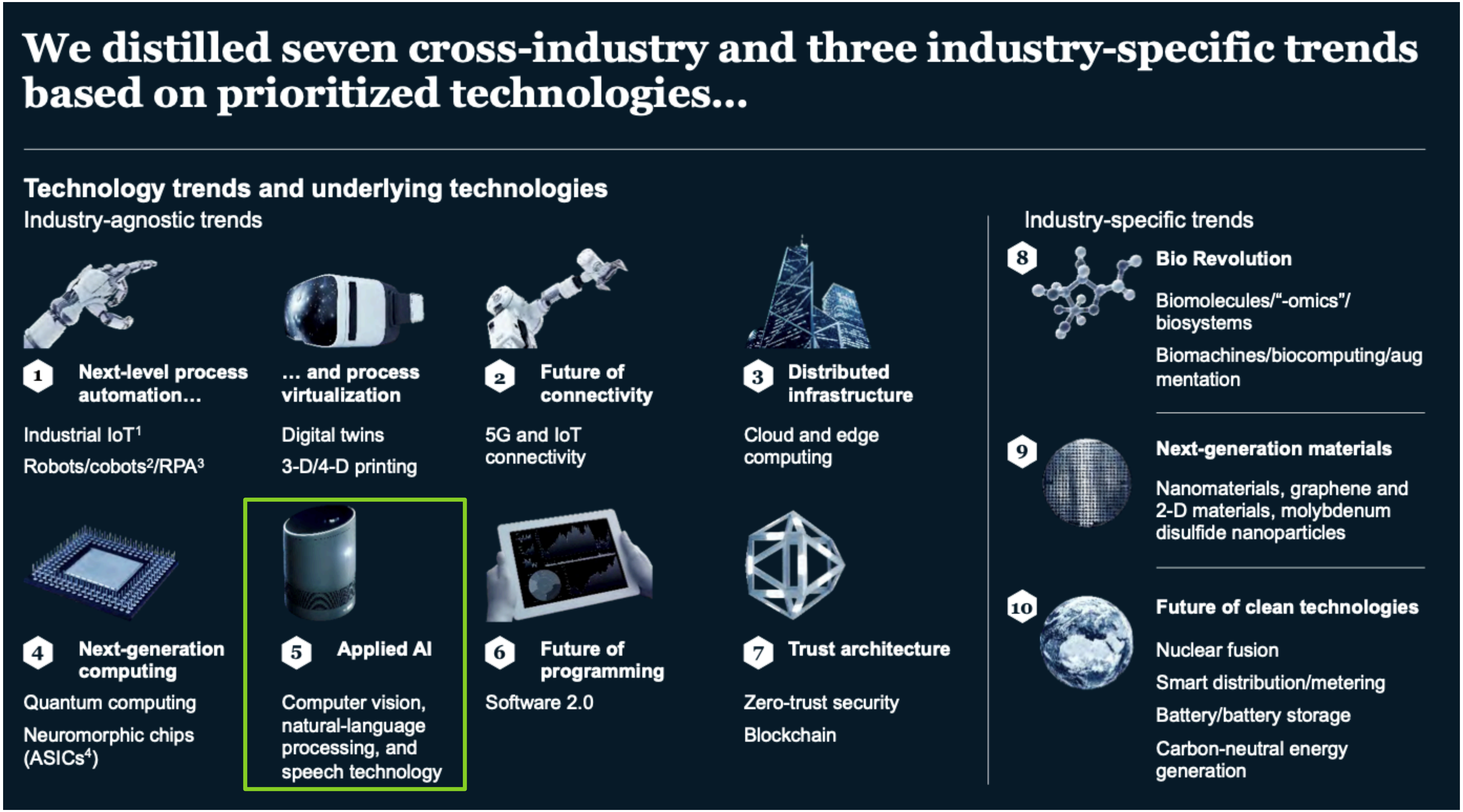
Battery/battery storage

Carbon-neutral energy generation

University of Cyprus
Department of Computer Science

"Top Trends in Tech," McKinsey, 2021

Top 10 Trends Shaping Next Decade



We distilled seven cross-industry and three industry-specific trends based on prioritized technologies...

Technology trends and underlying technologies

Industry-agnostic trends

1 **Next-level process automation...**

Industrial IoT[1]

Robots/cobots[2]/RPA[3]

... and process virtualization

Digital twins

3-D/4-D printing

2 **Future of connectivity**

5G and IoT connectivity

3 **Distributed infrastructure**

Cloud and edge computing

4 **Next-generation computing**

Quantum computing

Neuromorphic chips (ASICs[4])

5 **Applied AI**

Computer vision, natural-language processing, and speech technology

6 **Future of programming**

Software 2.0

7 **Trust architecture**

Zero-trust security

Blockchain

Industry-specific trends

8 **Bio Revolution**

Biomolecules/"-omics"/ biosystems

Biomachines/biocomputing/augmentation

9 **Next-generation materials**

Nanomaterials, graphene and 2-D materials, molybdenum disulfide nanoparticles

10 **Future of clean technologies**

Nuclear fusion

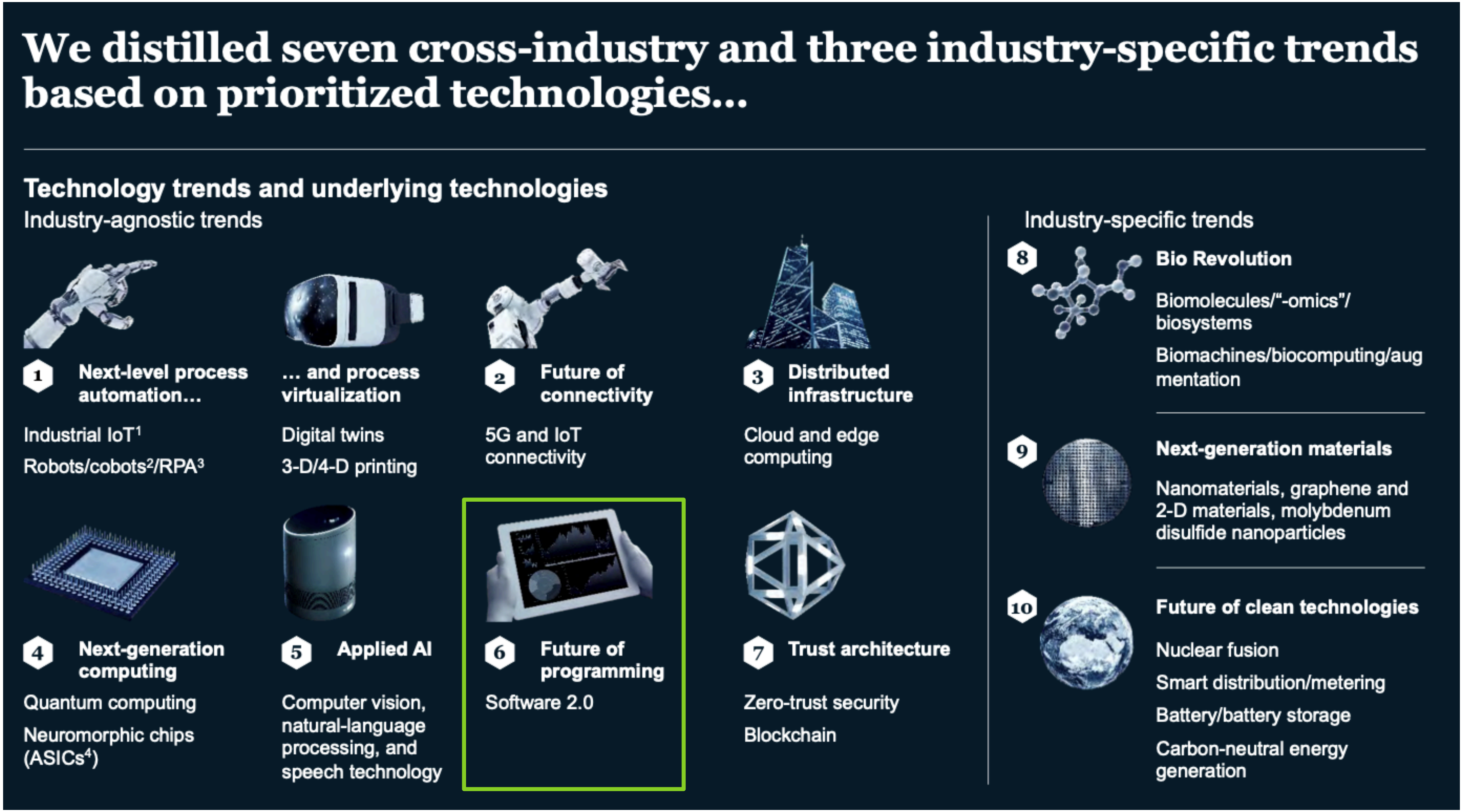
Smart distribution/metering

Battery/battery storage

Carbon-neutral energy generation

"Top Trends in Tech," McKinsey, 2021

Top 10 Trends Shaping Next Decade



We distilled seven cross-industry and three industry-specific trends based on prioritized technologies...

Technology trends and underlying technologies

Industry-agnostic trends

1 **Next-level process automation...**

Industrial IoT[1]

Robots/cobots[2]/RPA[3]

... and process virtualization

Digital twins

3-D/4-D printing

2 **Future of connectivity**

5G and IoT connectivity

3 **Distributed infrastructure**

Cloud and edge computing

4 **Next-generation computing**

Quantum computing

Neuromorphic chips (ASICs[4])

5 **Applied AI**

Computer vision, natural-language processing, and speech technology

6 **Future of programming**

Software 2.0

7 **Trust architecture**

Zero-trust security

Blockchain

Industry-specific trends

8 **Bio Revolution**

Biomolecules/"-omics"/ biosystems

Biomachines/biocomputing/augmentation

9 **Next-generation materials**

Nanomaterials, graphene and 2-D materials, molybdenum disulfide nanoparticles

10 **Future of clean technologies**

Nuclear fusion

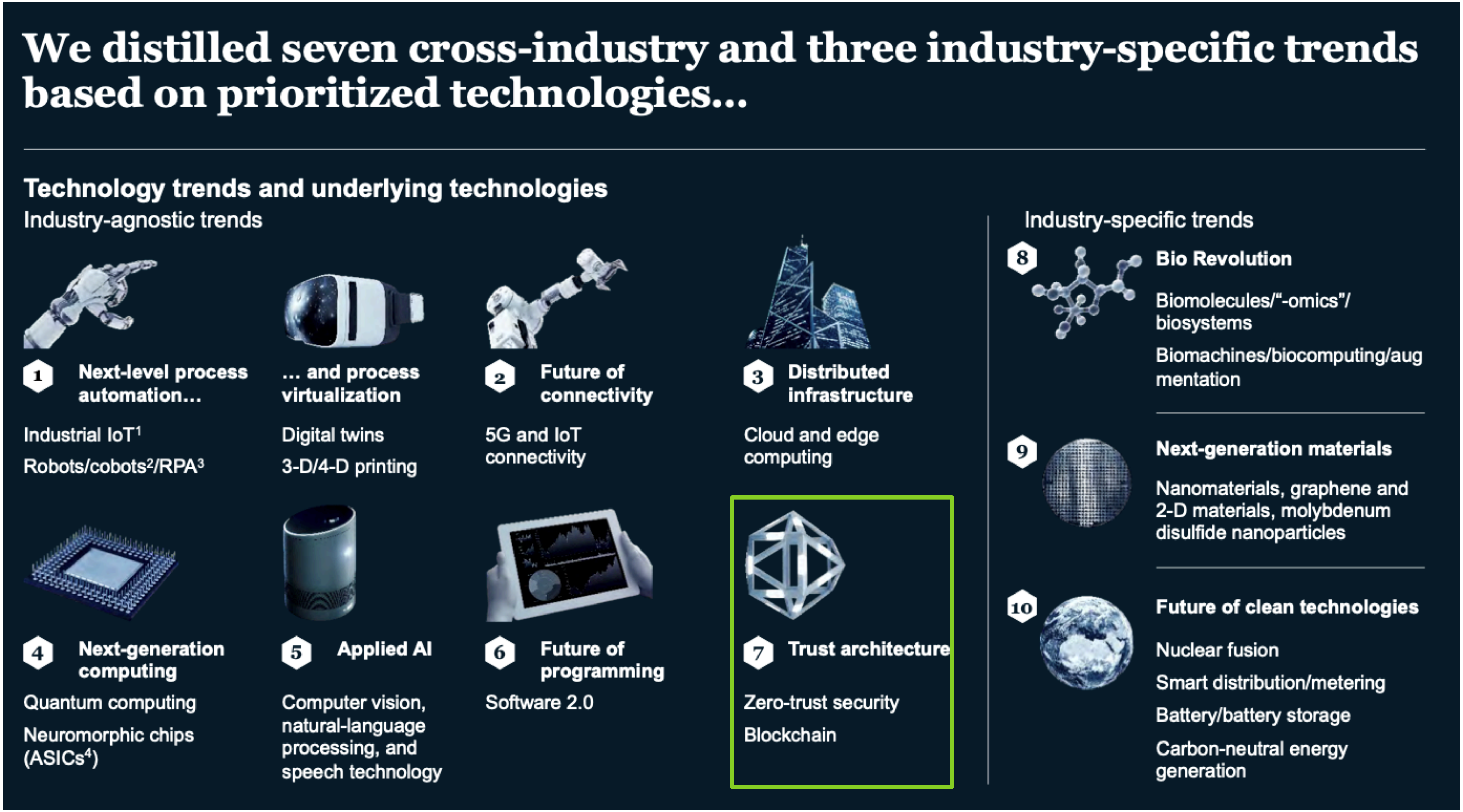
Smart distribution/metering

Battery/battery storage

Carbon-neutral energy generation

"Top Trends in Tech," McKinsey, 2021

Top 10 Trends Shaping Next Decade



We distilled seven cross-industry and three industry-specific trends based on prioritized technologies...

Technology trends and underlying technologies

Industry-agnostic trends

1 Next-level process automation...

Industrial IoT[1]

Robots/cobots[2]/RPA[3]

... and process virtualization

Digital twins

3-D/4-D printing

2 Future of connectivity

5G and IoT connectivity

3 Distributed infrastructure

Cloud and edge computing

4 Next-generation computing

Quantum computing

Neuromorphic chips (ASICs[4])

5 Applied AI

Computer vision, natural-language processing, and speech technology

6 Future of programming

Software 2.0

7 Trust architecture

Zero-trust security

Blockchain

Industry-specific trends

8 Bio Revolution

Biomolecules/"-omics"/ biosystems

Biomachines/biocomputing/augmentation

9 Next-generation materials

Nanomaterials, graphene and 2-D materials, molybdenum disulfide nanoparticles

10 Future of clean technologies

Nuclear fusion

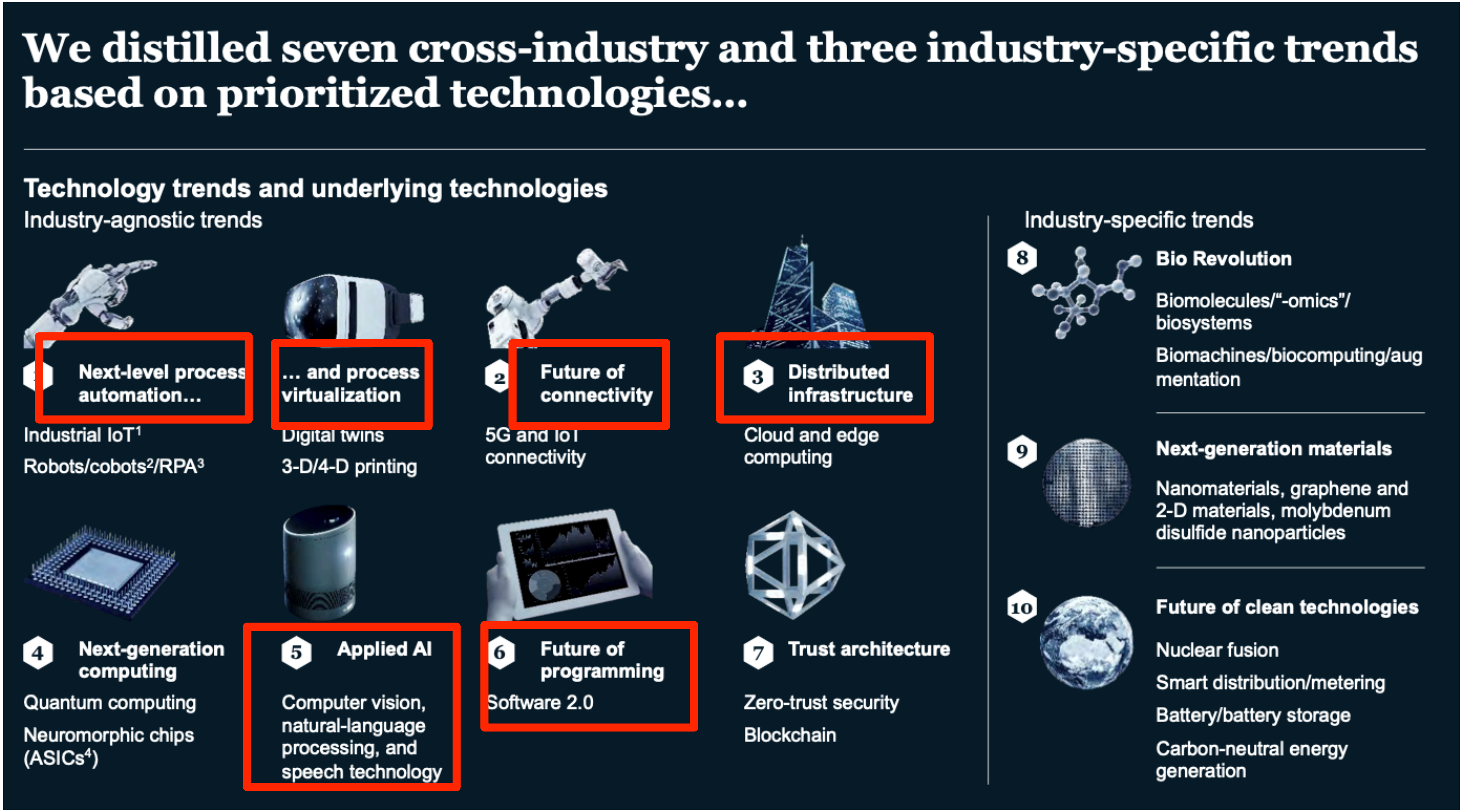
Smart distribution/metering

Battery/battery storage

Carbon-neutral energy generation

"Top Trends in Tech," McKinsey, 2021

Top 10 Trends Shaping Next Decade



We distilled seven cross-industry and three industry-specific trends based on prioritized technologies...

Technology trends and underlying technologies

Industry-agnostic trends

1 Next-level process automation...

Industrial IoT[1]

Robots/cobots[2]/RPA[3]

... and process virtualization

Digital twins

3-D/4-D printing

2 Future of connectivity

5G and IoT connectivity

3 Distributed infrastructure

Cloud and edge computing

4 Next-generation computing

Quantum computing

Neuromorphic chips (ASICs[4])

5 Applied AI

Computer vision, natural-language processing, and speech technology

6 Future of programming

Software 2.0

7 Trust architecture

Zero-trust security

Blockchain

Industry-specific trends

8 Bio Revolution

Biomolecules/"-omics"/ biosystems

Biomachines/biocomputing/augmentation

9 Next-generation materials

Nanomaterials, graphene and 2-D materials, molybdenum disulfide nanoparticles

10 Future of clean technologies

Nuclear fusion

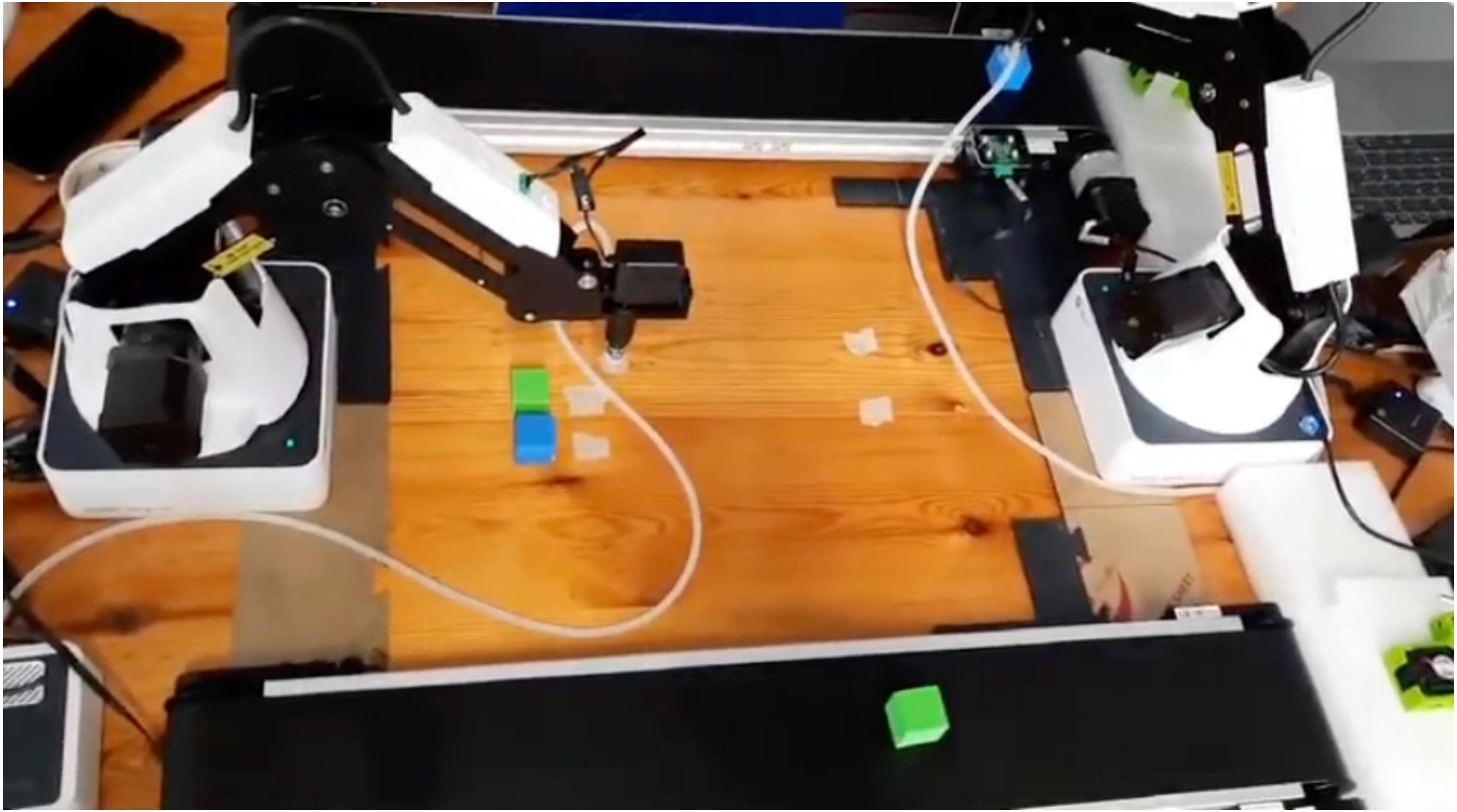
Smart distribution/metering

Battery/battery storage

Carbon-neutral energy generation

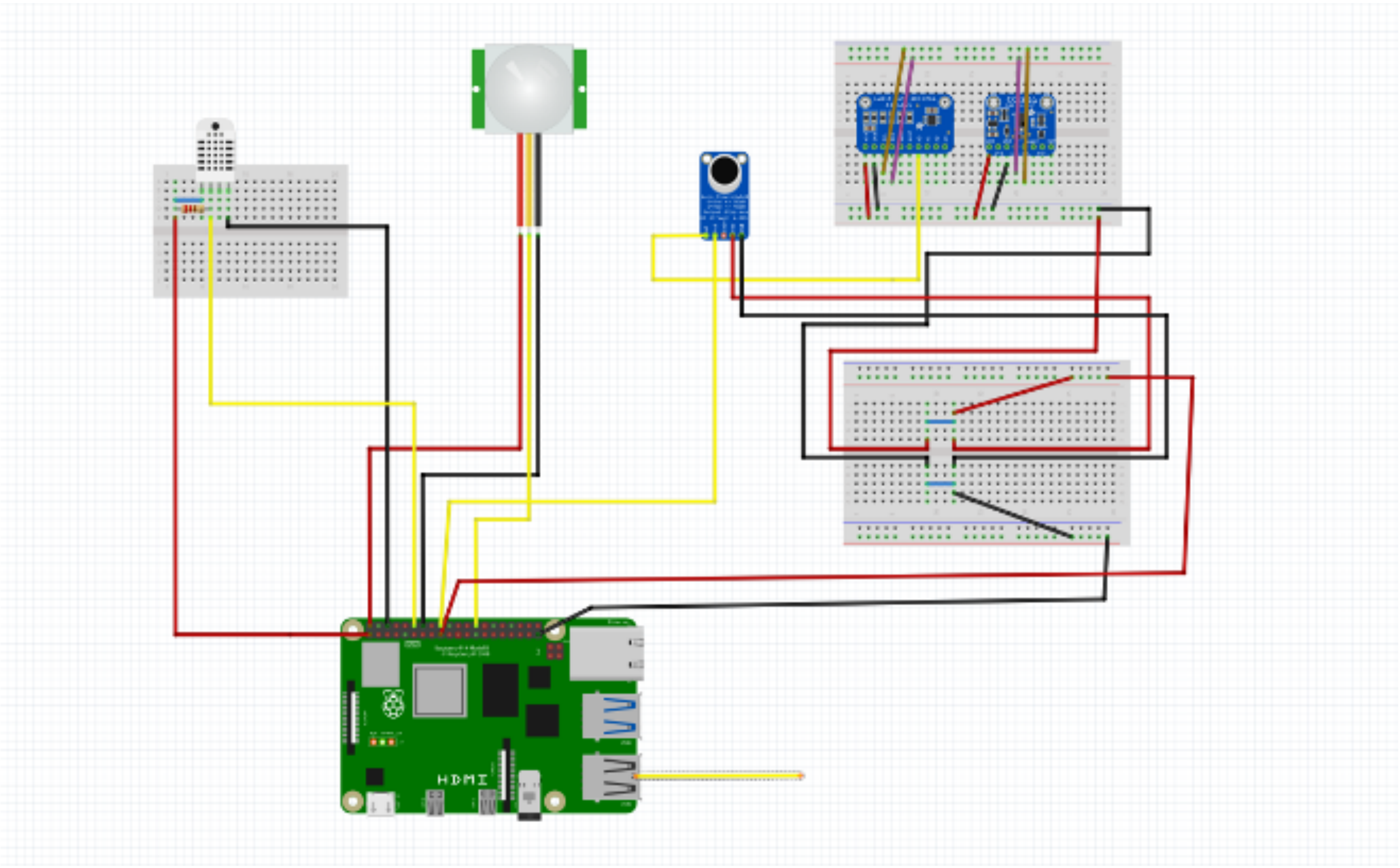
"Top Trends in Tech," McKinsey, 2021

An Industrial IoT Prototype



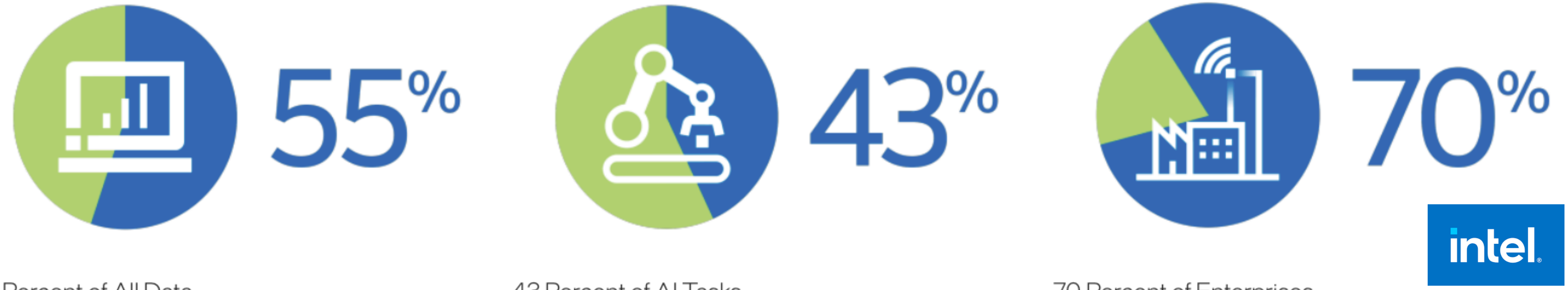
Source: Ioannis Panteli, “ΕΦΑΡΜΟΓΕΣ ΡΟΜΠΟΤΙΚΟΥ ΒΡΑΧΙΟΝΑ” BSc Thesis, UCY, 2020

