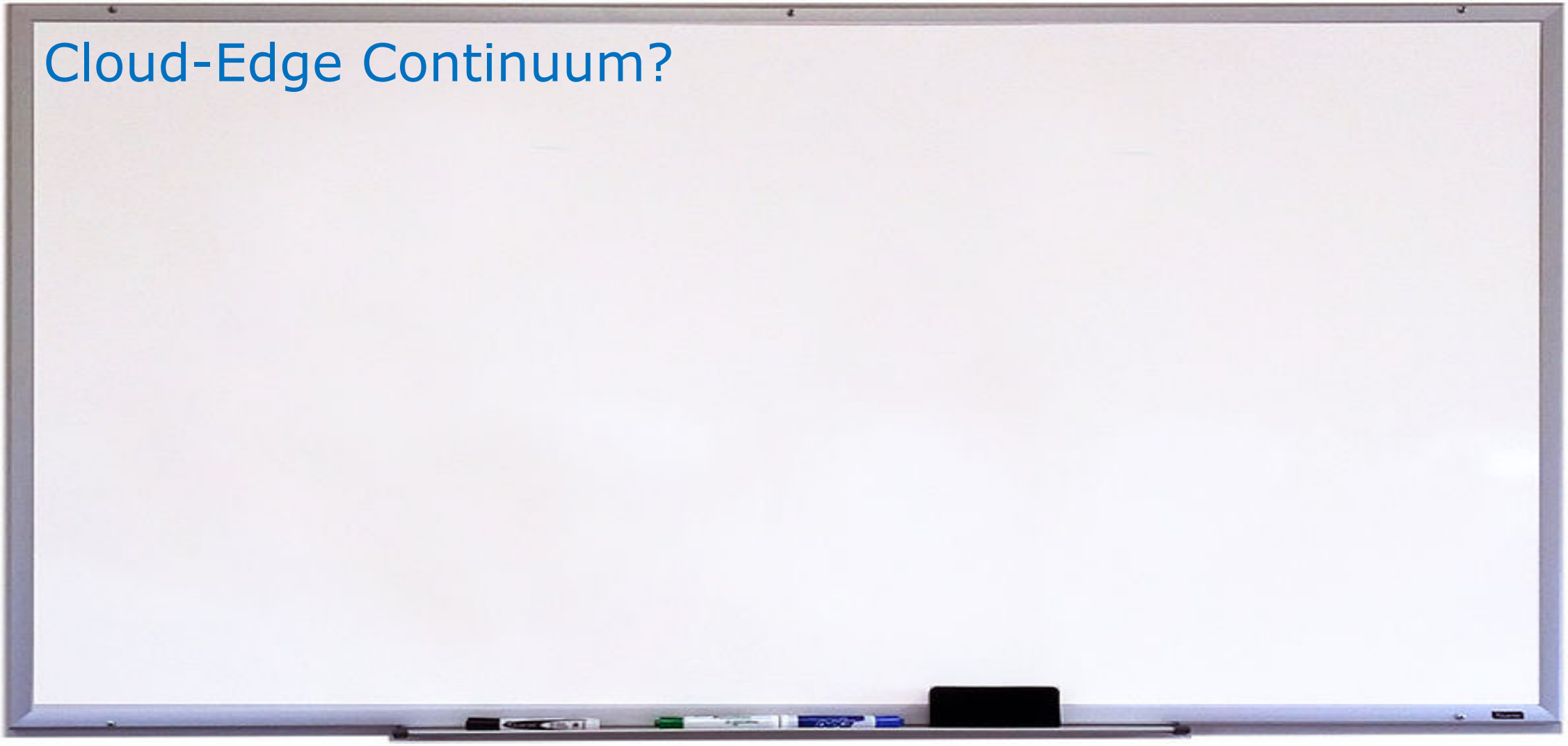
Cloud-Edge Continuum

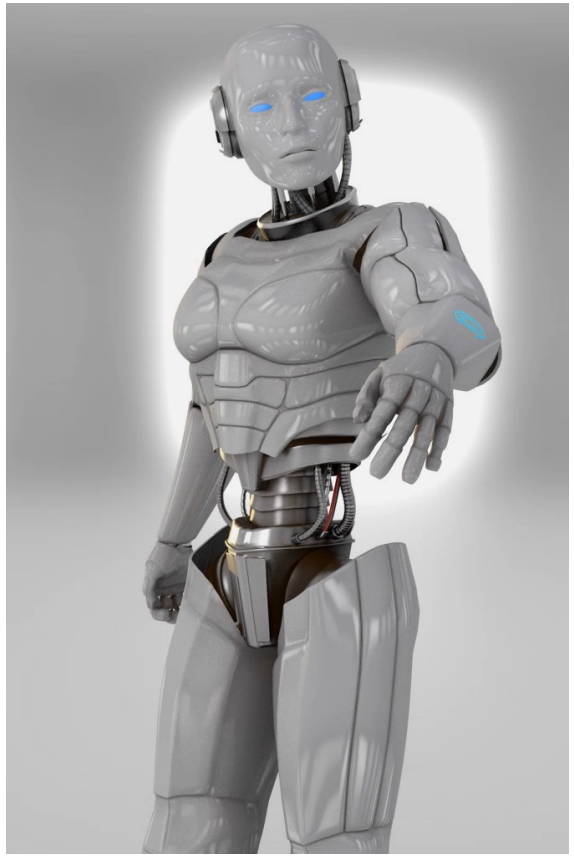
Antonio Brogi

Department of Computer Science
University of Pisa

Cloud-Edge Continuum?