Environmental Monitoring for Covid Risk Assessment inside Rooms



Source: Christina Hadjizorzi, "System to assess Covid spread in closed rooms using Raspberry PI" BSc Thesis, UCY, 2022

Huge Economic Impact



55 Percent of All Data
Is forecast to be generated by IoT in 2025.[1]

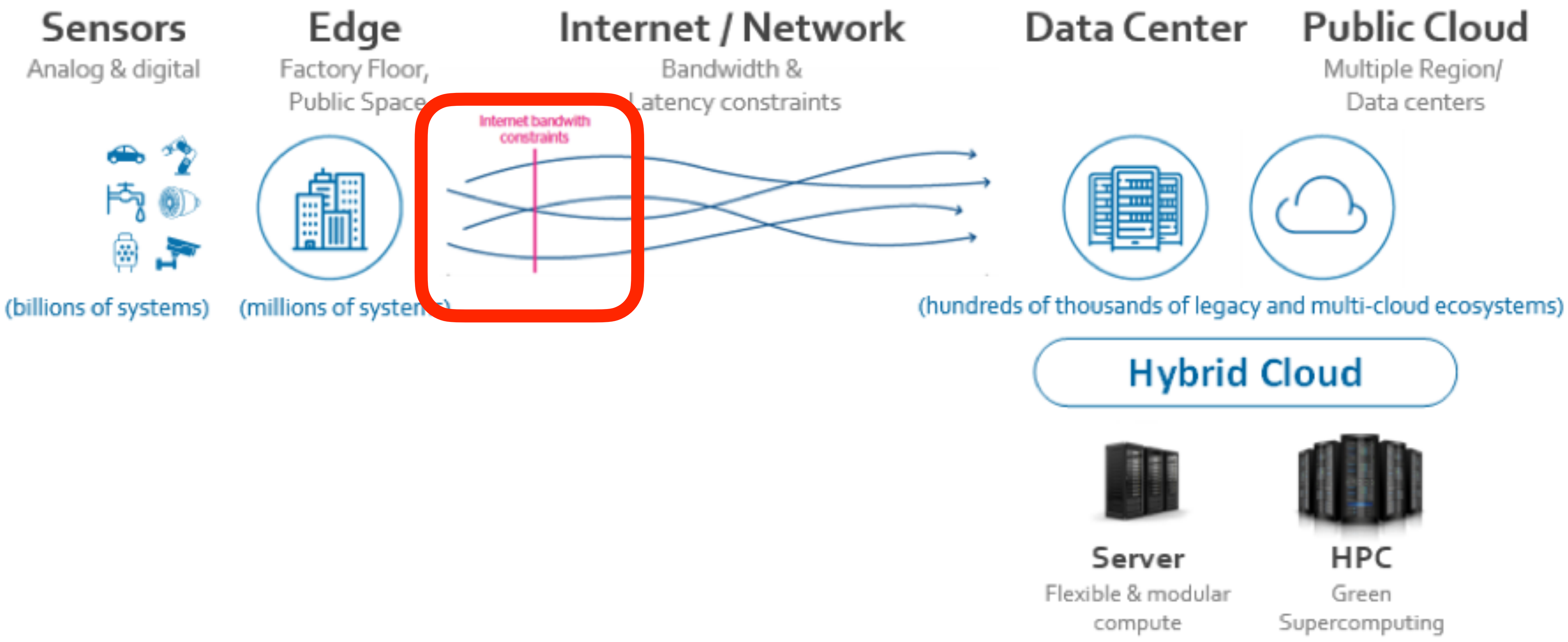
43 Percent of AI Tasks
Will happen on edge devices in 2023.[2]

70 Percent of Enterprises
Will run varying levels of data processing at the IoT edge by 2023.[3]

- IoT will have an economic impact of between $4 trillion-$11 trillion by 2025.
- Companies can capture value by creating new revenue streams from:
  - providing connected solutions and services to consumers/enterprises
  - reducing costs in operations.
- Much of the IoT data currently captured, however, is under leveraged.

McKinsey
Global Institute

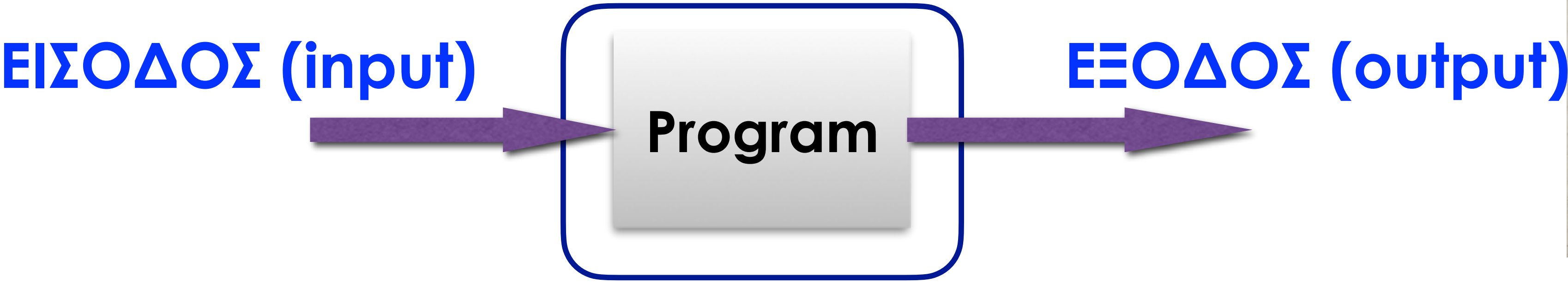
Expected Approach, but..



Source: European Industrial Technology Roadmap for the next Generation Cloud-Edge Offering. 2021.

HOW DO WE PROGRAM THE NEW IT LANDSCAPE?

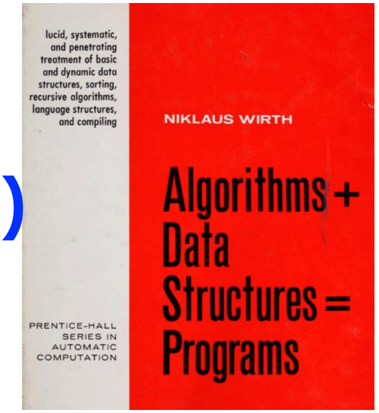
In Search of a Computing Model



ΕΙΣΟΔΟΣ (input)

Program

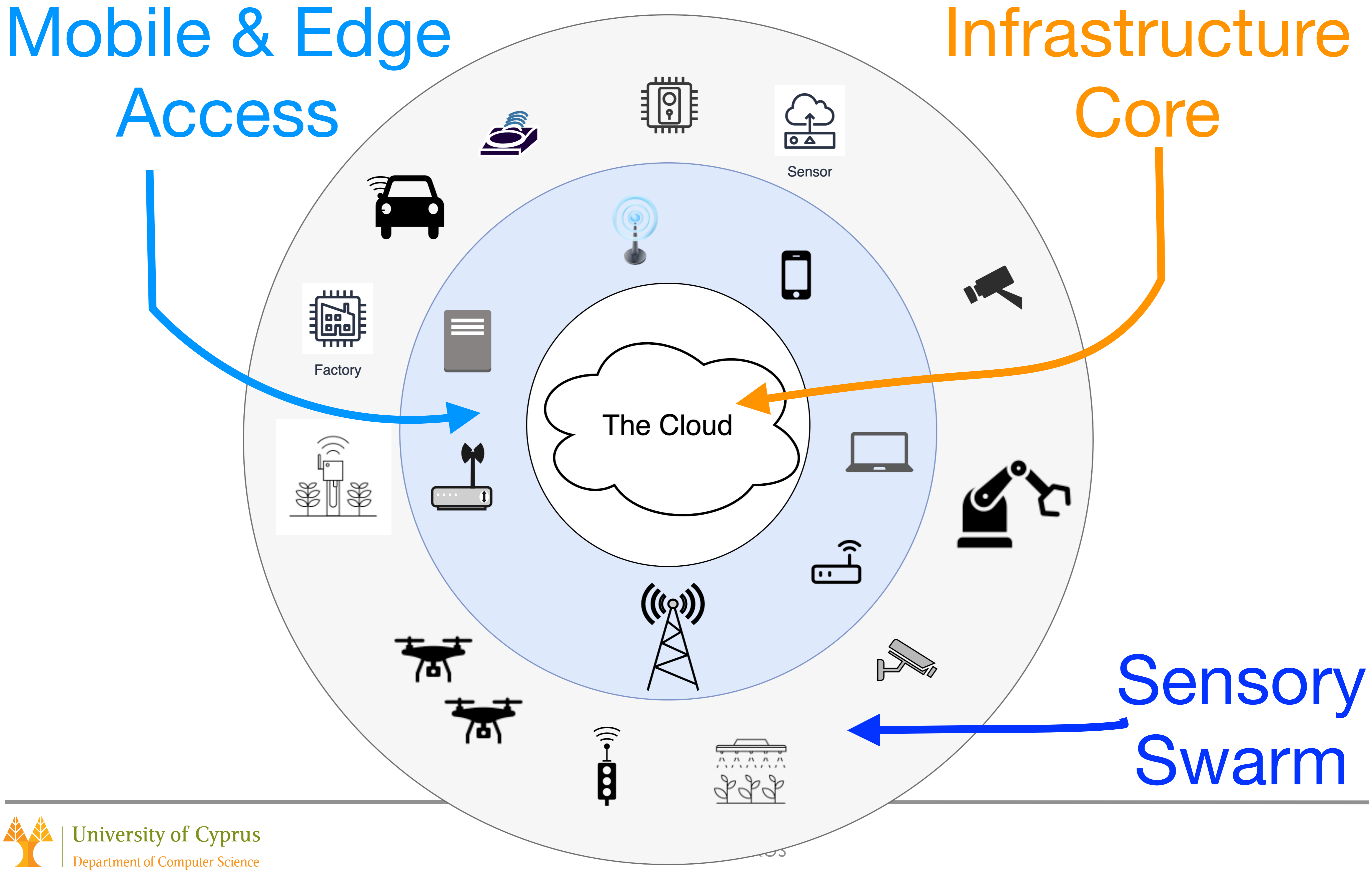
ΕΞΟΔΟΣ (output)


lucid, systematic, and penetrating treatment of basic and dynamic data structures, sorting, recursive algorithms, language structures, and compiling
NIKLAUS WIRTH
Algorithms + Data Structures = Programs
PRENTICE-HALL SERIES IN AUTOMATIC COMPUTATION

- Which is the computer?
- What is an application?
- Where do we get the input from and what do we do with the output?
- Where is the operating system and what does it need to do?
- Which programming language / programming abstractions?

WHERE IS THE COMPUTER?

A New IT Landscape



Mobile & Edge Access

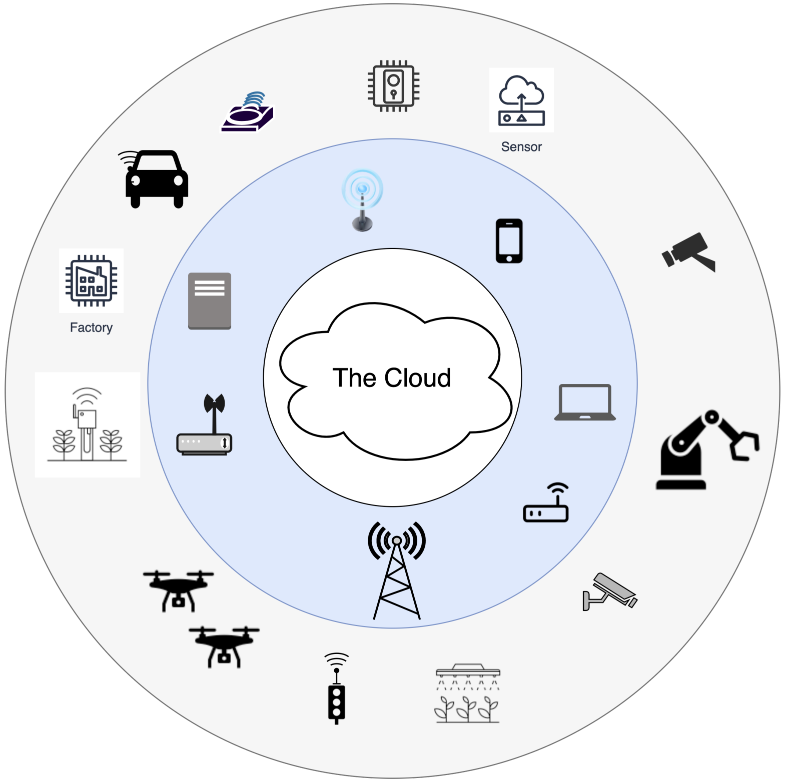
Infrastructure Core

Sensory Swarm

Issues to consider

- Location of devices
- Device characteristics & constraints
- Communication capabilities and constraints
- Ownership, maintenance, governance
- Energy and Cost
- Application development, deployment, management

HOW MIGHT WE ORGANIZE OUR COMPUTING UNIVERSE INTO A SIMPLE FRAMEWORK WITH EXPLANATORY POWER AND PREDICTIVE VALUE?

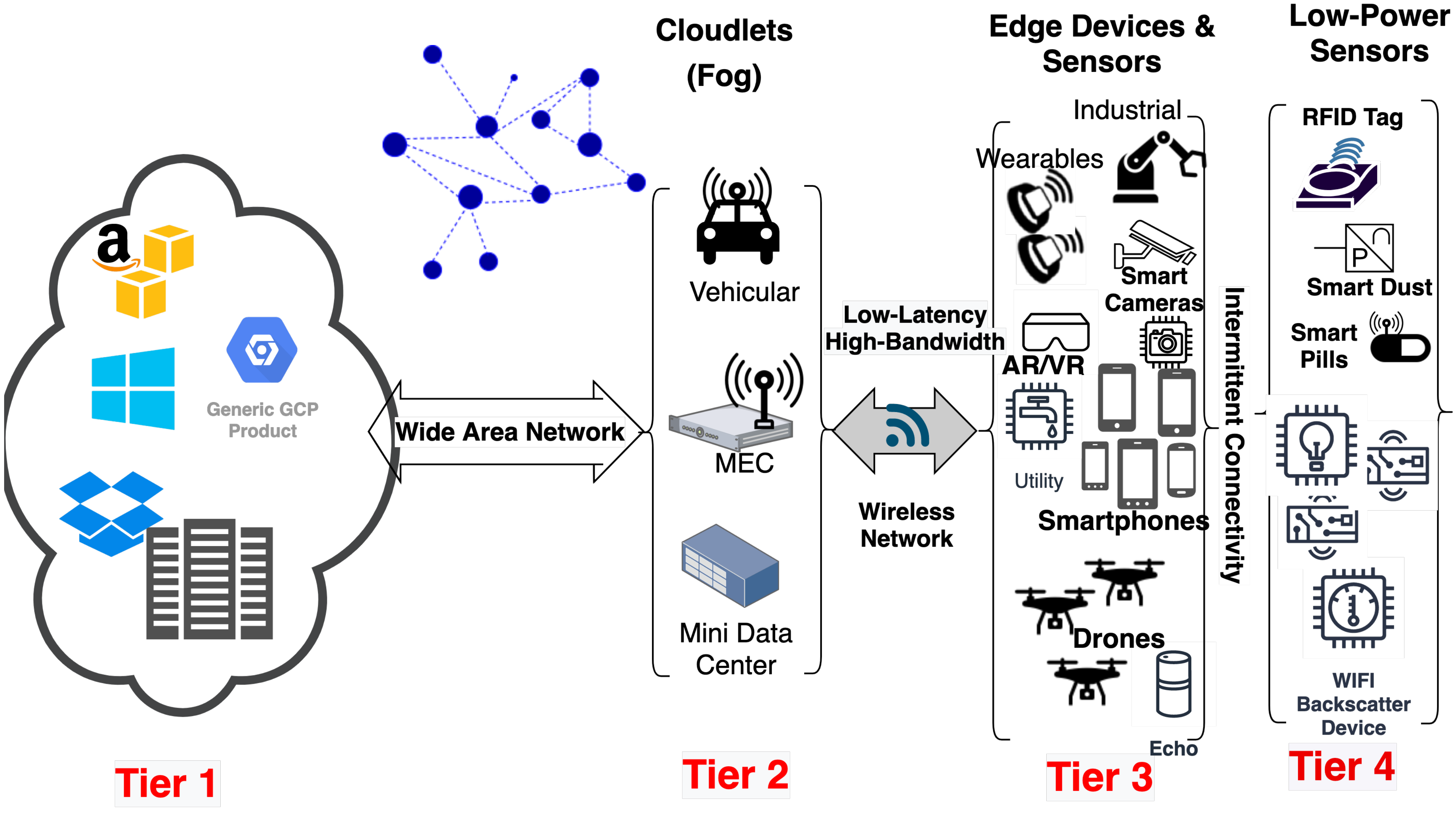


Mahadev Satyanarayanan, Wei Gao, and Brandon Lucia. 2019. ***The Computing Landscape of the 21st Century****.HotMobile '19*

Modern Computing Landscape

A Tiered Model of Internet Computing

A Tiered Model of Computing



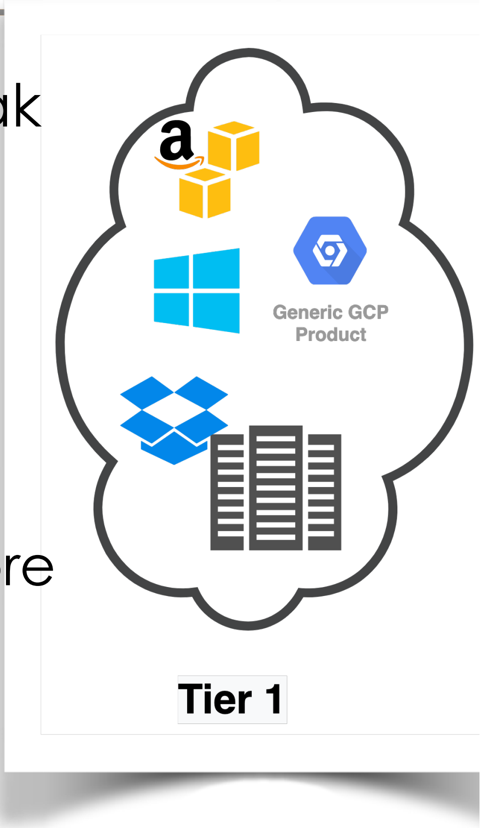
Mahadev Satyanarayanan, Wei Gao, and Brandon Lucia. 2019. ***The Computing Landscape of the 21st Century****.HotMobile '19*

What is a Tier?

- A tier is defined by a common set of important design constraints.
- Many alternative hw & sw implementations
  - Subject to the same set of design constraints
- No expectation of full interoperability across tiers
  - Randomly choosing one component from each tier unlikely to result in a functional system.
  - Many sets of compatible choices across tiers.
  - A single company will ensure that its products at each tier work well with its own products in other tiers, but not necessarily with products of other companies.

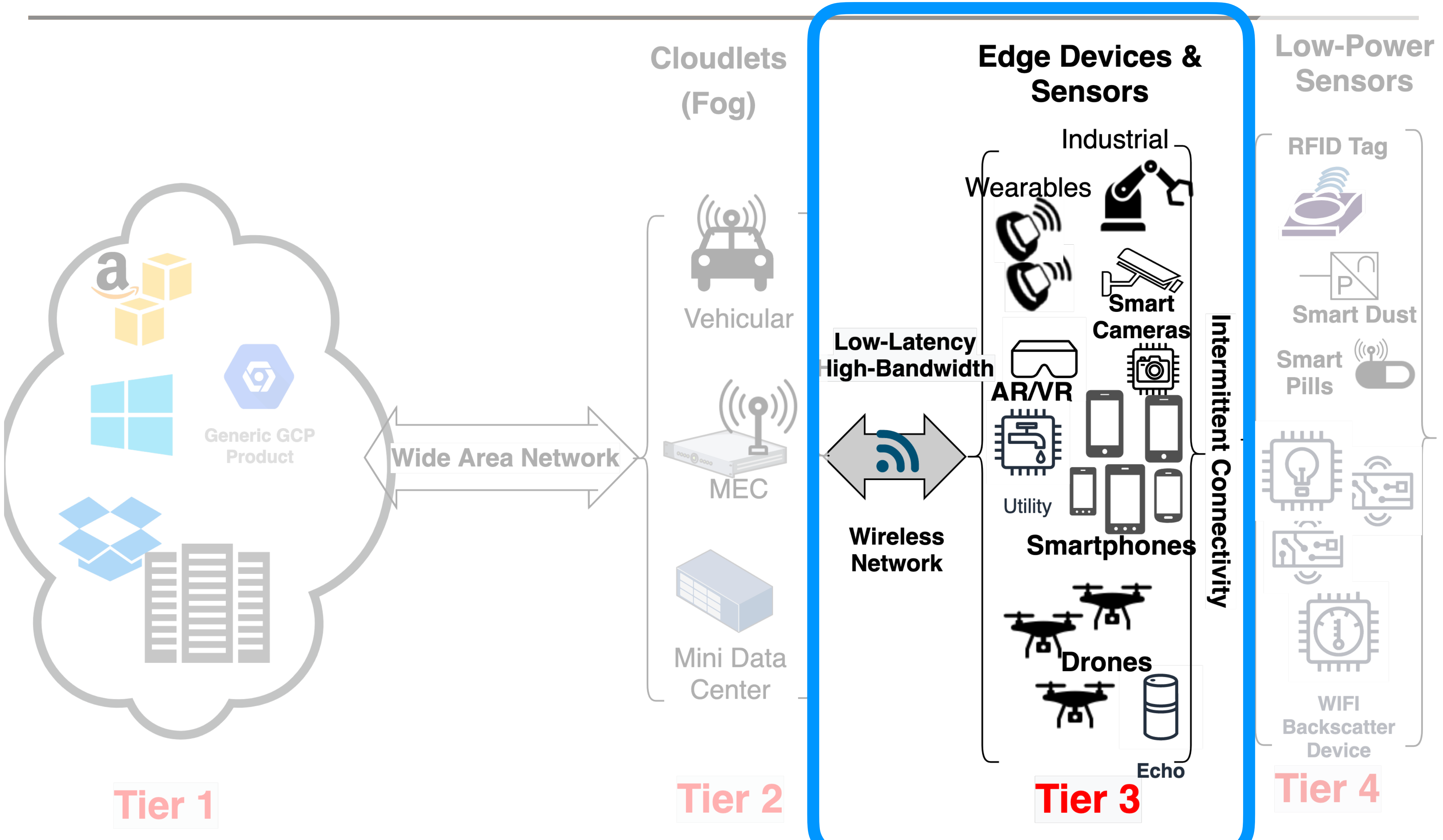
Tier 1 - The Cloud

- Compute Elasticity: Nearly unlimited, able to meet peak demand
  - Other tiers have limited elasticity
- Storage Permanence: storage redundancy (RAID), infrastructure stability, management practices
  - Other tiers offer less security; often, it is imperative to store data captured at those tiers to the cloud
- Consolidation
  - Economies of scale and very low total cost of computing;
  - IT personnel costs amortised over many machines in a very large data center



For large tasks without strict timing, data ingress volume, or data privacy requirements, Tier-1 is typically the optimal place to perform the task

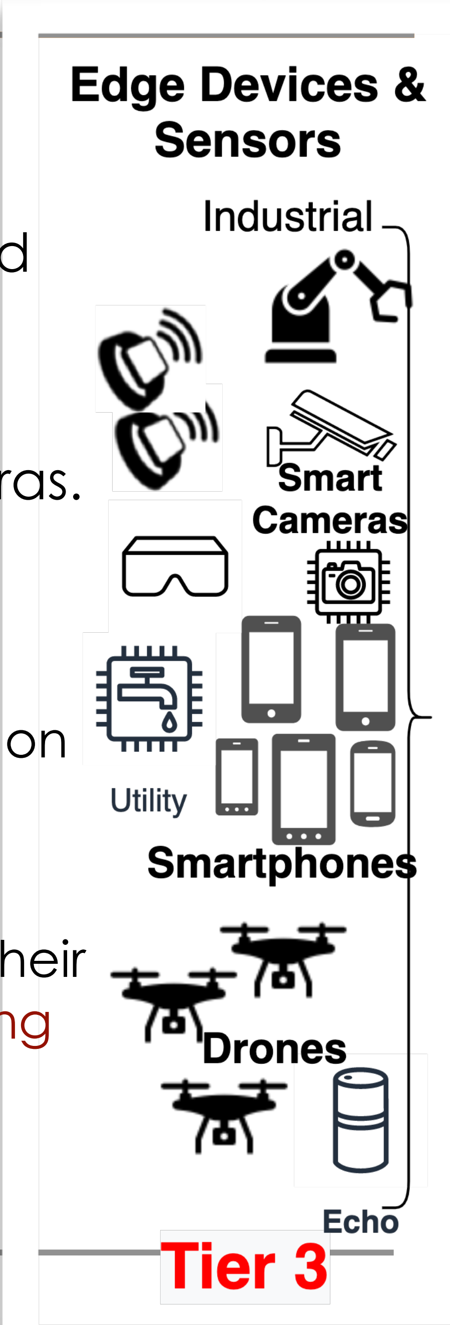
Tier 3: The Edge



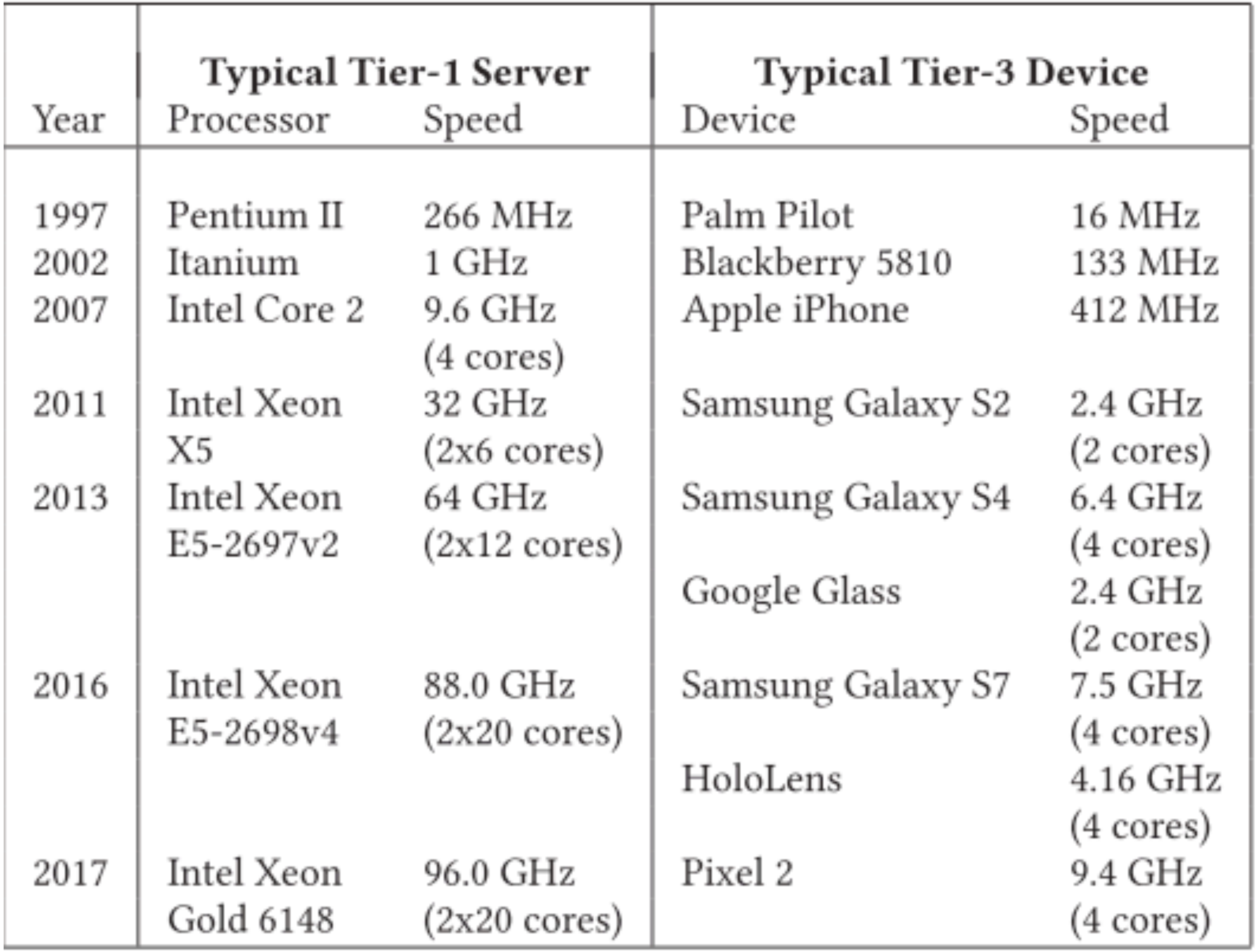
Mahadev Satyanarayanan, Wei Gao, and Brandon Lucia. 2019. ***The Computing Landscape of the 21st Century****.HotMobile '19*

Tier 3: Constraints & Concerns

- **Mobility**: stringent constraints on weight, size, and heat dissipation of devices that a user carries or wears.
- **Battery life**: places time constraints on duration of mobility and use, connectivity & computation intensity.
- **Sensing**: mobile devices rich in sensors such as GPS, microphones, accelerometers, gyroscopes, and video cameras.
  - Not powerful enough to perform real-time analysis of data captured by their on-board sensors (e.g., video analytics).
  - Large gap between what is feasible on a mobile device and on a server of the same era.
  - The vision of low-cost embedded sensing with tiny sensing-computing-communication platforms continuously report on their environment has proved elusive: replacing batteries or charging sensors is time-consuming and/or difficult.



Mobility penalty map: processing

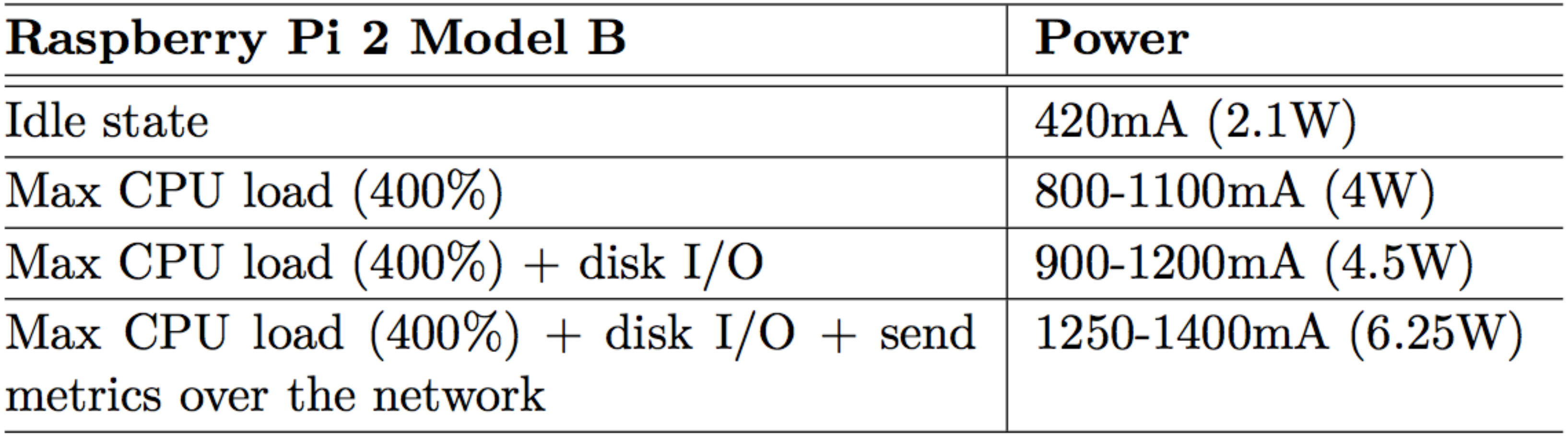


	Typical Tier-1 Server		Typical Tier-3 Device	
Year	Processor	Speed	Device	Speed
1997	Pentium II	266 MHz	Palm Pilot	16 MHz
2002	Itanium	1 GHz	Blackberry 5810	133 MHz
2007	Intel Core 2	9.6 GHz (4 cores)	Apple iPhone	412 MHz
2011	Intel Xeon X5	32 GHz (2x6 cores)	Samsung Galaxy S2	2.4 GHz (2 cores)
2013	Intel Xeon E5-2697v2	64 GHz (2x12 cores)	Samsung Galaxy S4	6.4 GHz (4 cores)
			Google Glass	2.4 GHz (2 cores)
2016	Intel Xeon E5-2698v4	88.0 GHz (2x20 cores)	Samsung Galaxy S7	7.5 GHz (4 cores)
			HoloLens	4.16 GHz (4 cores)
2017	Intel Xeon Gold 6148	96.0 GHz (2x20 cores)	Pixel 2	9.4 GHz (4 cores)

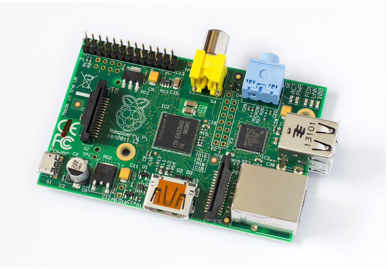
Source: Adapted from Chen [3] and Flinn [8]

"Speed" metric = number of cores times per-core clock speed.

Mobility penalty map: power

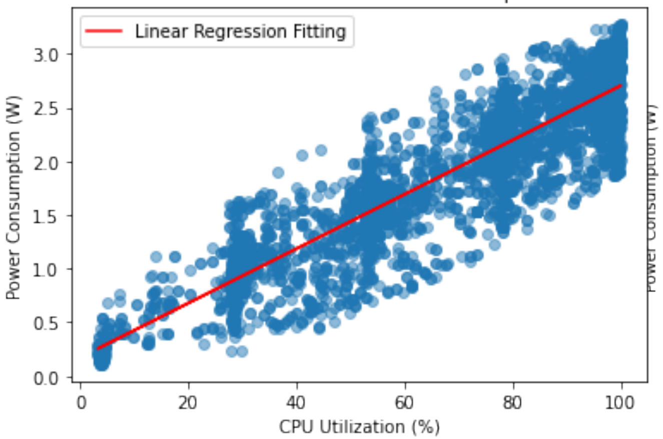


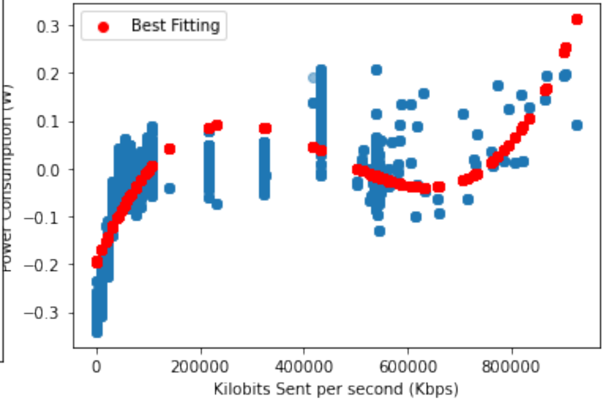
Raspberry Pi 2 Model B	Power
Idle state	420mA (2.1W)
Max CPU load (400%)	800-1100mA (4W)
Max CPU load (400%) + disk I/O	900-1200mA (4.5W)
Max CPU load (400%) + disk I/O + send metrics over the network	1250-1400mA (6.25W)

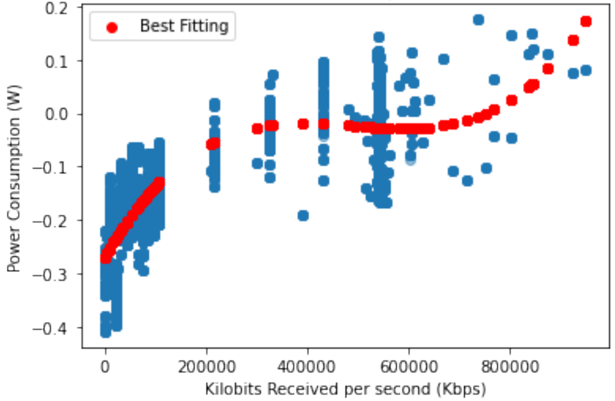


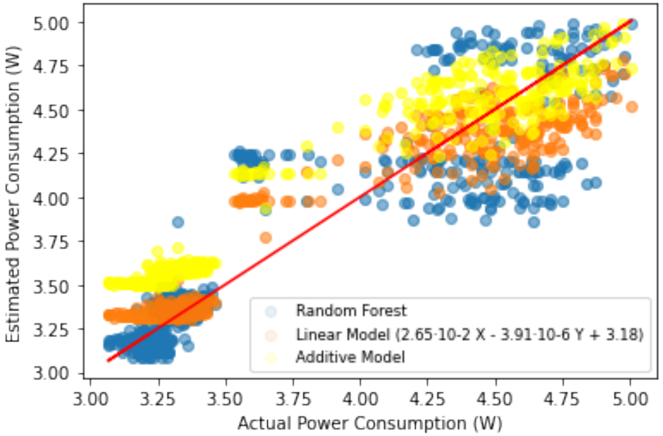
- Processing and data dissemination are the main energy drains in embedded and mobile devices

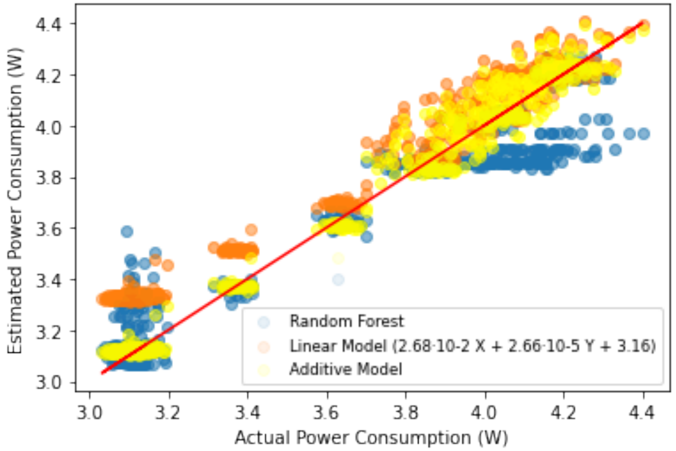
Mobility penalty map: power

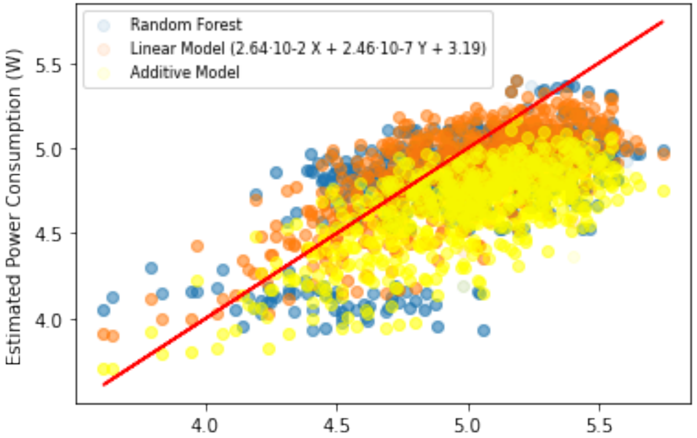

CPU Utilization vs Power Consumption

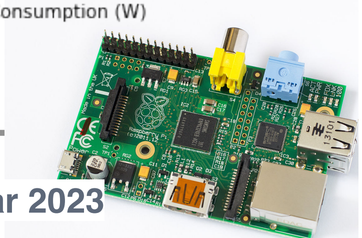

Kilobits Sent vs Power Consumption (Ethernet Upload)


Kilobits Received vs Power Consumption (Ethernet Download)


Estimated vs. Actual Power Consumption (WiFi Download)


Estimated vs. Actual Power Consumption (WiFi Upload)


Estimated vs. Actual Power Consumption (Ethernet Download)



Power Estimation Models for Edge Computing Devices, Kasioulis et al. Heteropar 2023

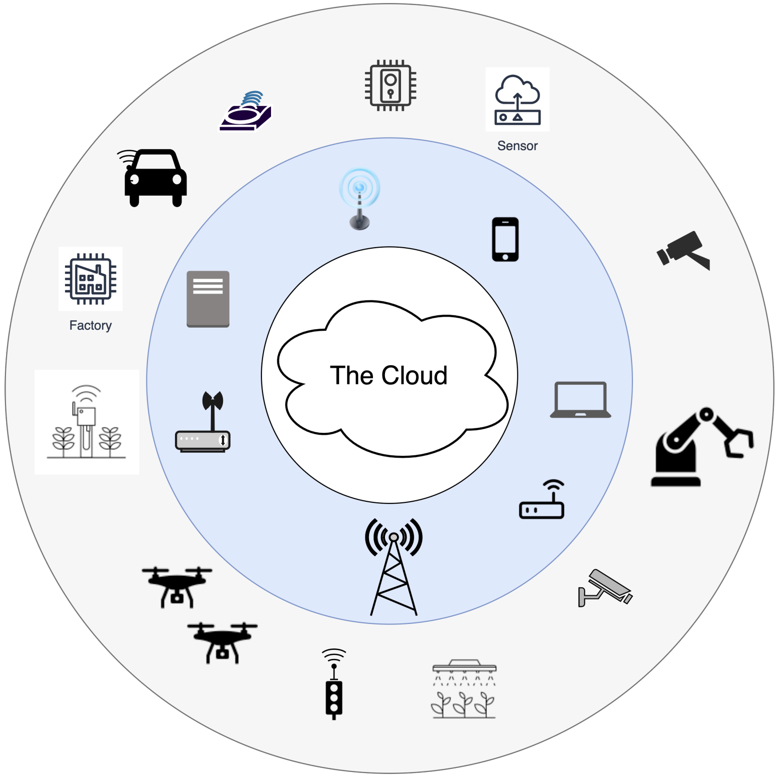
Overcoming the mobility penalty

- Offload computation over a wireless network to Tier-1, which offers:
  - Unlimited virtual resources
  - Well established tools and programming models
  - Stable execution environment and more...
    - E.g.: speech recognition and NLP in iOS, Android

Non-mobile Tier-3 Devices

- Typically, IoT devices are viewed as Tier-3 devices.
- Some are **not mobile**, but there is strong incentive for them to be inexpensive:
- Since this typically implies meagre processing capability:
  - offloading computation to Tier-1 is again attractive.

WHAT IS THE BLUEPRINT OF IOT APPLICATIONS?

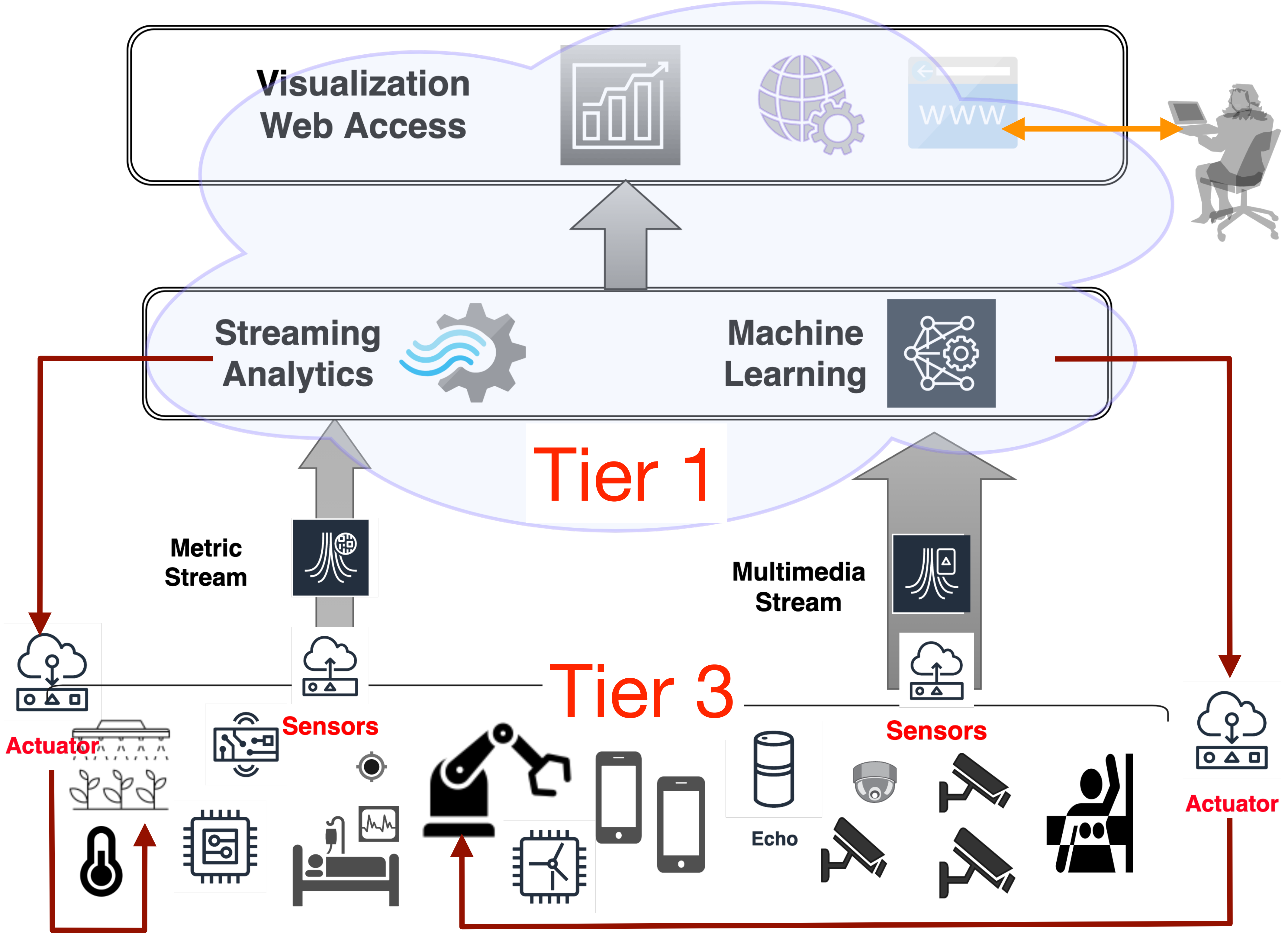


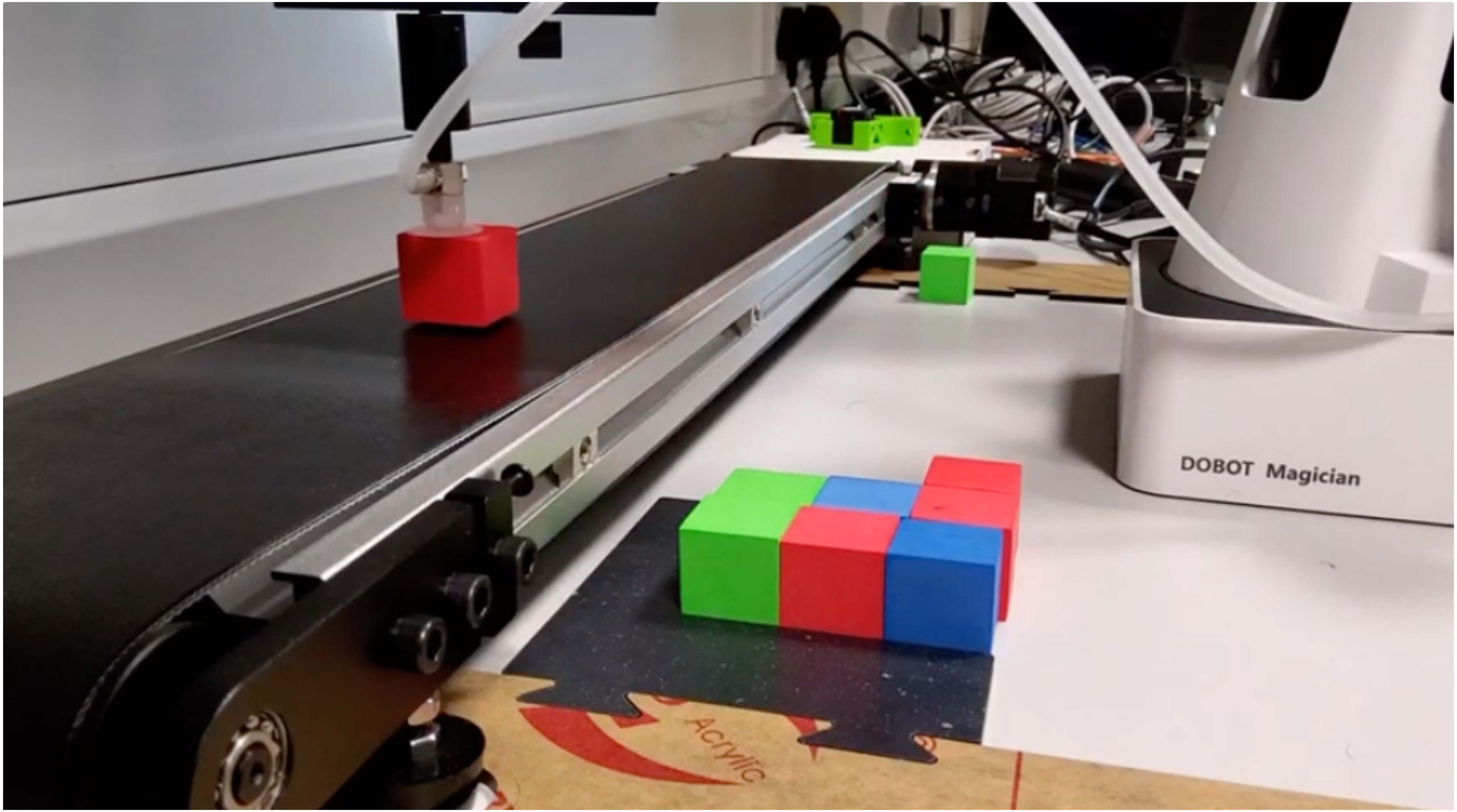
University of Cyprus
Department of Computer Science

M. D. Dikaiakos

IOT Applications Blueprint

Sensing on the IoT: Data-driven Applications


Visualization
Web Access
www
Streaming
Analytics
Machine
Learning
Tier 1
Metric
Stream
Multimedia
Stream
Tier 3
Actuator
Sensors
Sensors
Echo
Actuator


DOBOT Magician
Acrylic

IoT Applications' Functionalities

- Data collection from IoT sensors: Sensing the Physical (and Social, Urban, Human…) environment
- Storage, cleaning, curation, fusion, annotation & modeling of data: Data Management & Machine Learning
- Data processing & analysis: Analytics, Knowledge extraction, Inference, Prediction
- Data Visualization and Decision Support: typically via Web portals with RESTful APIs
- Actuation: Adapting the behavior of physical systems

IoT Applications' Requirements

- Low-latency when moving data from millions of distributed data sources to "central location" for processing to ensure analytics timeliness
- Location awareness: IoT sensors operate and sense locales
- Widespread geographical distribution
- Real-time or near real-time requirements to support timely actuation and decision-support

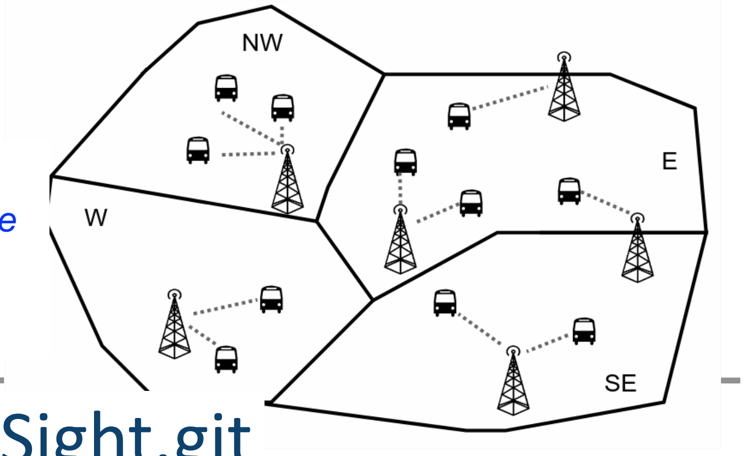
Data-driven Applications

Case Study: City Buses Application

Application Example: City Buses

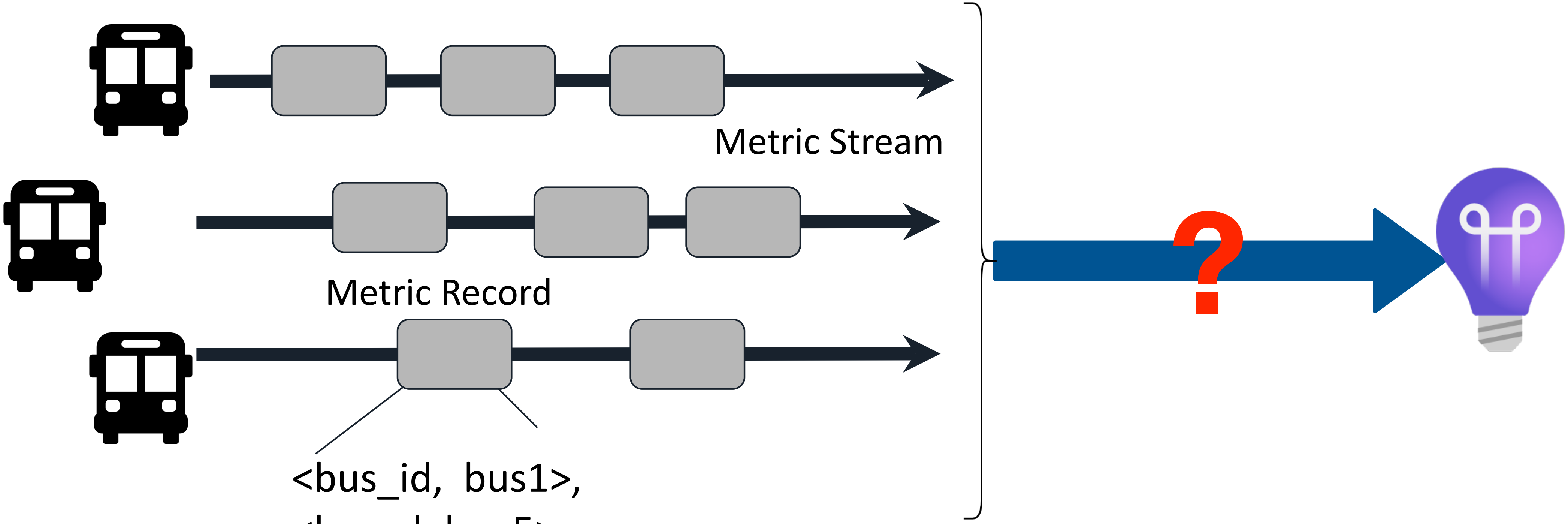
- Buses equipped with GPS tracking devices emitting updates of location to central server.
- Bus updates include: bus id, location coordinates, operating city region, an estimation of the current bus route delay, etc.

"***StreamSight: A Query-Driven Framework for Streaming Analytics in Edge Computing***", *Georgiou, Symeonides, Trihinas, Pallis, Dikaiakos, Proceedings of the 11th International Conference on Utility and Cloud Computing" (UCC '18), 2018*



https://github.com/UCY-LINC-LAB/StreamSight.git

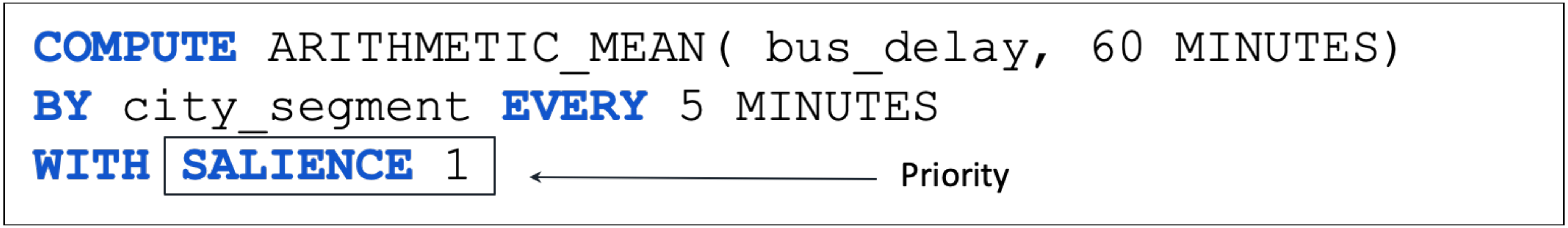
Application Example: City Buses



<bus_id, bus1>,
<bus_delay, 5>
<bus_region, nw>

16 metrics/record including: bus_id, bus_delay, city_segment

City Bus Analytics



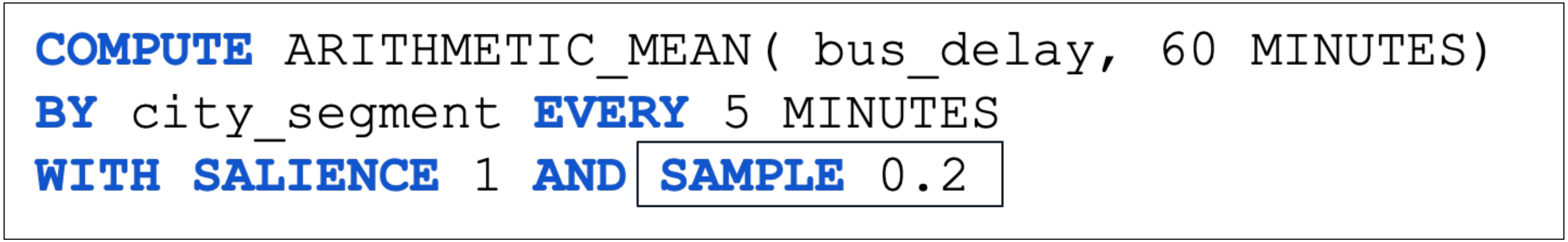
```
COMPUTE ARITHMETIC_MEAN( bus_delay, 60 MINUTES)
BY city_segment EVERY 5 MINUTES
WITH SALIENCE 1
```

Priority

- Query prioritisation

Query-Driven Descriptive Analytics for IoT and Edge Computing, *Georgiou, Symeonides, Trihinas, Pallis, Dikaiakos, IEEE International Conference on Cloud Engineering (IC2E 2019)*

City Bus Analytics

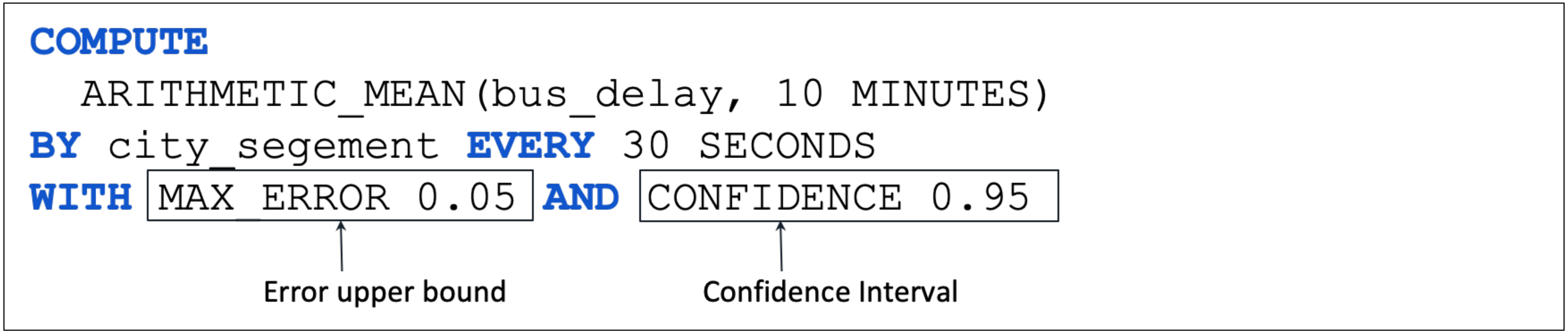


```
COMPUTE ARITHMETIC_MEAN( bus_delay, 60 MINUTES)
BY city_segment EVERY 5 MINUTES
WITH SALIENCE 1 AND SAMPLE 0.2
```

- Uniform sampling: apply the query on a portion of the data stream to increase query execution time.

***Query-Driven Descriptive Analytics for IoT and Edge Computing**, Georgiou, Symeonides, Trihinas, Pallis, Dikaiakos, IEEE International Conference on Cloud Engineering (IC2E 2019)*

City Bus Analytics



```
COMPUTE
  ARITHMETIC_MEAN(bus_delay, 10 MINUTES)
BY city_segement EVERY 30 SECONDS
WITH MAX_ERROR 0.05 AND CONFIDENCE 0.95
```

- Sampling with Error Margin & Confidence

***Query-Driven Descriptive Analytics for IoT and Edge Computing**, Georgiou, Symeonides, Trihinas, Pallis, Dikaiakos, IEEE International Conference on Cloud Engineering (IC2E 2019)*

City Bus Analytics

```
COMPUTE
  PEWMA[0.5](bus_delay) BY bus_id
EVERY 30 SECONDS
WITH MAX_ERROR 0.05 AND CONFIDENCE 0.95
     AND AWARENESS ON ACCURACY
```

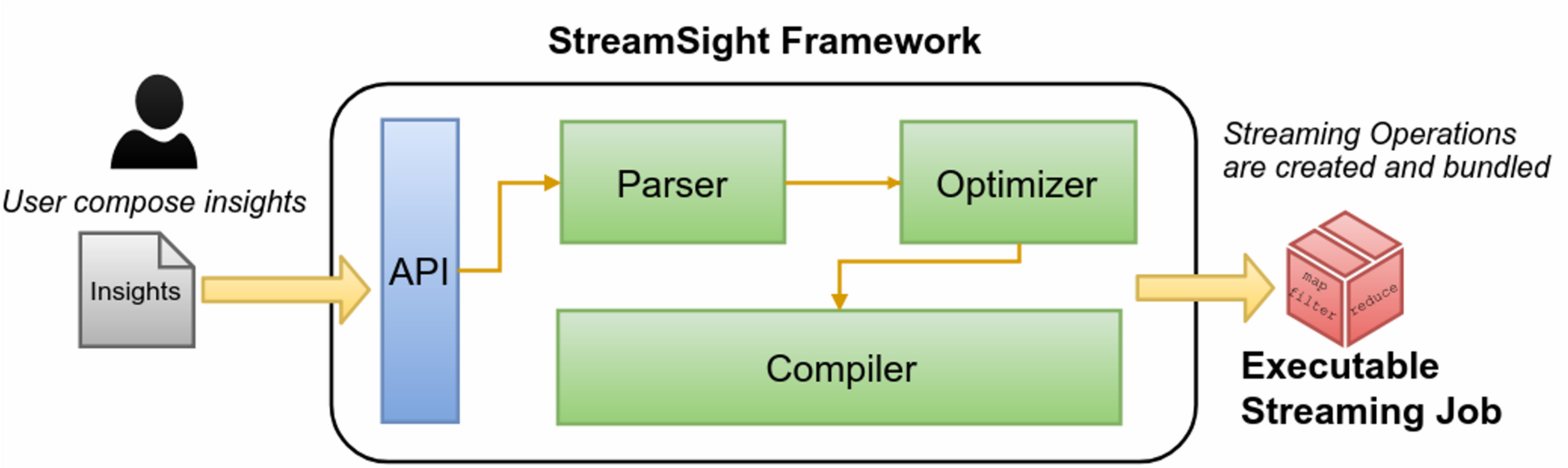
- PEWMA: probabilistic exponentially weighted moving average

***Query-Driven Descriptive Analytics for IoT and Edge Computing**, Georgiou, Symeonides, Trihinas, Pallis, Dikaiakos, IEEE International Conference on Cloud Engineering (IC2E 2019)*

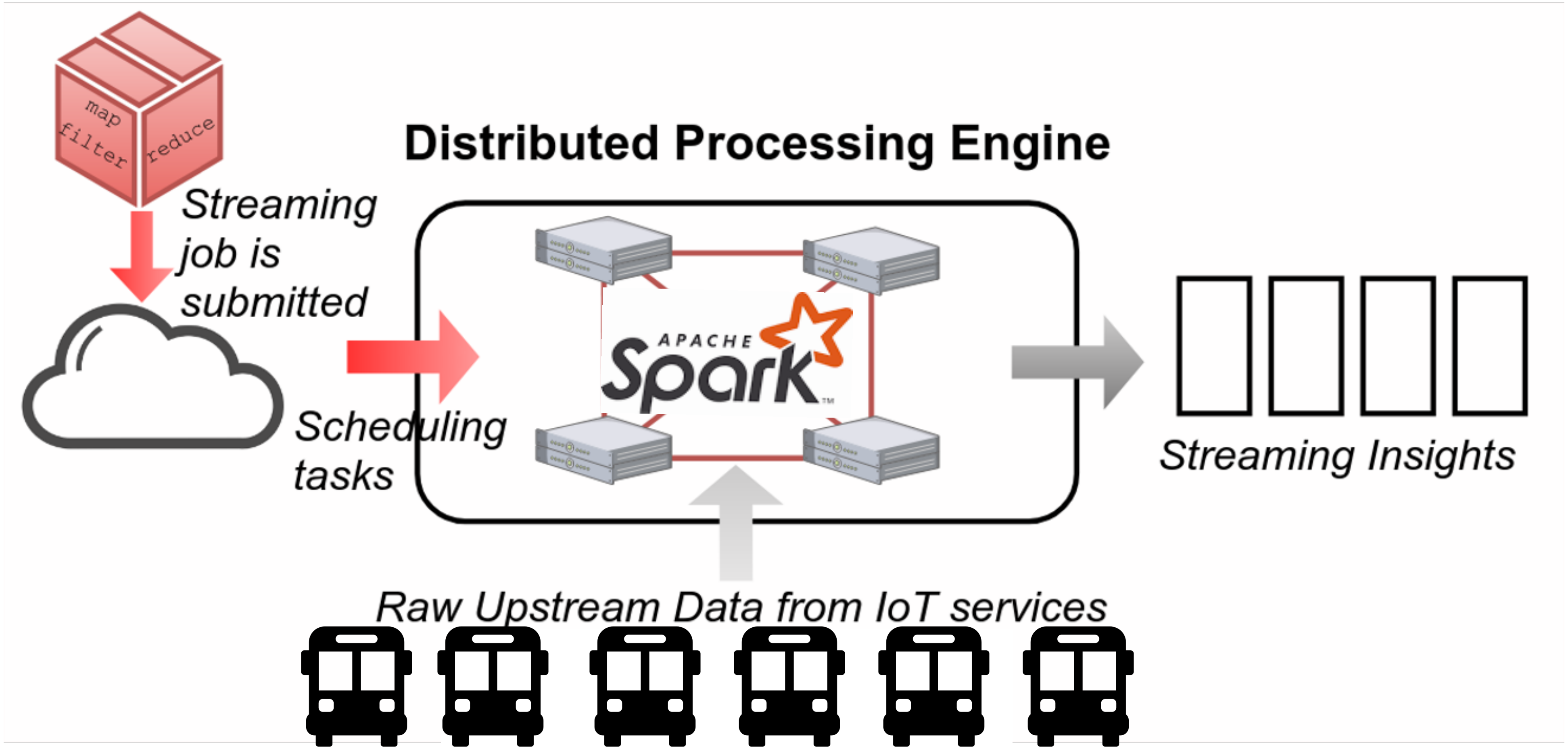
***StreamSight: A Query-Driven Framework for Streaming Analytics in Edge Computing**, Georgiou, Symeonides, Trihinas, Pallis, Dikaiakos, "Proceedings of the 11th International Conference on Utility and Cloud Computing" (UCC '18), 2018*

StreamSight Framework

- A framework for the specification, the compilation, and execution of streaming analytic queries on distributed processing engines optimized for edge computing environments

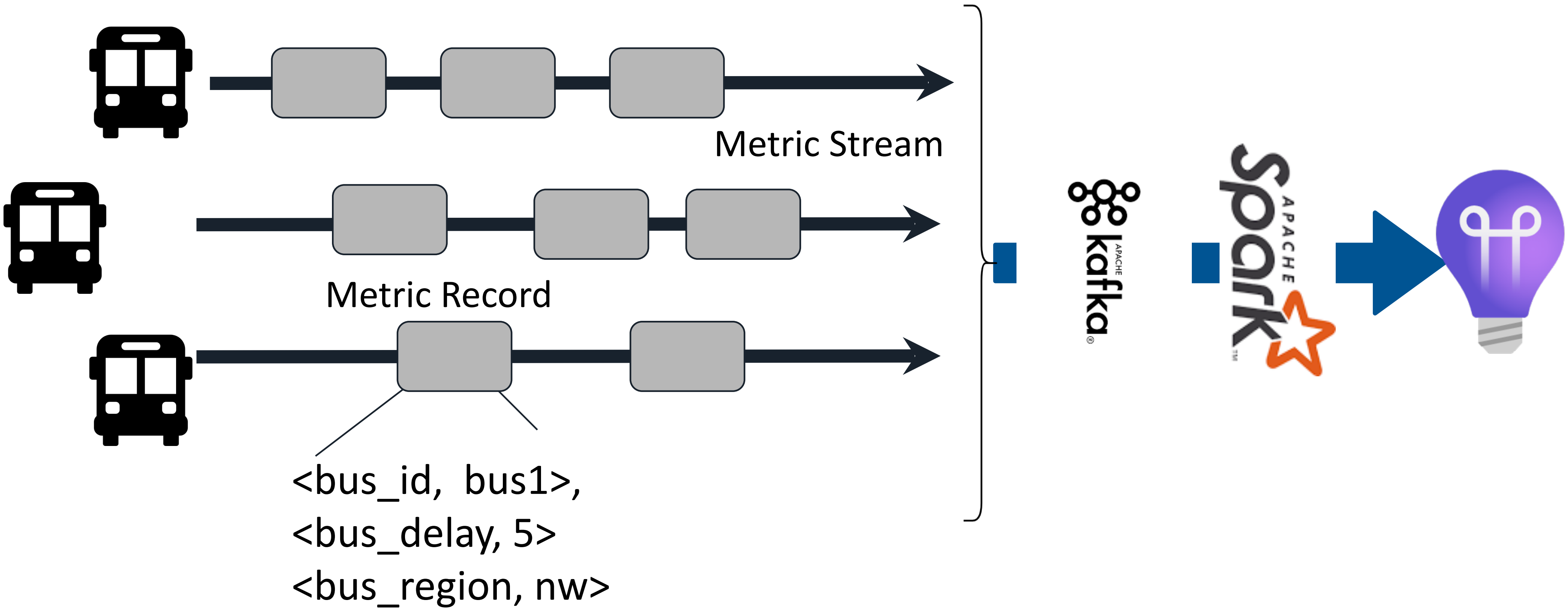


Streaming Analytics Pipeline



map
filter
reduce
Streaming job is submitted
Distributed Processing Engine
Scheduling tasks
APACHE Spark
Streaming Insights
Raw Upstream Data from IoT services

Application Example: City Buses



Metric Stream

Metric Record

<bus_id, bus1>,
<bus_delay, 5>
<bus_region, nw>

16 metrics/record including: bus_id, bus_delay, city_segment

StreamSight: A Query-Driven Framework for Streaming Analytics in Edge Computing, *Georgiou, Symeonides, Trihinas, Pallis, Dikaiakos, "Proceedings of the 11th International Conference on Utility and Cloud Computing" (UCC '18), 2018*

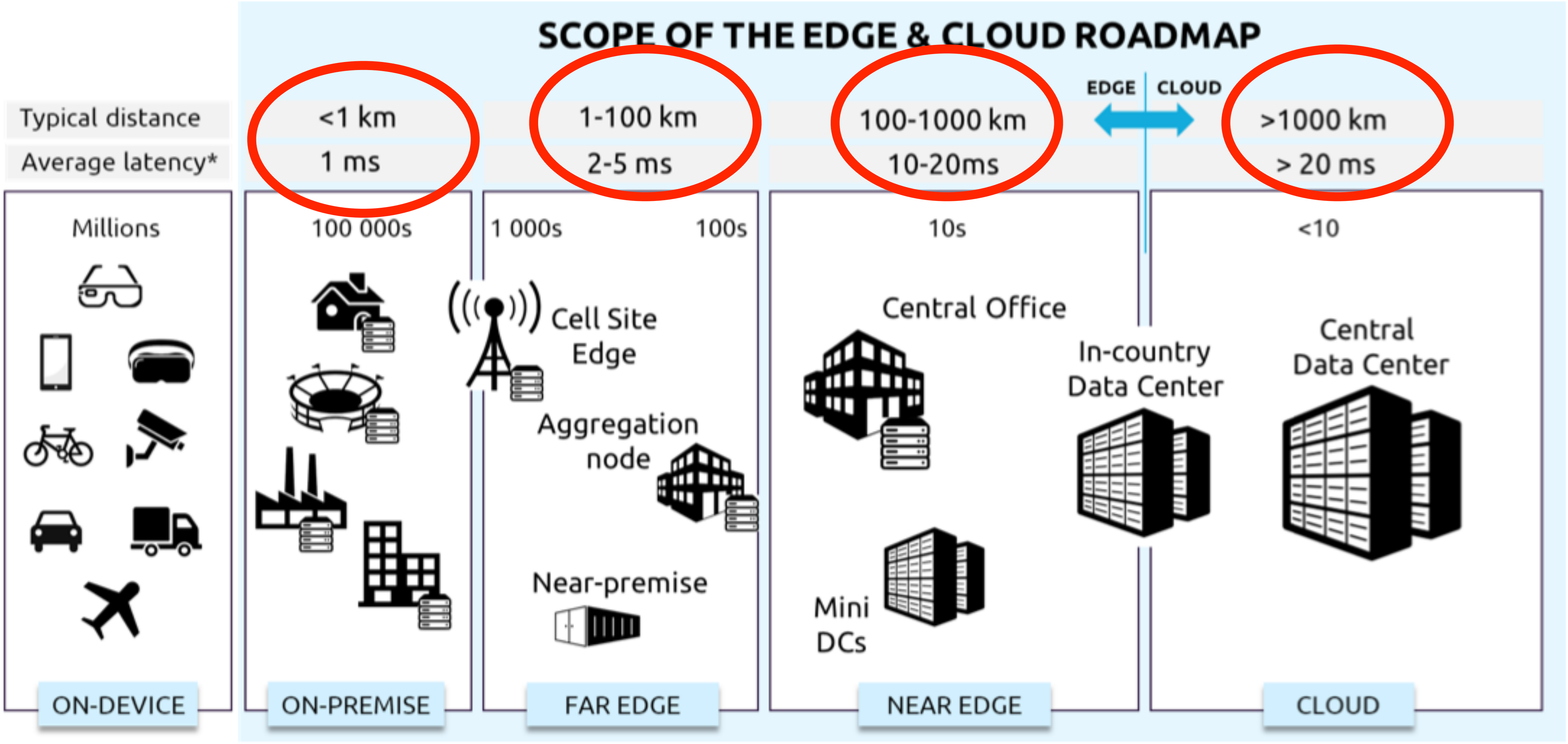
https://github.com/UCY-LINC-LAB/StreamSight.git

CAN THE CLOUD COPE WITH THESE REQUIREMENTS?

Cloud/Tier-1 Shortcomings

- Due to its extreme consolidation, the Cloud:
  - Offers services out of huge data centers in “centralized” locations
  - Must accommodate streams from millions of IoT devices, which are immersed into our physical, urban, social environment
- Shortcomings:
  - Longer network round-trip times (RTT) to Tier-1 from Tier-3
    - Delays can be critical in actuation scenarios with near-real time constraints.
  - Huge cumulative **ingress** bandwidth demand into Tier-1 data centers
    - Cloud network connections become a bottleneck

Latency to the Cloud



SCOPE OF THE EDGE & CLOUD ROADMAP

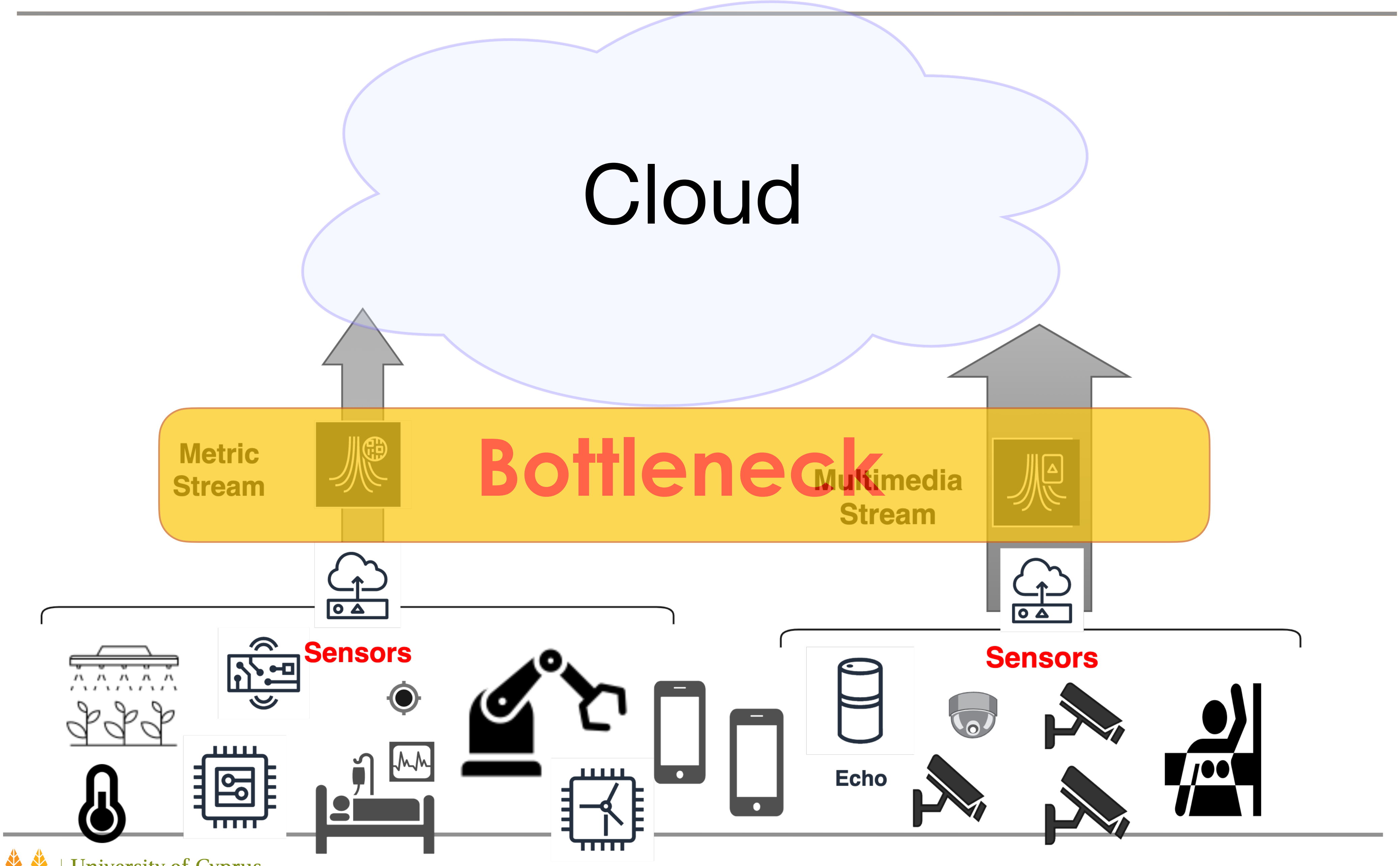
	ON-DEVICE	ON-PREMISE	FAR EDGE	NEAR EDGE	CLOUD
Typical distance		<1 km	1-100 km	100-1000 km	>1000 km
Average latency*		1 ms	2-5 ms	10-20ms	> 20 ms
	Millions	100 000s	1 000s 100s	10s	<10

* Latency does not depend only on distance. Other factors influencing latency are a) access technology (latency in 5G or FTTH much lower than in 4G), b) transport topology and technology, c) core network configuration (user plane location, breakout point), d) network optimization (traffic prioritization, bandwidth allocation, Edge node selection).

***Figure**: Scope of the Industrial Roadmap in the cloud-edge continuum*

Source: European Industrial Technology Roadmap for the next Generation Cloud-Edge Offering. 2021.

Cloud/Tier-1 Shortcomings



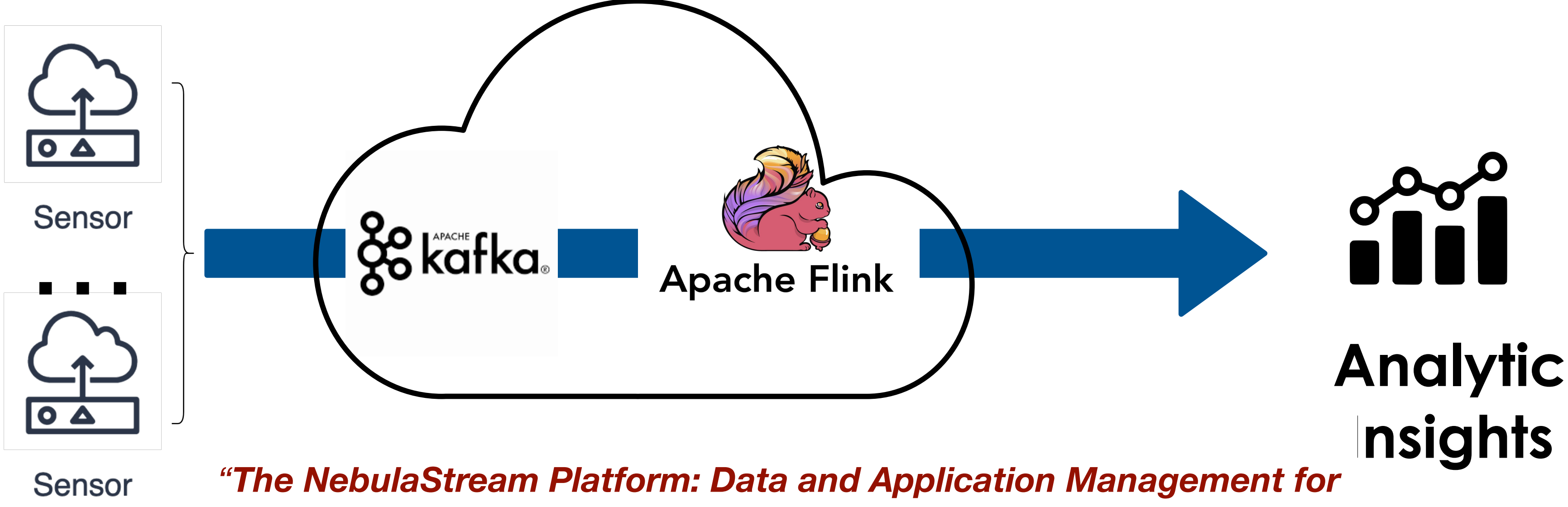
Case Study

IOT TO CLOUD BOTTLENECKS

University of Cyprus
Department of Computer Science

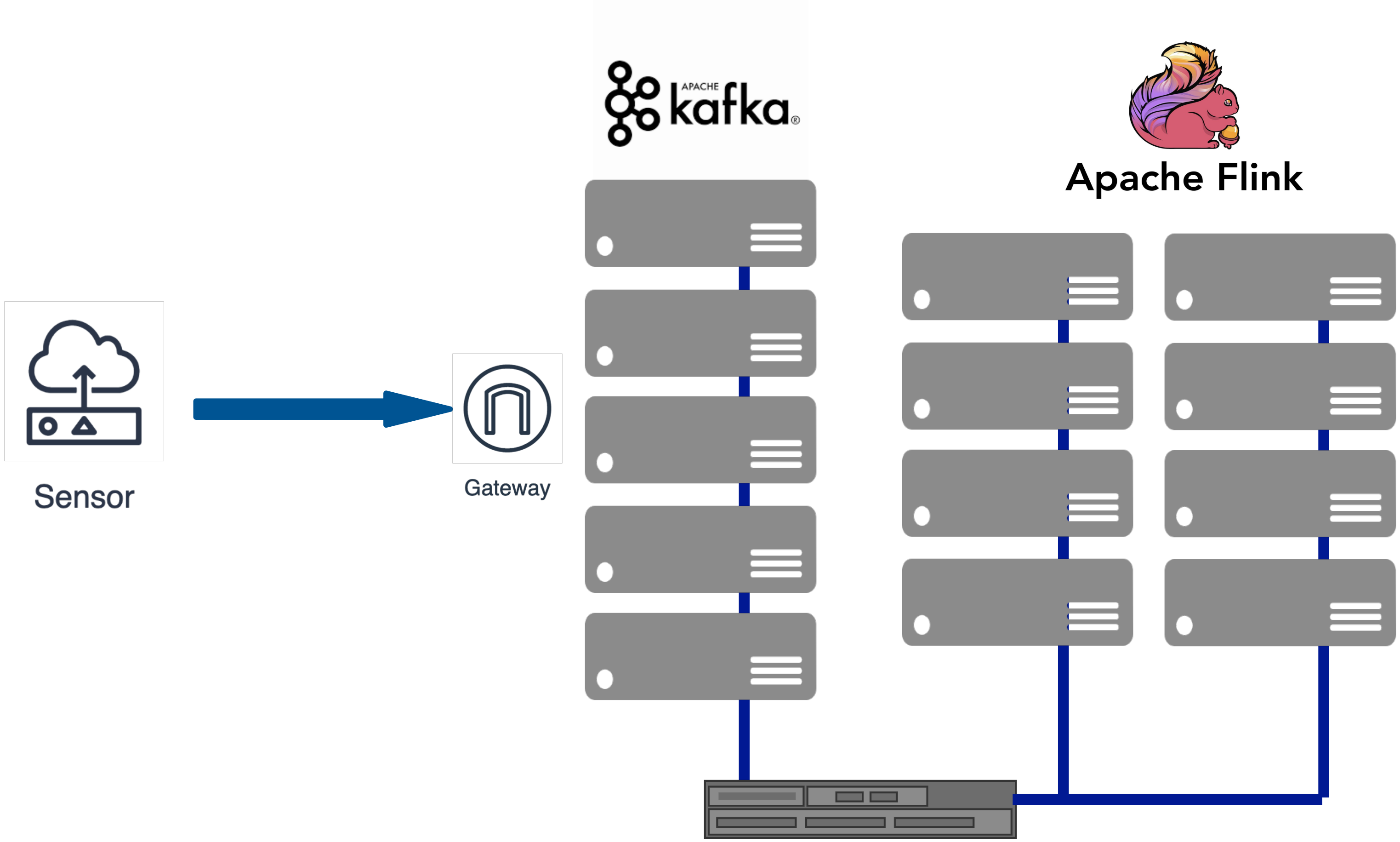
Bottleneck: An Example

- 1 to 80 IoT data producers:
  - Each producer generates data at a constant speed of 50K record/sec.
  - Producers send their data over a gateway to a Cloud.
- Cloud service comprises:
  - Kafka cluster with five nodes (receiving incoming metric stream)
  - Flink cluster with eight nodes w. 1 Gbit Ethernet connection.
  - Flink job reads data from Kafka and executes a simple windowed aggregation of 10 seconds to count distinct events.



"The NebulaStream Platform: Data and Application Management for the Internet of Things." *Zeuch, Chaudhary, Del Monte, Gavriilidis, Giouroukis, Grulich, Breß, Traub, Mark. CIDR 2020.*

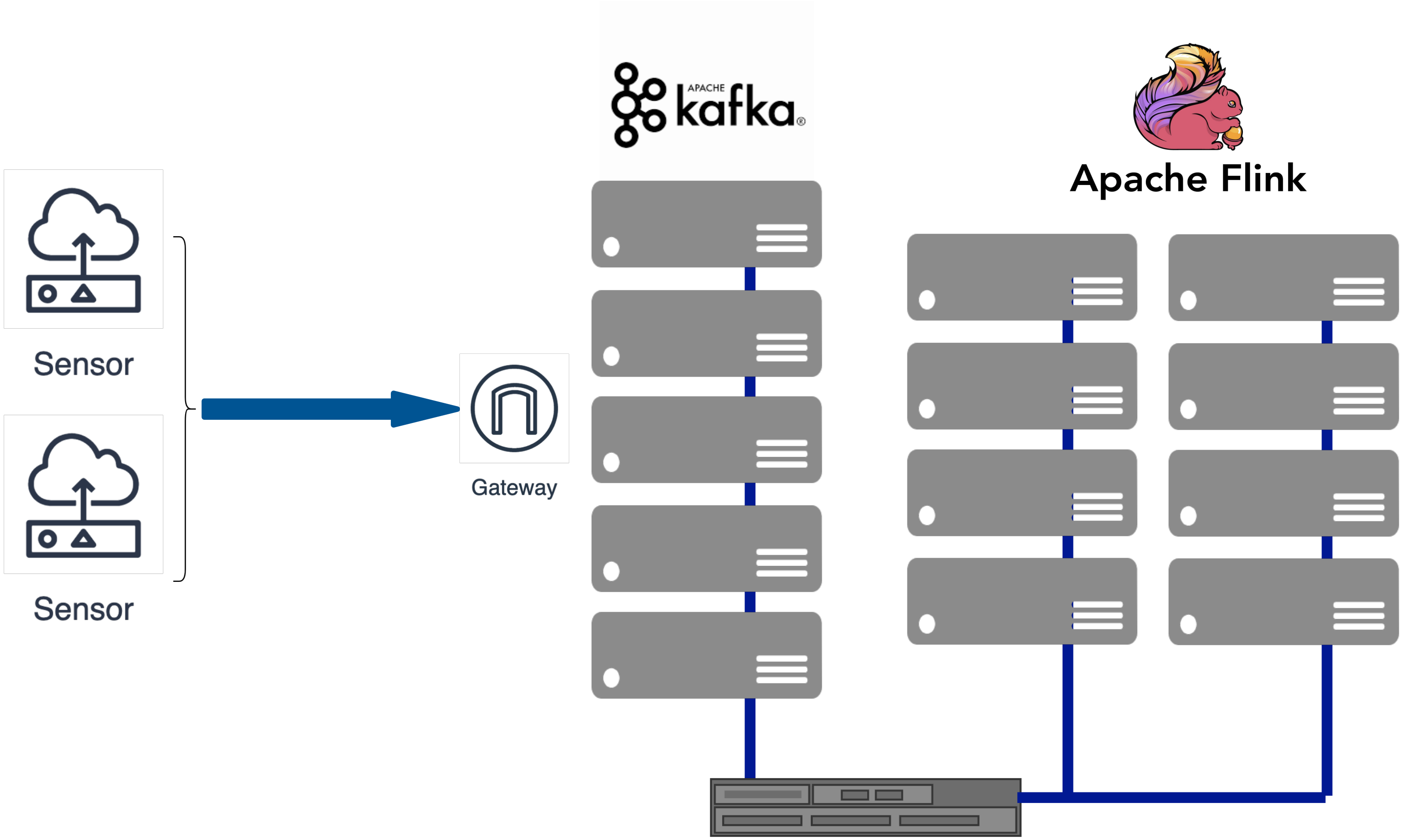
Bottleneck: An Example



Let experiment run for 10 minutes and measure end-to-end processing latency

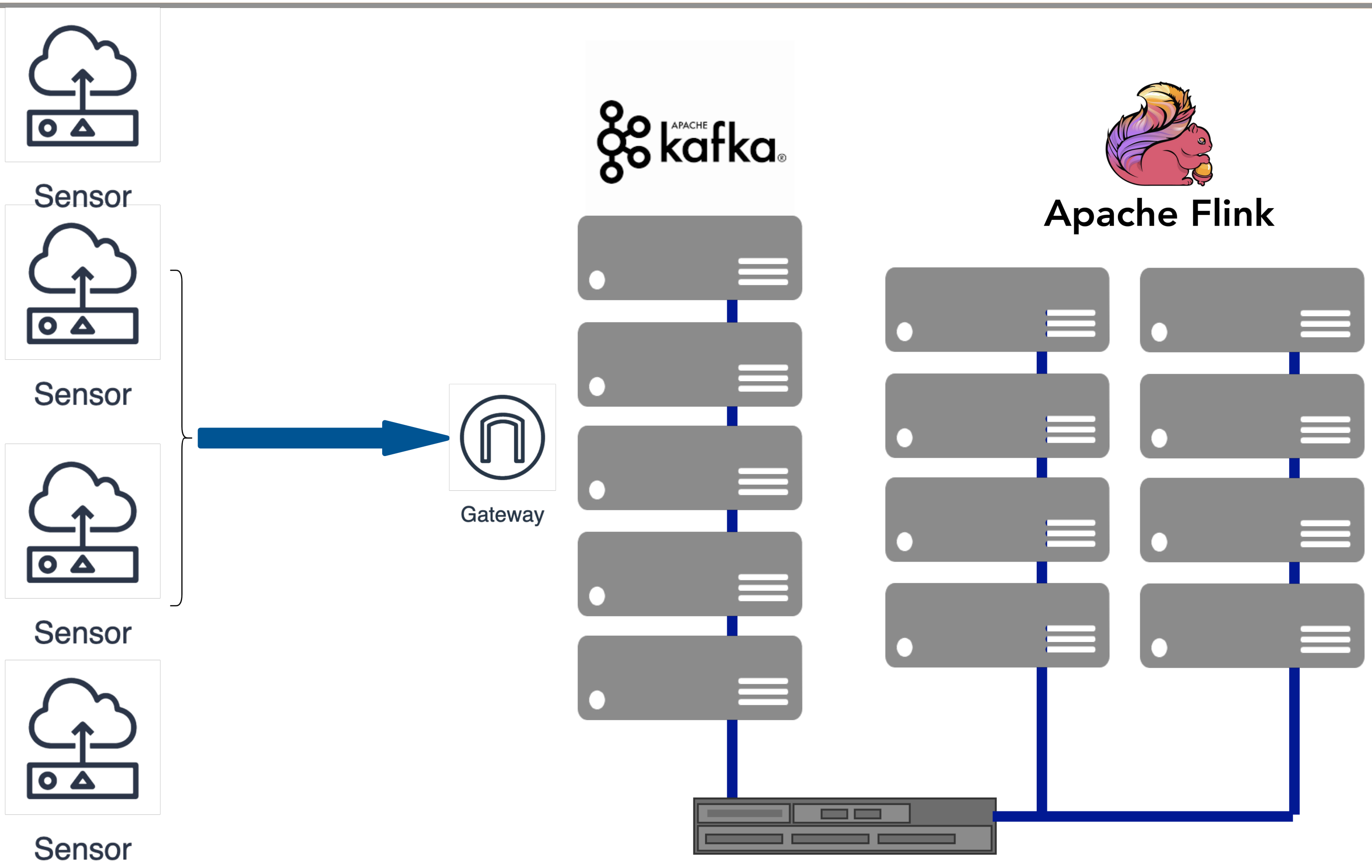
***"The NebulaStream Platform: Data and Application Management for the Internet of Things**." Zeuch, Chaudhary, Del Monte, Gavriilidis, Giouroukis, Grulich, Breß, Traub, Mark. CIDR 2020.*

Bottleneck: An Example



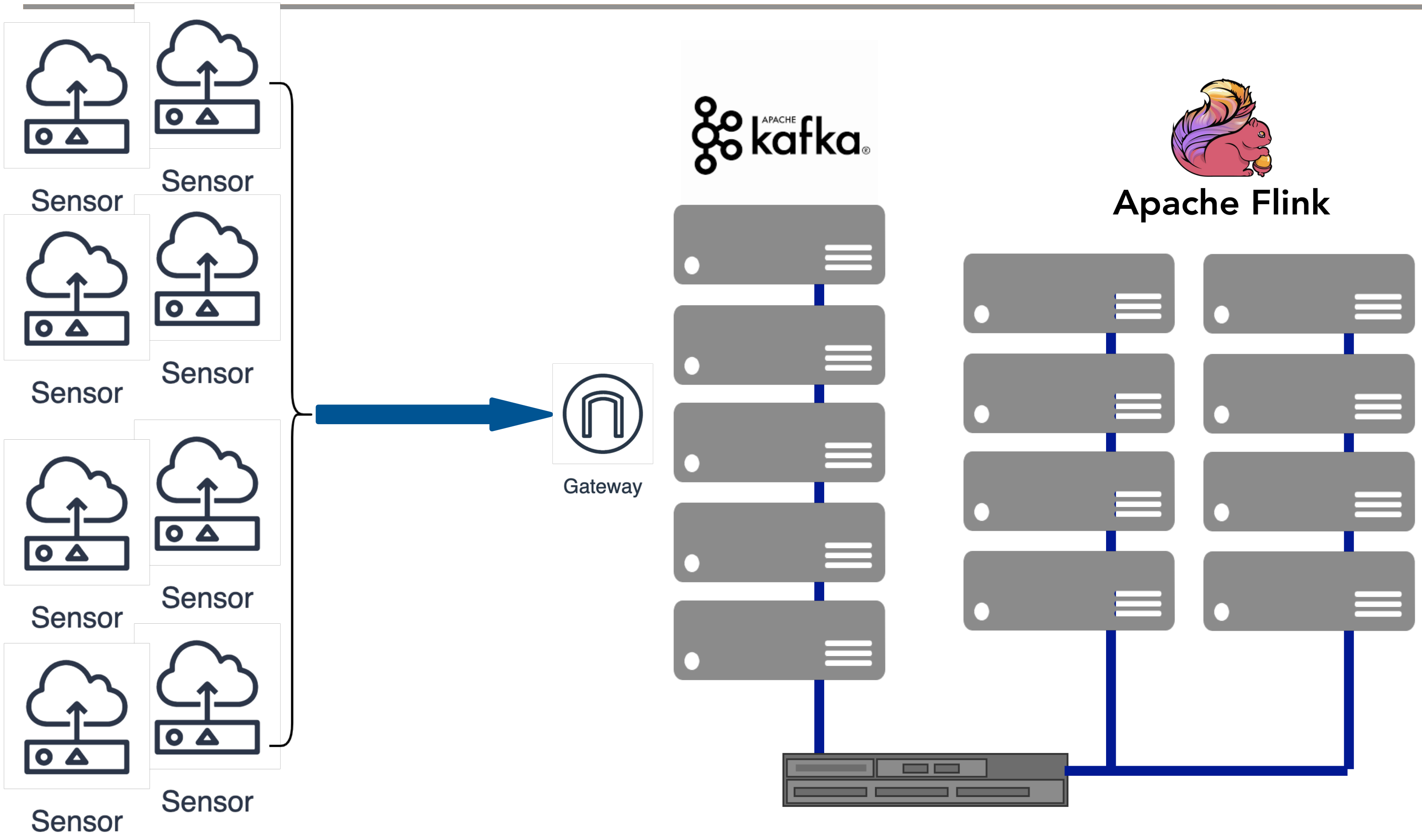
"The NebulaStream Platform: Data and Application Management for the Internet of Things." *Zeuch, Chaudhary, Del Monte, Gavriilidis, Giouroukis, Grulich, Breß, Traub, Mark. CIDR 2020.*

Bottleneck: An Example



*“**The NebulaStream Platform: Data and Application Management for the Internet of Things.**” Zeuch, Chaudhary, Del Monte, Gavriilidis, Giouroukis, Grulich, Breß, Traub, Mark. CIDR 2020.*

Bottleneck: An Example



“The NebulaStream Platform: Data and Application Management for the Internet of Things.” Zeuch, Chaudhary, Del Monte, Gavriilidis, Giouroukis, Grulich, Breß, Traub, Mark. CIDR 2020.

Bottleneck: An Example

- Latency increases with the number of producers.
- The cloud-based IoT application scenario can sustain up to 20 producers with constant latency.
- Beyond 20 producers, the application saturates and latency increases gradually.
- Centralized cloud approach does not scale for IoT applications and thus future IoT applications require a new system.

*“**The NebulaStream Platform: Data and Application Management for the Internet of Things**.” Zeuch, Chaudhary, Del Monte, Gavriilidis, Giouroukis, Grulich, Breß, Traub, Mark. CIDR 2020.*

Bottleneck: An Example

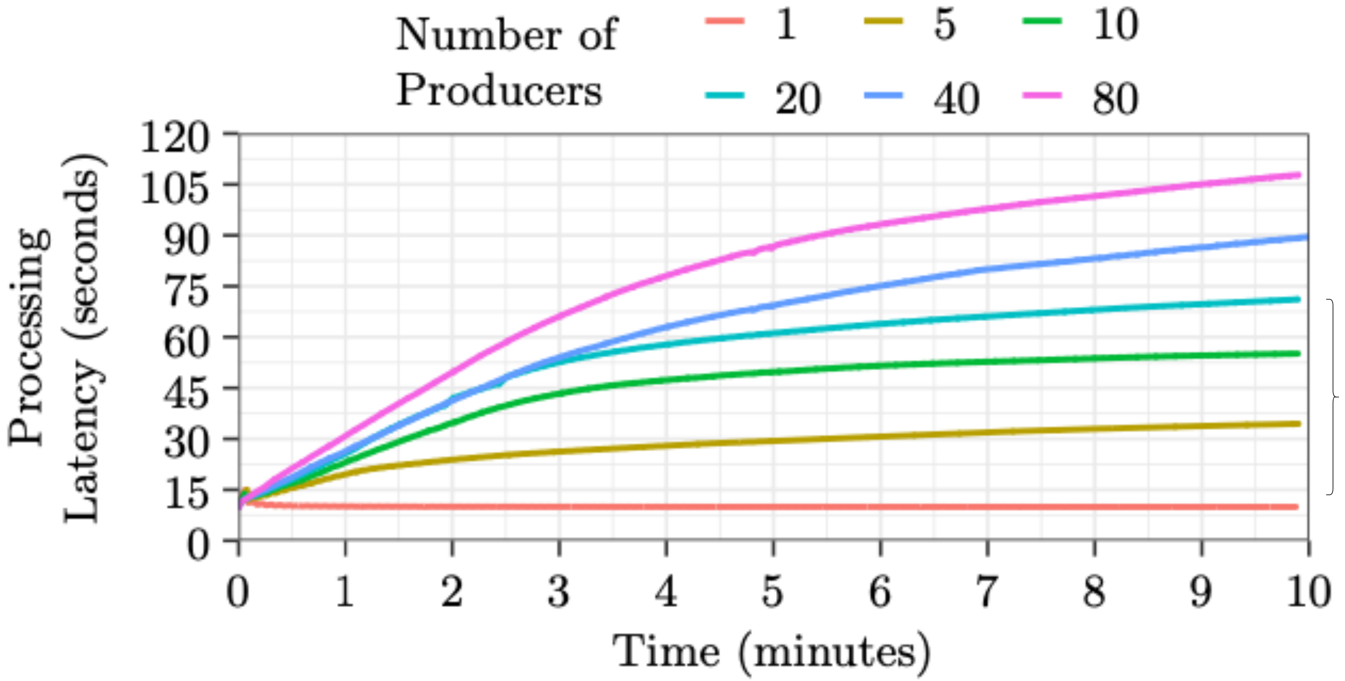
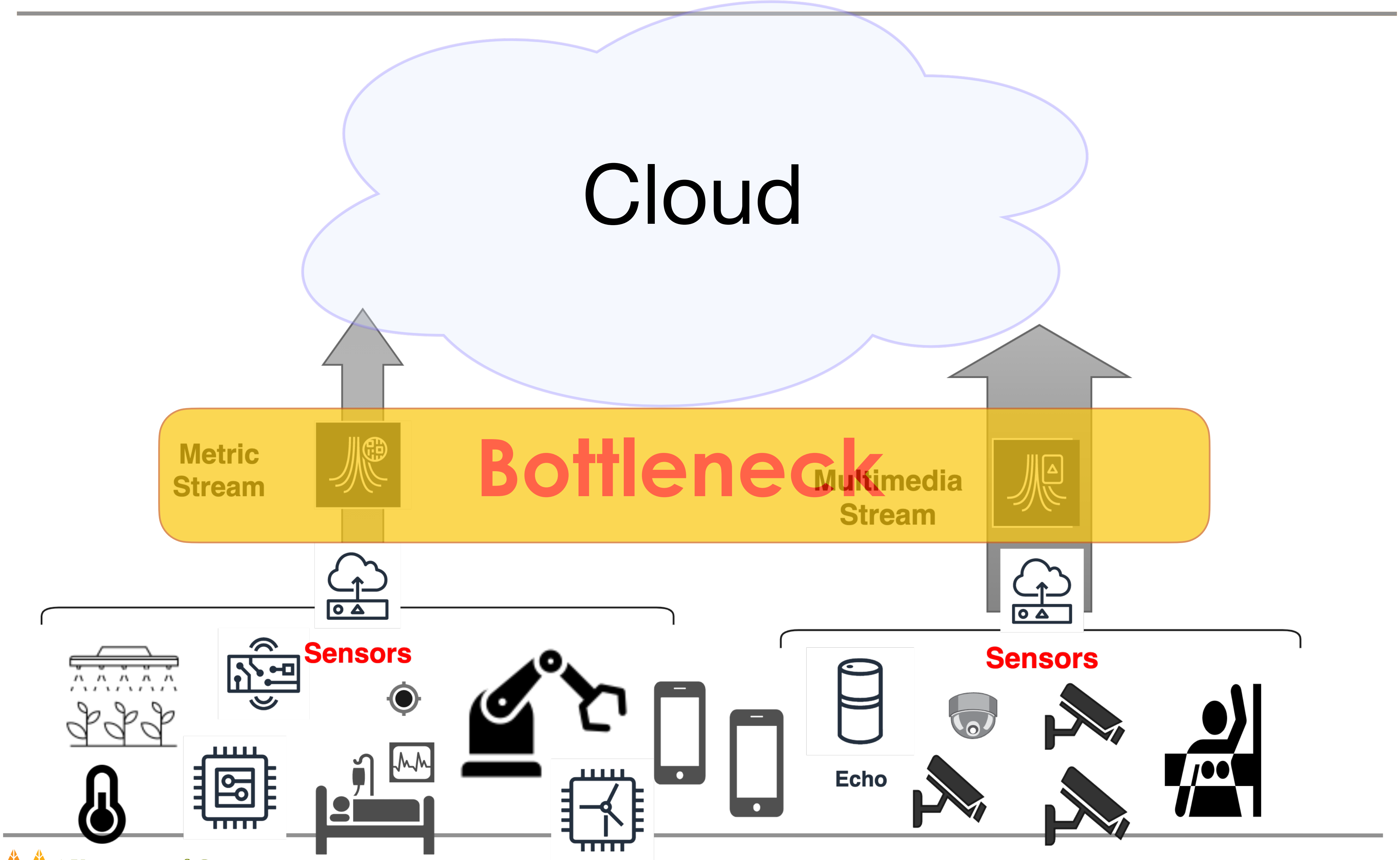


Figure 1: IoT application using a cloud-centric SPE.

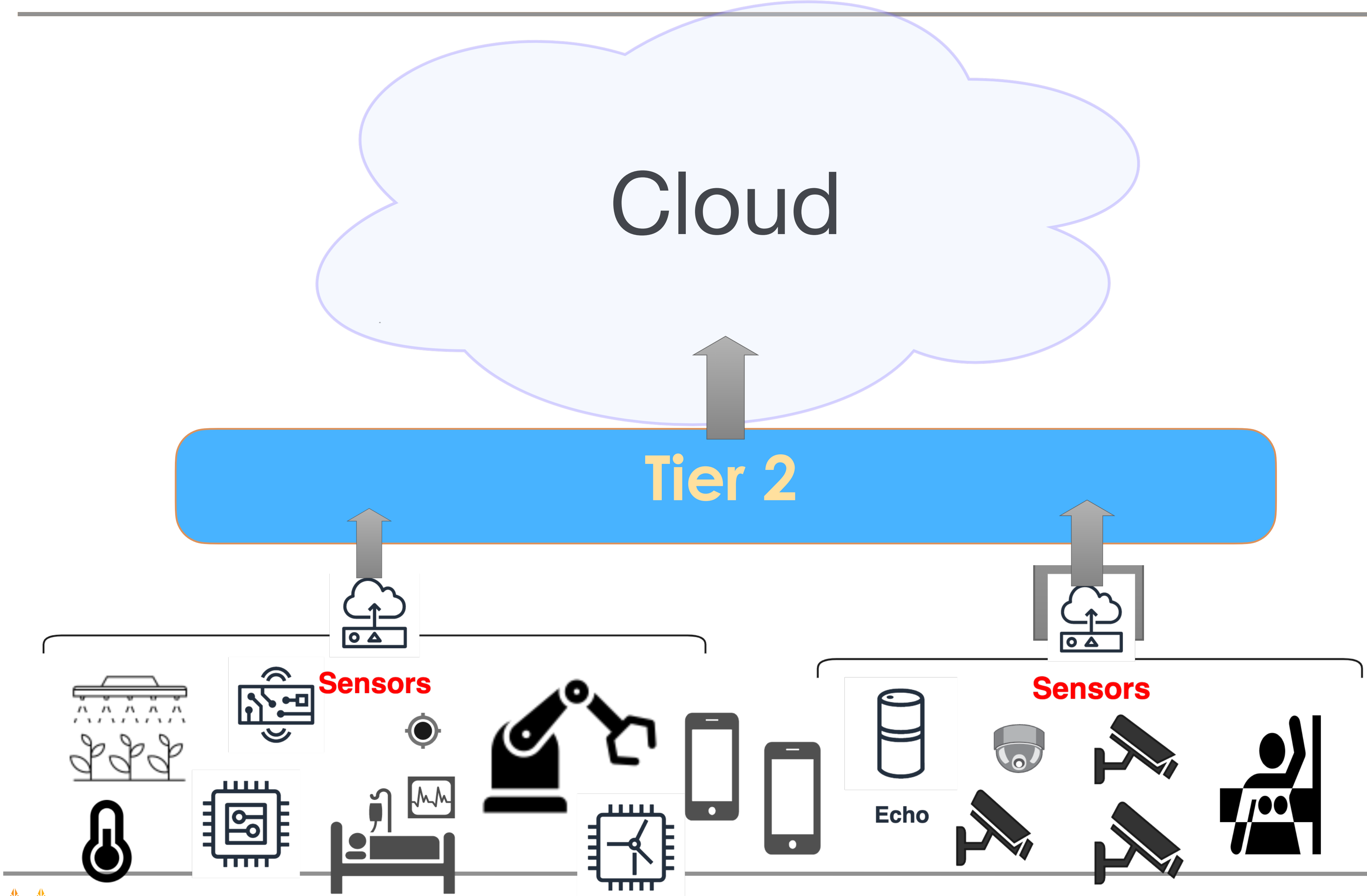
Beyond 20 producers, the application saturates and latency increases gradually.

*“**The NebulaStream Platform: Data and Application Management for the Internet of Things.**” Zeuch, Chaudhary, Del Monte, Gavriilidis, Giouroukis, Grulich, Breß, Traub, Mark. CIDR 2020.*

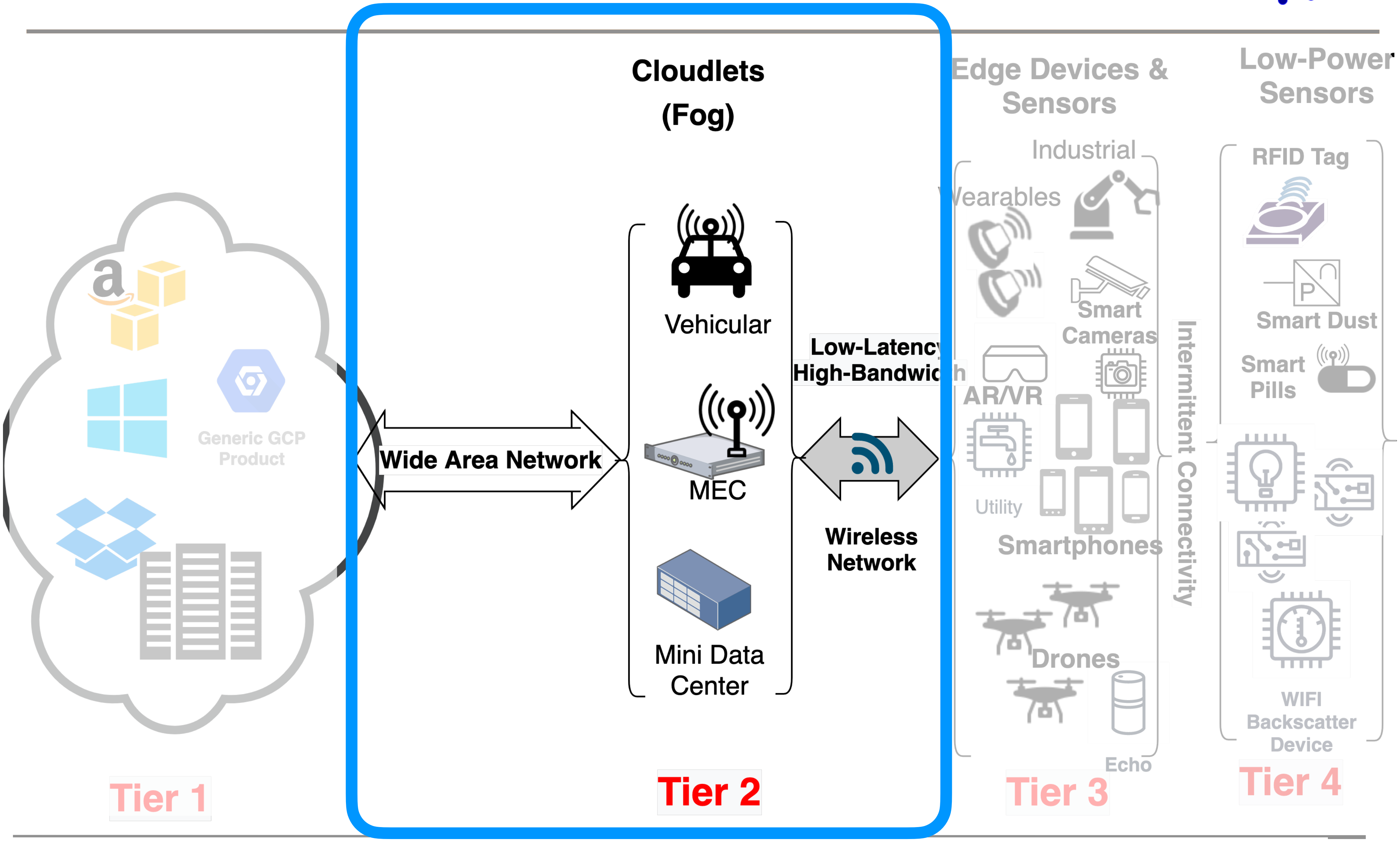
Cloud/Tier-1 Shortcomings



Tier 2: Fog



Tier 2



Mahadev Satyanarayanan, Wei Gao, and Brandon Lucia. 2019. ***The Computing Landscape of the 21st Century****.HotMobile '19*

Tier 2 - Network proximity

- Offers "Network Proximity" to address Tier-1 shortcomings
- Creates the illusion of bringing Tiers 1 and 3 "closer"
- **Offloading** of tier-3 compute-intensive operations to Tier-2, at very low latency, preserving **tight response time bounds** needed for immersive user experience and cyber-physical systems.
  - Much smaller fan-in between Tiers-3 and -2 than when Tier-3 devices connect directly to Tier-1.
  - Tier-2 processing of data captured at Tier-3 avoids excessive bandwidth demand.

A note on “proximity”

- Network proximity rather than physical proximity.
- It is crucial that RTT be low and end-to- end bandwidth be high.
  - This is achievable by using a fiber link between a wireless access point and a cloudlet that is many tens or even hundreds of kilometers away.
- Physical proximity does not guarantee network proximity.
  - A highly congested WiFi network may have poor RTT, even if Tier-2 is physically near Tier-3.

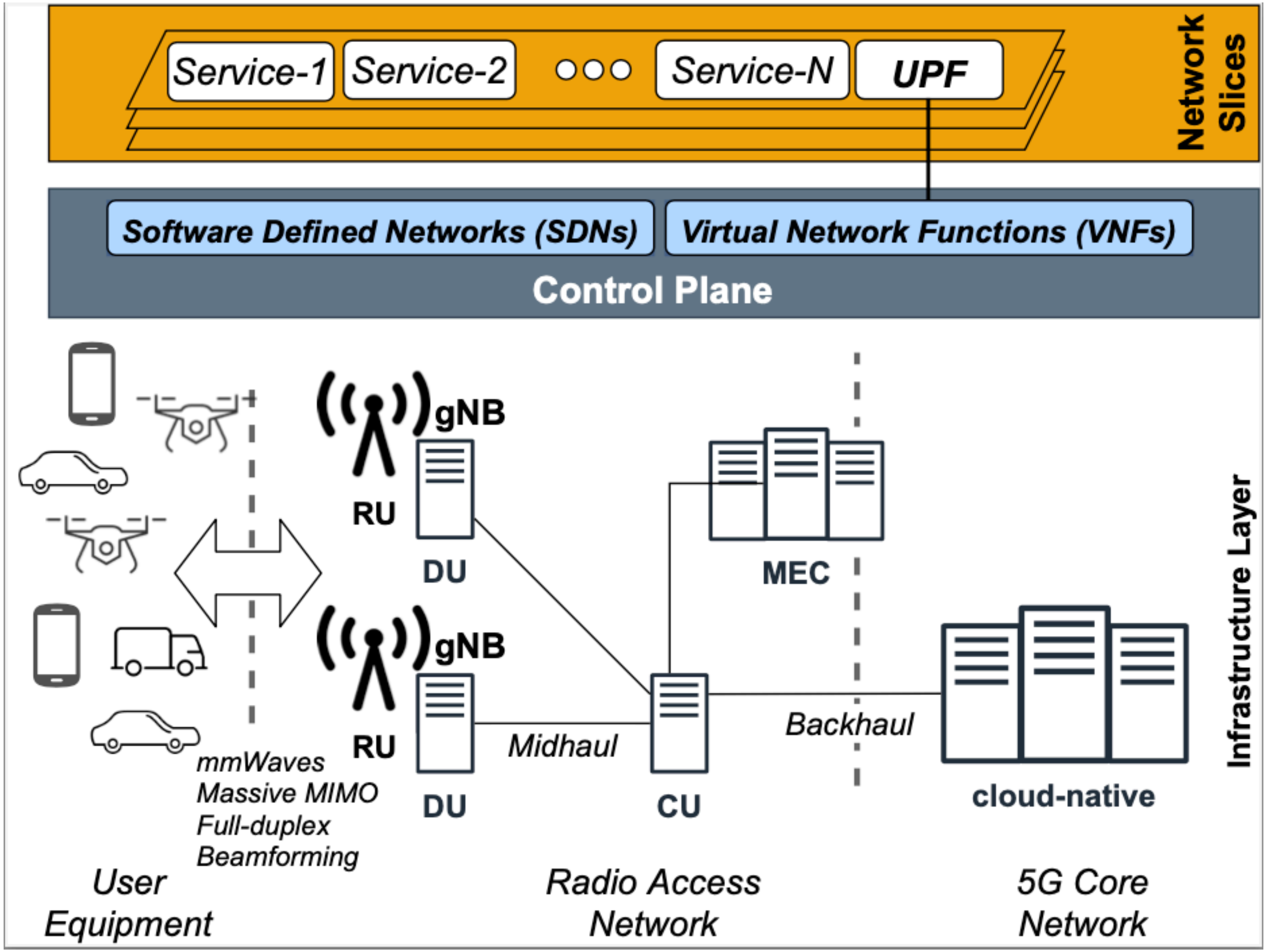
A note on hardware

- Server hardware at Tier-2 is essentially the same as at Tier-1, but engineered differently:
  - No extreme consolidation
  - Servers organized into small, dispersed data centers called cloudlets ("a data center in a box")

A note on wireless: 5G

- 5G targets emerging IoT applications by offering:
  - 1000x higher mobile data volume per unit area
  - support for 10–100x more connected devices
- Network slicing facilitates the provision of multiple logical networks on top of a physical network.
- Multi-access Edge Computing (MEC) servers can host network slices and receive offloaded tasks that cannot be executed on the edge devices.

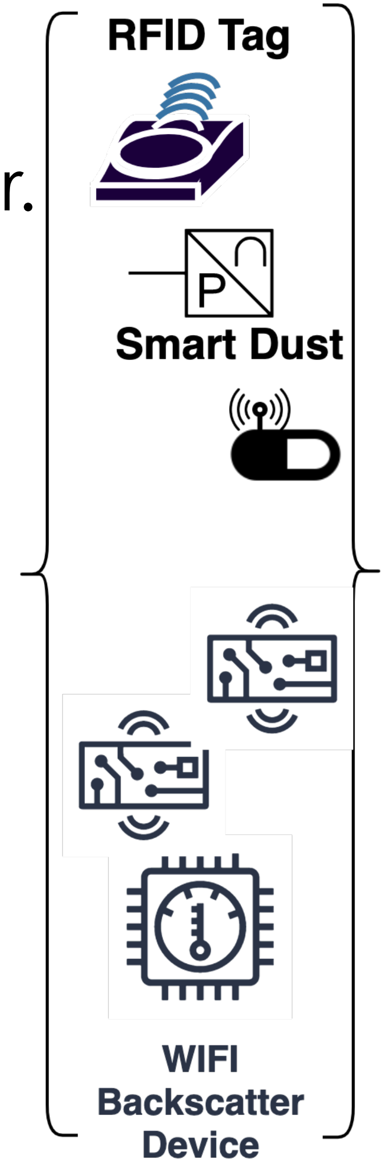
A note on wireless: 5G



“5G-Slicer: An emulator for mobile IoT applications deployed over 5G network slices.” Symeonides, Trihinas, Pallis, Dikaiakos, Psomas, Krikides, ACM/IEEE International Conference on Internet of Things Design and Implementation (IoTDI'22)

Tier 4: Longevity and Opportunism

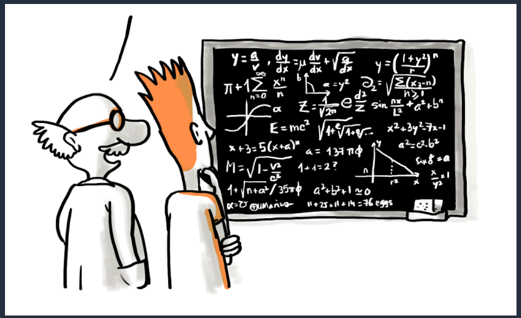
- No chemical energy source (battery)
- Harvest incident EM energy to charge capacitor.
  - Capacitor powers a brief episode of sensing, computation and wireless transmission.
  - Device remains passive until next occasion to harvest sufficient energy.
- **Intermittent computing:** no need for energy-related maintenance of devices in the field.
  - **Longevity** of deployment, combined with **opportunism** in energy harvesting.



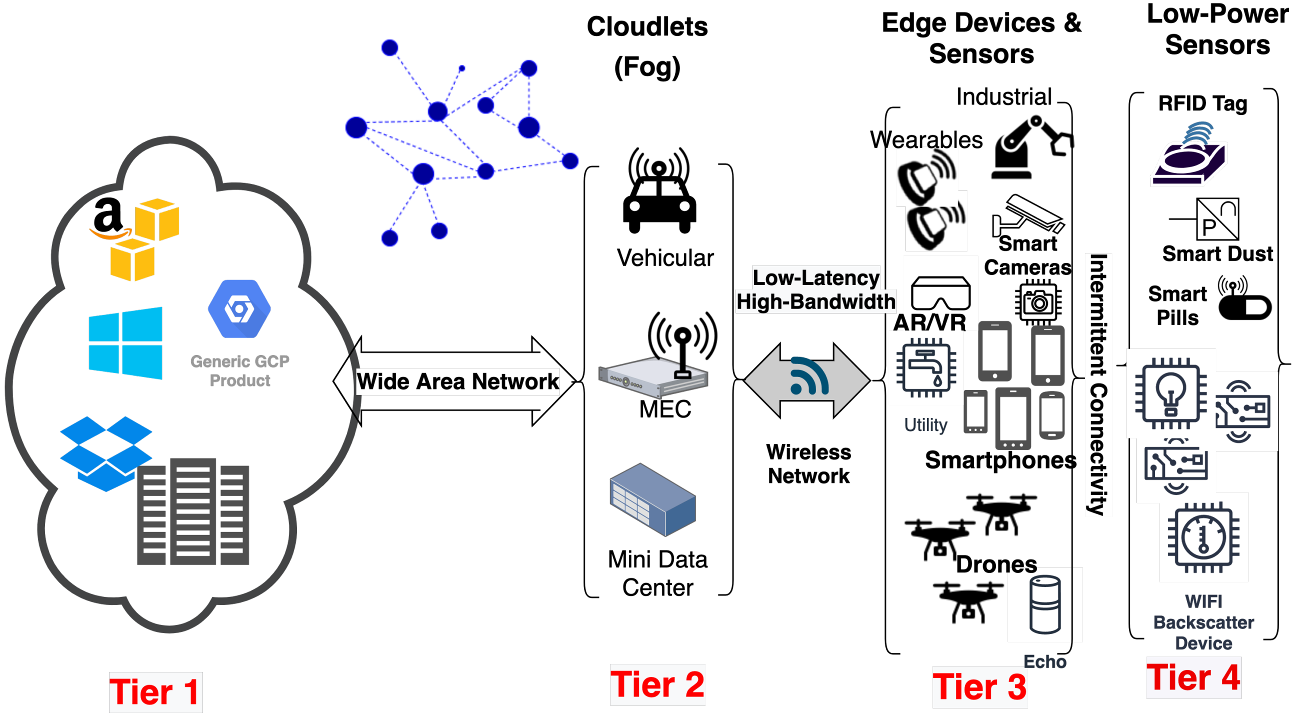
Tier 4 & Tier 3: Immersive Proximity

- Tier-3 devices (e.g., RFID reader) provide the energy that is harvested by a Tier-4 device.
- Immersive proximity defines the relationship between Tier-4 and Tier-3 devices:
  - they have to be physically close enough for the Tier-4 device to harvest sufficient energy for an episode of intermittent computation.

In previous meeting



- Discussed the emerging distributed computing landscape and the technological underpinnings that shape it.
- Explained the concept of cyber physical systems and the Internet of Things
- Explained the 4 Tier model that is used to model the emerging landscape, reason about systems and applications deployed on it, and explored the constraints that define each tier.
- Reviewed and discussed examples of data-driven applications that process data derived from the IOT.
- Explained the role of Tier 2 and 5G.



University of Cyprus
Department of Computer Science

M. D. Dikaiakos

Using the Model

- Not every distributed system will have all tiers.
- Each tier embodies a small set of salient properties that define the reason for the existence of that tier and constrain its range of acceptable designs.
- The same reasoning applies to software at each tier:
  - Tier-3 to Tier-1 communication is (by definition) over a WAN and may involve a wireless first hop that is unreliable and/or congested:
    - embody support for disconnected and weakly-connected operation.
- Tier-3 to Tier-2 communication is expected to be LAN or WLAN quality at all times.
- Tier-1 and Tier-2 server hardware may be identical; only their placement and communication assumptions are different.

Limitations

- The fog computing paradigm exploits processing capabilities of edge devices:
  - These devices apply data reduction techniques, e.g., pre-selection or pre-aggregation, to reduce data volume as early as possible in the processing pipeline, i.e., close to the sensor.
  - However, they only scale within the fog : do not exploit the virtually unlimited cloud resources.
- Wireless sensor networks exploit small battery-powered sensors to create a network of nodes to capture physical phenomena. WSN data management systems:
  - apply acquisitional query processing techniques to optimize the execution for battery lifetimes and deploy a small set of specialized queries to capture the physical phenomena.
  - scale only within the sensor networks and do not exploit the resources of the attached cloud and fog environments:
    - they do not consider offloading computation to external nodes
    - do not provide general-purpose query execution capabilities

Challenges for IoT Platforms

- A data management system for the IoT has to combine the cloud, the fog, and the sensors in a single unified platform to leverage their individual advantages and enable cross-paradigm optimizations (e.g., fusing, splitting, or operator reordering).
- From a system point of view, this unified environment imposes three unique characteristics:
  - **Heterogeneity**
  - **Unreliability**
  - **Elasticity**
- These are not supported by state-of-the-art data management systems.

"The NebulaStream Platform: Data and Application Management for the Internet of Things." *Zeuch, Chaudhary, Del Monte, Gavriilidis, Giouroukis, Grulich, Breß, Traub, Mark. CIDR 2020.*

Heterogeneity

- Processing nodes range
  - from low-end battery-powered sensors over SBCs to
  - high-end rack-scale servers.
- To exploit the individual capacities of each node, an IoT data management system has to take their individual capabilities and resource restrictions into account.
- Current data management systems abstract the underlying hardware with VMs and managed runtimes.
- These abstractions hinder the exploitation of specialized instructions and processing units and prevent important optimizations.

Unreliability

- The fog introduces a highly dynamic runtime environment with unreliable nodes that might change their geo-spatial position, resulting in many transient errors or changes in latency/throughput.
- WSNs exacerbate this highly dynamic runtime even further by turning-off sensors temporally to save energy and allowing reads only following a dedicated read schedule.
- In contrast, a cloud infrastructure is a relatively stable environment where node failures are rare.
- Current approaches for load balancing, fault-tolerance, and correctness only concentrate on one particular environment:
  - miss important cross-paradigm optimization potential

Elasticity

- In a unified environment, data move from the sensors via intermediate nodes to the cloud, and finally to the consumer, e.g., a user device or another system.
- The fog topology is commonly built as a tree-like network topology with several dataflow paths.
  - Data processing in the fog topology has to be network-aware because only nodes on the path from the sensors to the cloud can participate.
- A WSN, all sensors send their data to the next sensor in range until all data end up at the root of the network.
- In contrast, in the cloud, every node has access to all data, e.g., via a distributed file system, e.g., HDFS.
- Current approaches allow optimizations, scaling, and load balancing only within nodes of the same environment and thus miss out important cross-paradigm optimization potential.

Future Evolution

- Possible modalities not covered by the 4 tiers:
  - Biological computer systems
  - Blurring boundaries between tiers, driven by advances in manufacturing technology, leading to a continuum of devices with different power budgets, computing workloads and manufacturing costs.
  - Quantum computing.

Modern Computing Landscape

Energy

Central Role of Energy

- Power concerns at different tiers span many *orders of magnitude*:
  - from a few nanowatts (e.g., a passive RFID tag) to
  - tens of megawatts (e.g., an exascale data center).
- Energy is the most critical factor when making design choices:
  - limited availability of energy could severely limit performance
  - power budget of a system design could be a major barrier to reductions of system cost and form factor.
    - **Form factor:** the physical size and shape of a piece of computer hardware

Central Role of Energy

Energy plays a central role in segmentation across tiers.

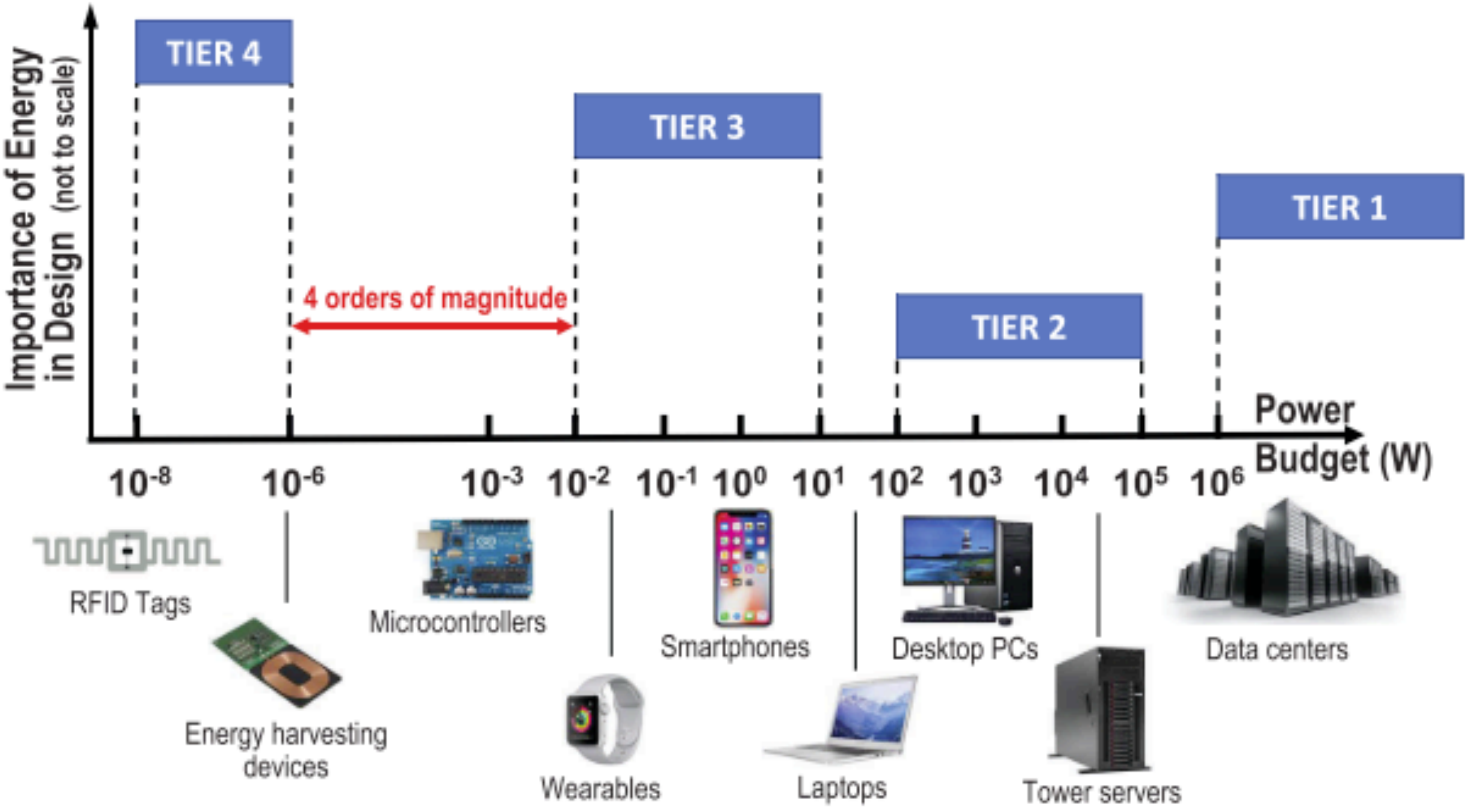


Figure 3: Importance of Energy as a Design Constraint

Energy concerns per Tier

- Tier-1 (Data Centers): Power is used in a data center for IT equipment and infrastructure, adding up to as much as **30 MW at peak hours.**
  - Current power saving techniques focus on load balancing and dynamically eliminating power peaks.
  - Power oversubscription enables more servers to be hosted than theoretically possible: peak demands rarely occur simultaneously.
- Tier-2 (Cloudlets): Cloudlets can span a wide range of form factors, from high-end laptops and desktop PCs to tower or rack servers. Power consumption can therefore vary from **<100W** to **several kilowatts.**
  - Power saving techniques such as CPU frequency scaling are applicable.
  - Techniques to reduce power consumption of attached hardware (GPUs), and to balance the power consumption among multiple interconnected cloudlets.

Energy concerns per Tier

- Tier-3 (Smartphones): dominant type of computing device at Tier-3.
  - Power consumption is >1000mW when idle, but can peak at 3500-4000mW.
  - Techniques such as frequency scaling, display optimization, and application partitioning are used to reduce power consumption.
  - Workload-specific techniques used in web browsing and mobile gaming.
- Tier-3 (Wearables): Energy consumption of smartwatches usually controlled to below 100 mW in stand-by mode with screen off.
  - When the screen is on or the device is wirelessly transmitting data, the energy consumption could surge to 150-200 mW.
  - Various techniques have been proposed to further reduce smartwatch power consumption to <100 mW in active modes via energy-efficient storage or display management.

Energy concerns per Tier

- Tier-4: Energy-harvesting enables infrastructure-free, low-maintenance operation for tiny devices that sense, compute and communicate.
- Energy harvesting presents unique challenges:
  - sporadic power is limited to **10^{-7}** to **10^{-8}** watts using, e.g., RF or biological sources.
  - A passive RFID tag consumes **hundreds of nA** at **1.5V**.
  - Wireless backscatter networking enables communication at extremely low power.
  - Intermittent computing allows sensing and complex processing on scarce energy.
  - Such capabilities enable a new breed of sensors and actuators deployed in the human body to monitor health signals, in civil infrastructure, and in adversarial environments like outer space.

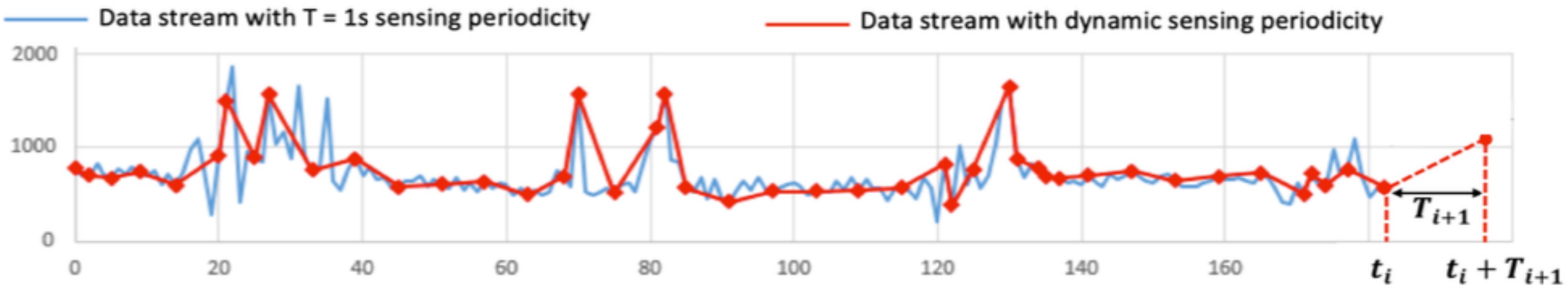
Energy Concerns in the Modern Computing Landscape

Case Study: ADAM, ADMIN

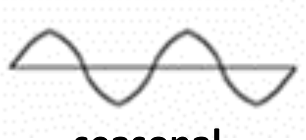
"***Adaptive Monitoring Dissemination for the Internet of Things***", *Demetris Trihinas and George Pallis and Marios Dikaiakos, "IEEE INFOCOM 2017.*

"***AdaM: an Adaptive Monitoring Framework for Sampling and Filtering on IoT Devices***", *D. Trihinas and G. Pallis and M. D. Dikaiakos, "2015 IEEE International Conference on Big Data" (IEEE BigData 2015), Santa Clara, CA, USA Pages: 717–726, 2015*

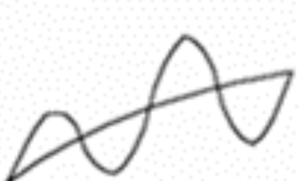
Saving Energy: Adaptive Techniques



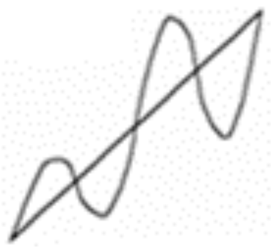
- Edge devices can adapt the monitoring intensity and the amount of data disseminated through the network based on the current evolution and variability of the metric stream
- Adaptation can also create in real-time **estimation models** of the evolution and seasonal behavior of the metric stream and rather than transmitting the entire stream, sending updates for its estimation model from which values can be inferred, triggering dissemination only when shifts in the stream evolution are detected.



seasonal



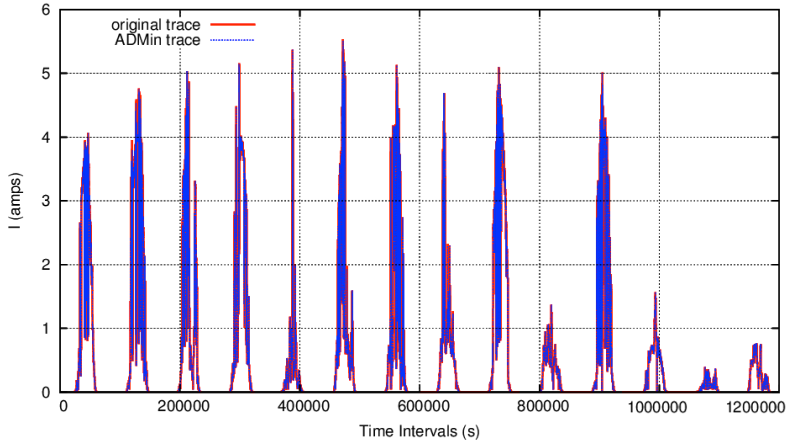
Seasonal with damped trend



Seasonal with exponential trend

*"**Adaptive Monitoring Dissemination for the Internet of Things**", Demetris Trihinas and George Pallis and Marios Dikaiakos, "IEEE INFOCOM 2017.*
*"**AdaM: an Adaptive Monitoring Framework for Sampling and Filtering on IoT Devices**", D. Trihinas and G. Pallis and M. D. Dikaiakos, "2015 IEEE International Conference on Big Data" (IEEE BigData 2015), Santa Clara, CA, USA Pages: 717–726, 2015*

Data Reduction vs Accuracy



Photovoltaic Panel Current (I_{DC}) Production

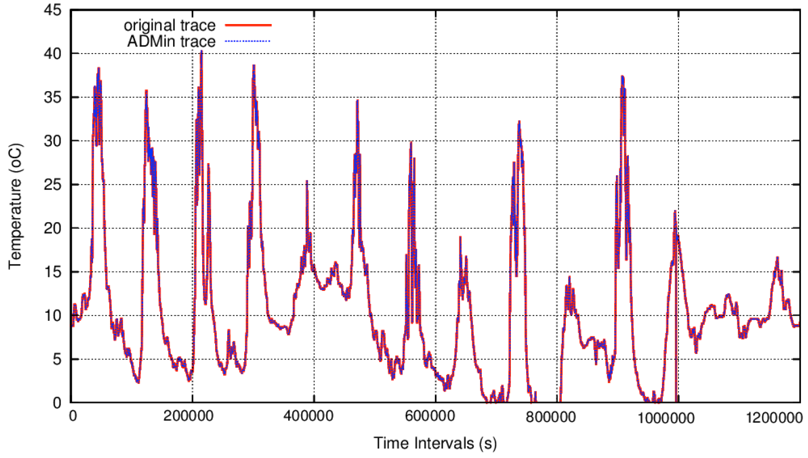
2 Weeks of data collected every 1 second

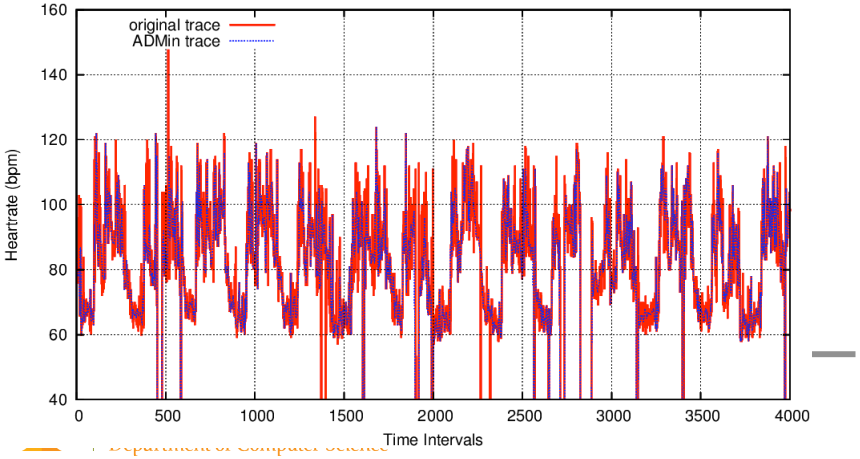
Data reduction: 87% -- Accuracy: 93%

Weather Station Air Temperature (oC)

2 Weeks of data collected every 1 second

Data reduction: 85% -- Accuracy: 92%



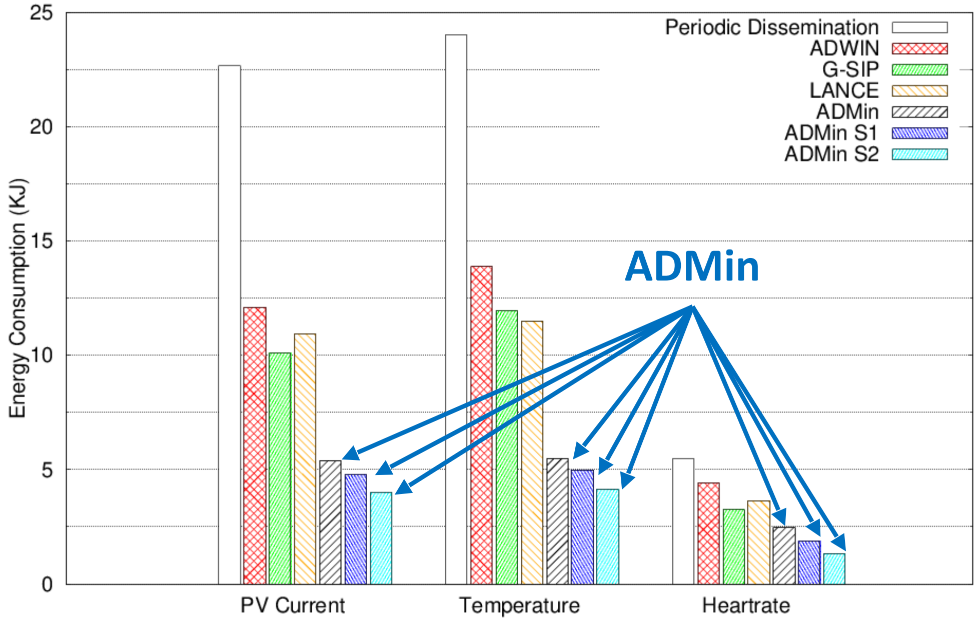


Wearable Human Heartrate (bpm)

1month of data collected every 1 minute

Data reduction: 80% -- Accuracy: 90%

Energy Savings



- ADMin reduces energy consumption by at least 76% and when incorporating seasonality knowledge by at least 83%

"***Adaptive Monitoring Dissemination for the Internet of Things***", *Demetris Trihinas and George Pallis and Marios Dikaiakos, "IEEE INFOCOM 2017.*

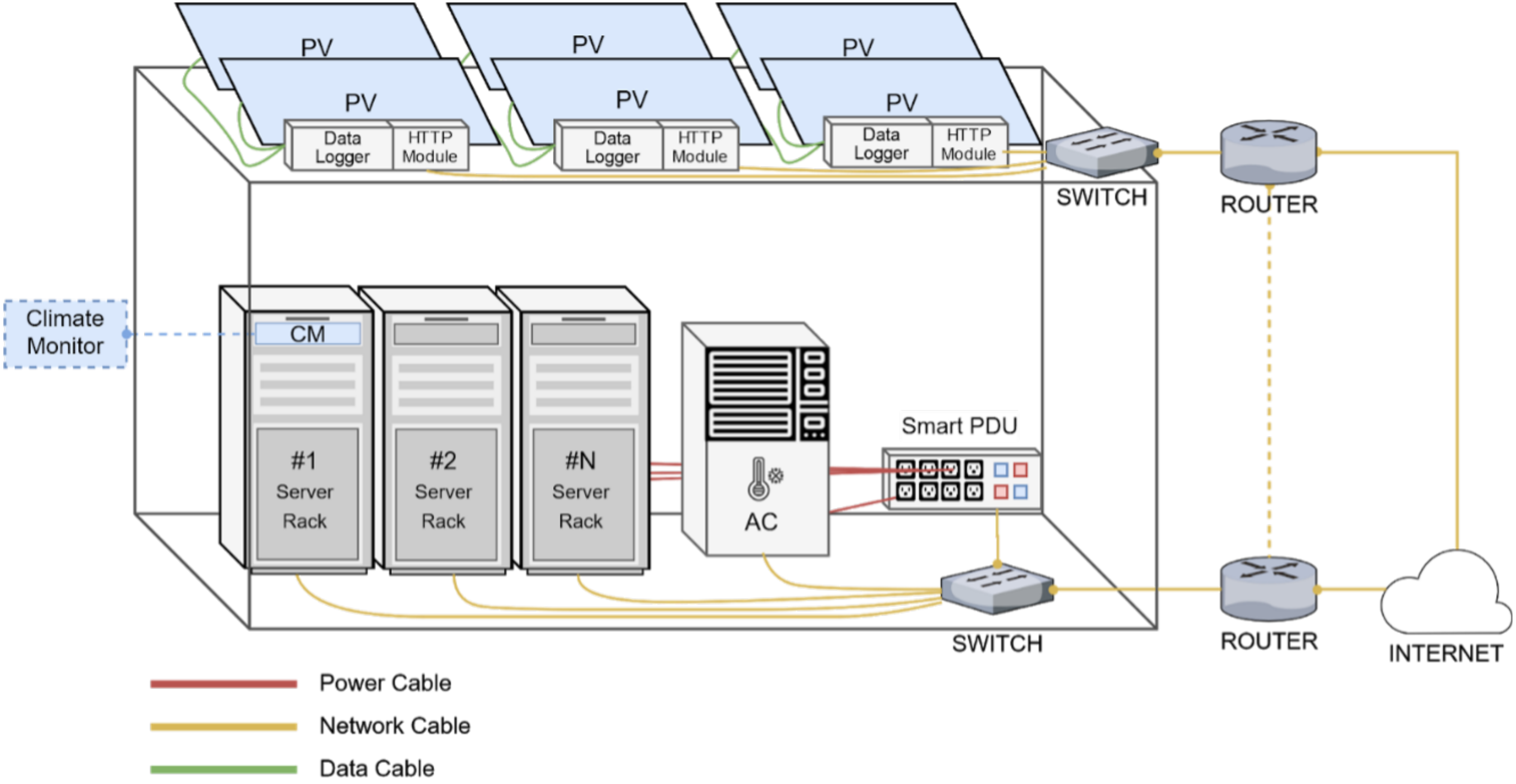
Energy Concerns in the Modern Computing Landscape

Case Study: ENEDI Platform

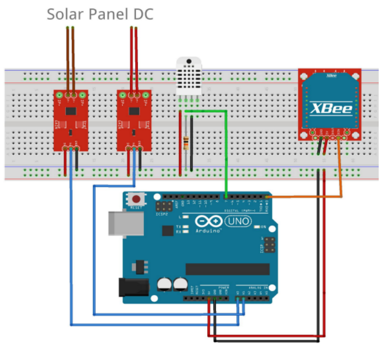
"ENEDI: Energy Saving in Datacenters"*, A. Tryfonos, A. Andreou, N. Loulloudes, G. Pallis, M. D. Dikaiakos, N. Chatzigeorgiou , G. E. Georghiou, "Global Conference on Internet of Things" (2018 IEEE GCIoT), Alexandria, Egypt 2018*

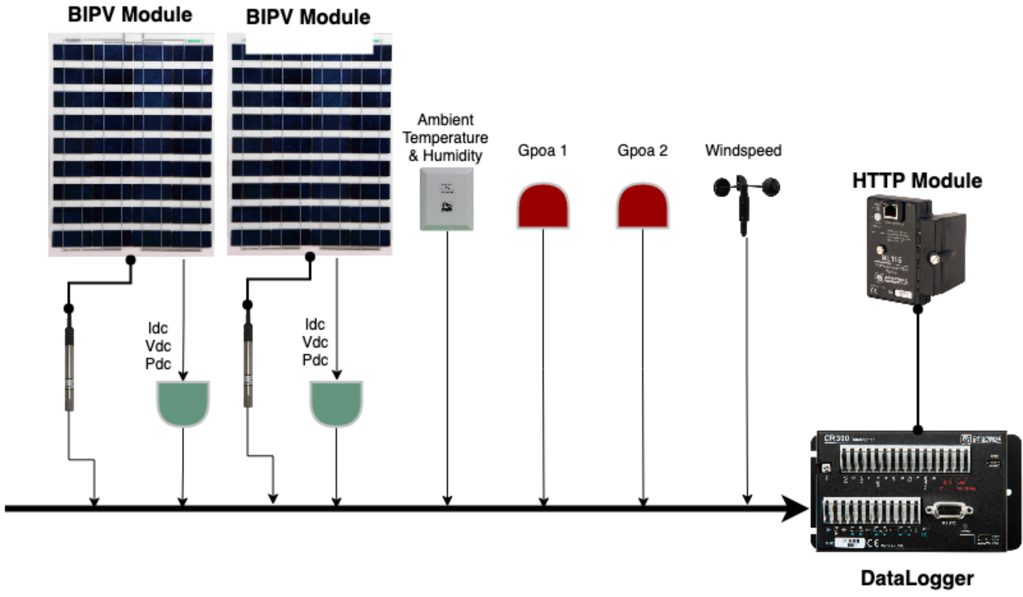
*"****A Cyber-Physical System for Solar-Powered Data Centers*** *", M. D. Dikaiakos, N. Chatzigeorgiou, A. Tryfonos, A. Andreou, N. Loulloudes, G. Pallis, G. E. Georghiou (submitted for publication)*

High-Level Architecture



IoT Sensors





Legend

DC Power, DC Current, DC VOltage Sensor

PV Temperature Sensor

Windspeed Sensor

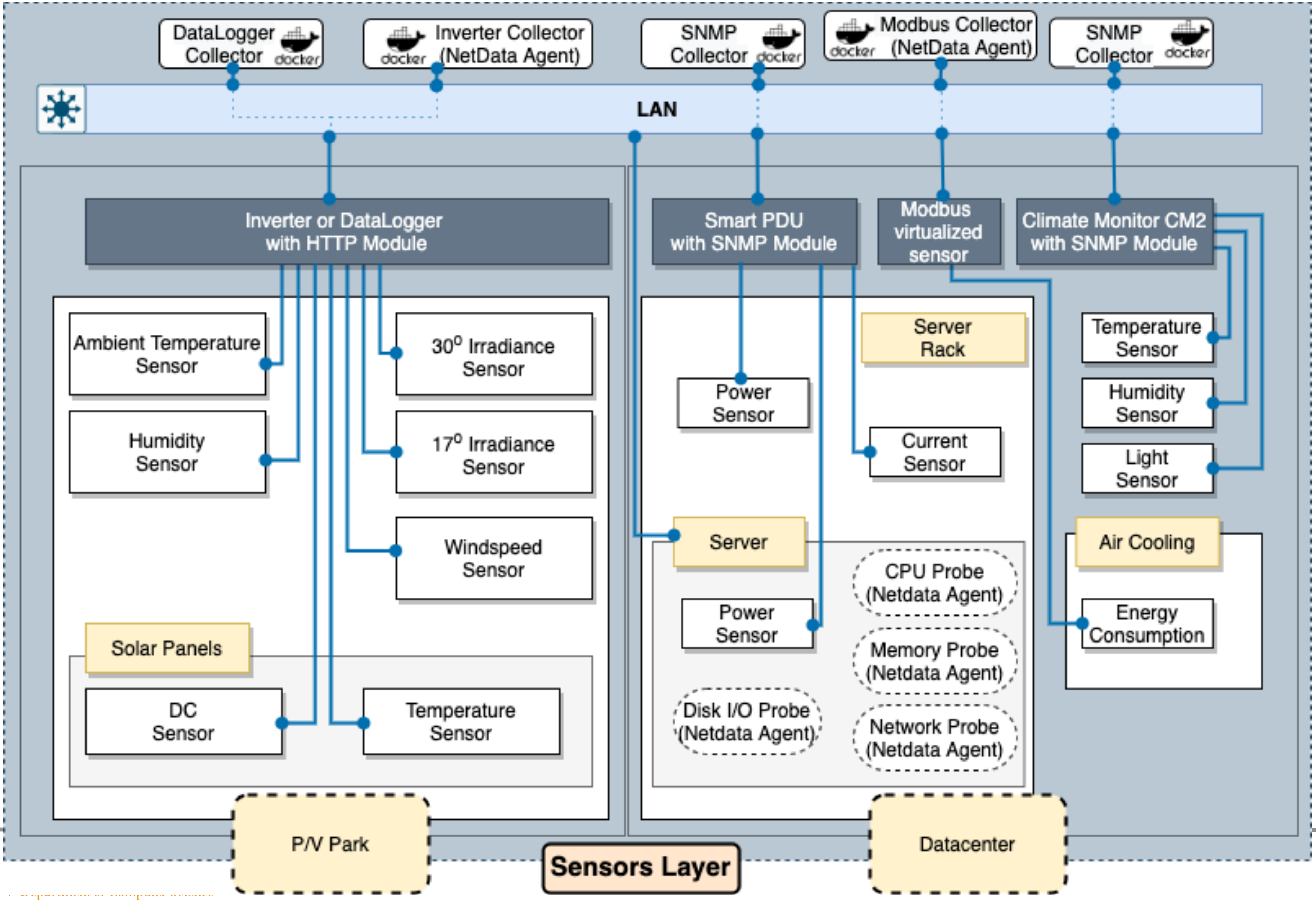
Global Pane of Array Irradiance

Campbell Scientific CR3000 or CR1000 DataLogger

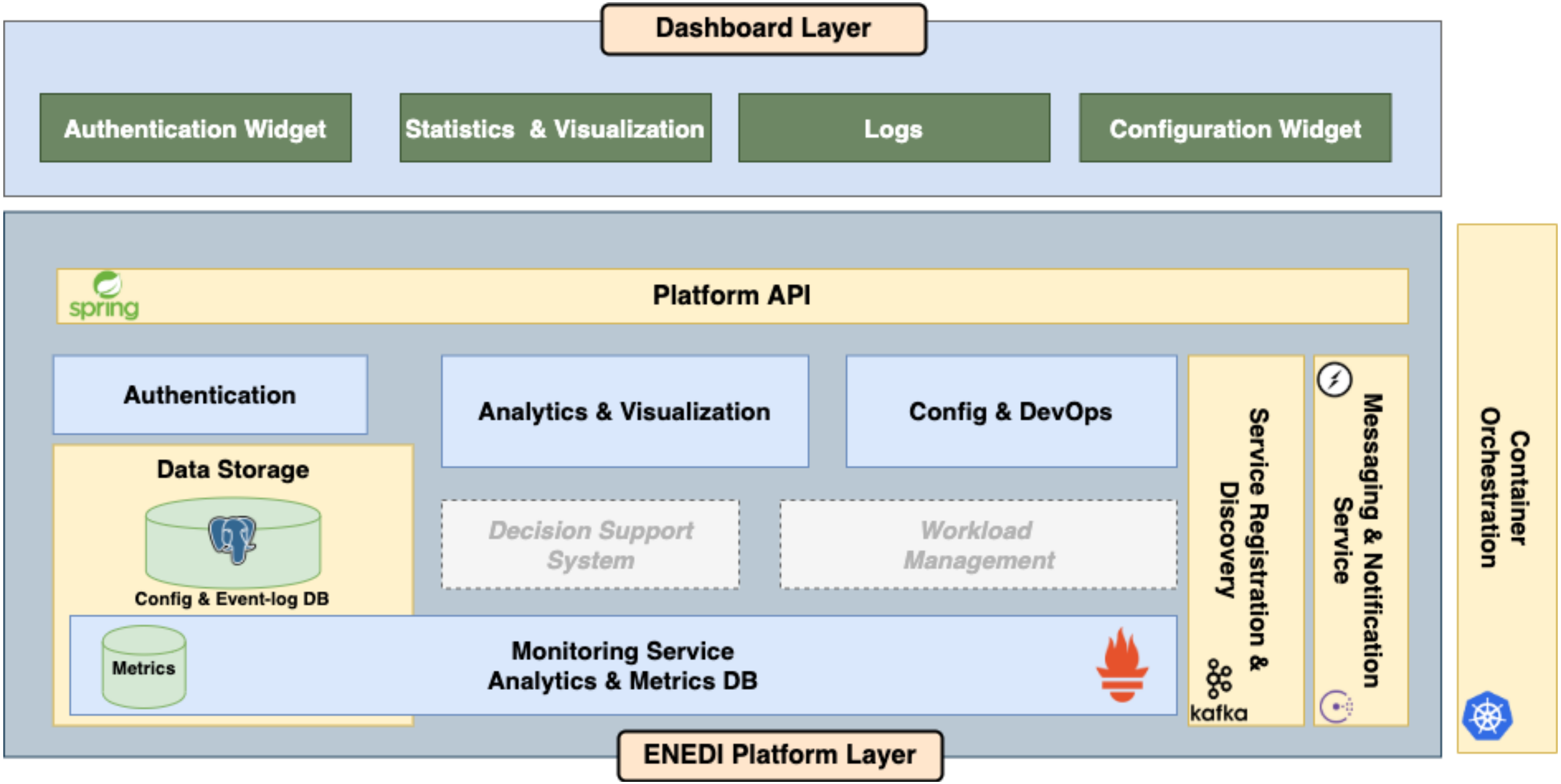
NL115 HTTP Module for Campbell Scientific DataLogger

Ambient Temperature and Humidity Sensor

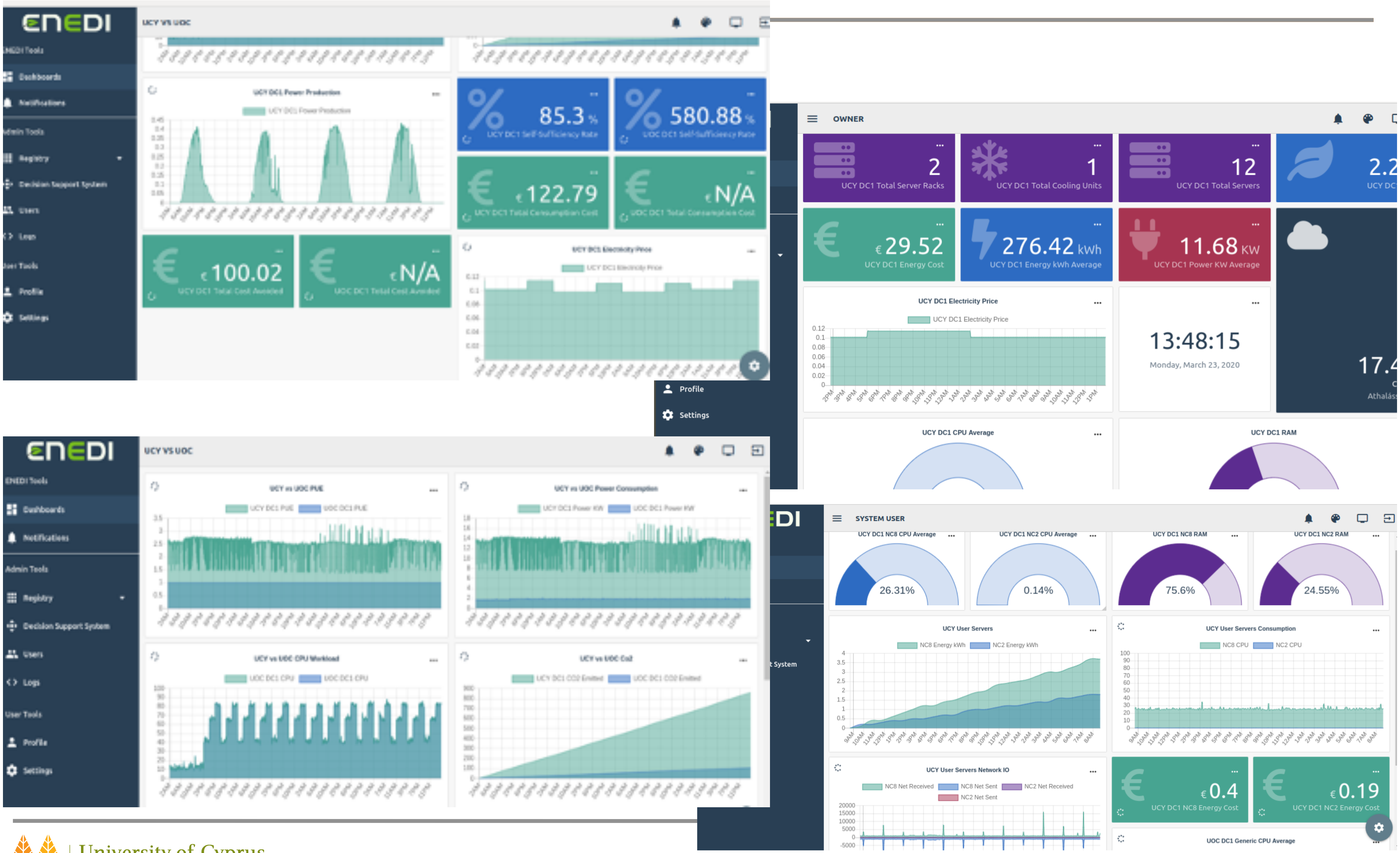
Sensor Integration



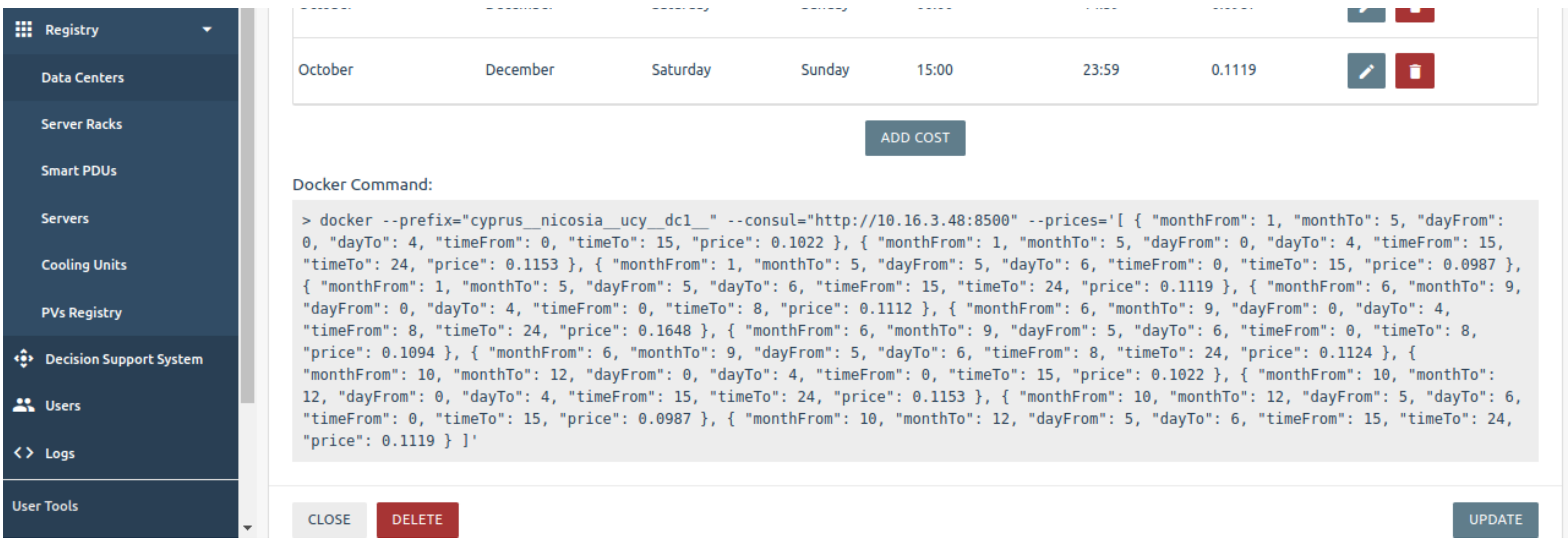
Middleware


Dashboard Layer
Authentication Widget
Statistics & Visualization
Logs
Configuration Widget
spring
Platform API
Authentication
Analytics & Visualization
Config & DevOps
Service Registration & Discovery
kafka
Messaging & Notification Service
Container Orchestration
Data Storage
Config & Event-log DB
Decision Support System
Workload Management
Metrics
Monitoring Service
Analytics & Metrics DB
ENEDI Platform Layer

Dashboard


ENEDI
ENEDI Tools
Dashboards
Notifications
Admin Tools
Registry
Decision Support System
Users
Logs
User Tools
Profile
Settings
UCY VS UOC
UCY DC1 Power Production
85.3 %
UCY DC1 Self-Sufficiency Rate
580.88 %
UOC DC1 Self-Sufficiency Rate
€ 122.79
UCY DC1 Total Consumption Cost
€ N/A
UOC DC1 Total Consumption Cost
€ 100.02
UCY DC1 Total Cost Avoided
€ N/A
UOC DC1 Total Cost Avoided
UCY DC1 Electricity Price
UCY vs UOC PUE
UCY DC1 PUE
UOC DC1 PUE
UCY vs UOC Power Consumption
UCY DC1 Power KW
UOC DC1 Power KW
UCY vs UOC CPU Workload
UCY DC1 CPU
UOC DC1 CPU
UCY vs UOC Co2
UCY DC1 CO2 Emitted
UOC DC1 CO2 Emitted
OWNER
2
UCY DC1 Total Server Racks
1
UCY DC1 Total Cooling Units
12
UCY DC1 Total Servers
2.2
€ 29.52
UCY DC1 Energy Cost
276.42 kWh
UCY DC1 Energy kWh Average
11.68 KW
UCY DC1 Power KW Average
13:48:15
Monday, March 23, 2020
17.4
Athalass
UCY DC1 CPU Average
UCY DC1 RAM
SYSTEM USER
UCY DC1 NC8 CPU Average
26.31%
UCY DC1 NC2 CPU Average
0.14%
UCY DC1 NC8 RAM
75.6%
UCY DC1 NC2 RAM
24.55%
UCY User Servers
NC8 Energy kWh
NC2 Energy kWh
UCY User Servers Consumption
NC8 CPU
NC2 CPU
UCY User Servers Network IO
NC8 Net Received
NC8 Net Sent
NC2 Net Received
NC2 Net Sent
€ 0.4
UCY DC1 NC8 Energy Cost
€ 0.19
UCY DC1 NC2 Energy Cost
UOC DC1 Generic CPU Average

Configuration & Management

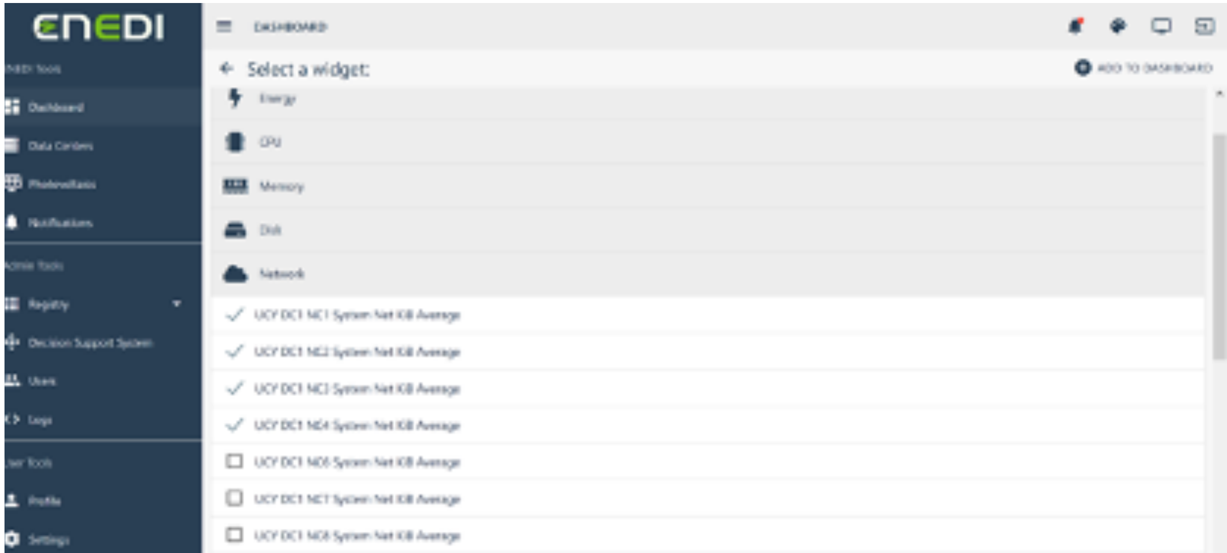


October	December	Saturday	Sunday	15:00	23:59	0.1119

ADD COST

Docker Command:

```
> docker --prefix="cyprus__nicosia__ucy__dc1__" --consul="http://10.16.3.48:8500" --prices='[ { "monthFrom": 1, "monthTo": 5, "dayFrom":
0, "dayTo": 4, "timeFrom": 0, "timeTo": 15, "price": 0.1022 }, { "monthFrom": 1, "monthTo": 5, "dayFrom": 0, "dayTo": 4, "timeFrom": 15,
"timeTo": 24, "price": 0.1153 }, { "monthFrom": 1, "monthTo": 5, "dayFrom": 5, "dayTo": 6, "timeFrom": 0, "timeTo": 15, "price": 0.0987 },
{ "monthFrom": 1, "monthTo": 5, "dayFrom": 5, "dayTo": 6, "timeFrom": 15, "timeTo": 24, "price": 0.1119 }, { "monthFrom": 6, "monthTo": 9,
"dayFrom": 0, "dayTo": 4, "timeFrom": 0, "timeTo": 8, "price": 0.1112 }, { "monthFrom": 6, "monthTo": 9, "dayFrom": 0, "dayTo": 4,
"timeFrom": 8, "timeTo": 24, "price": 0.1648 }, { "monthFrom": 6, "monthTo": 9, "dayFrom": 5, "dayTo": 6, "timeFrom": 0, "timeTo": 8,
"price": 0.1094 }, { "monthFrom": 6, "monthTo": 9, "dayFrom": 5, "dayTo": 6, "timeFrom": 8, "timeTo": 24, "price": 0.1124 }, {
"monthFrom": 10, "monthTo": 12, "dayFrom": 0, "dayTo": 4, "timeFrom": 0, "timeTo": 15, "price": 0.1022 }, { "monthFrom": 10, "monthTo":
12, "dayFrom": 0, "dayTo": 4, "timeFrom": 15, "timeTo": 24, "price": 0.1153 }, { "monthFrom": 10, "monthTo": 12, "dayFrom": 5, "dayTo": 6,
"timeFrom": 0, "timeTo": 15, "price": 0.0987 }, { "monthFrom": 10, "monthTo": 12, "dayFrom": 5, "dayTo": 6, "timeFrom": 15, "timeTo": 24,
"price": 0.1119 } ]'
```



Select a widget:

- Energy
- CPU
- Memory
- Disk
- Network

✓ UCY DC1 NC1 System Net KB Average
✓ UCY DC1 NC2 System Net KB Average
✓ UCY DC1 NC3 System Net KB Average
✓ UCY DC1 NC4 System Net KB Average
□ UCY DC1 NC6 System Net KB Average
□ UCY DC1 NC7 System Net KB Average
□ UCY DC1 NC8 System Net KB Average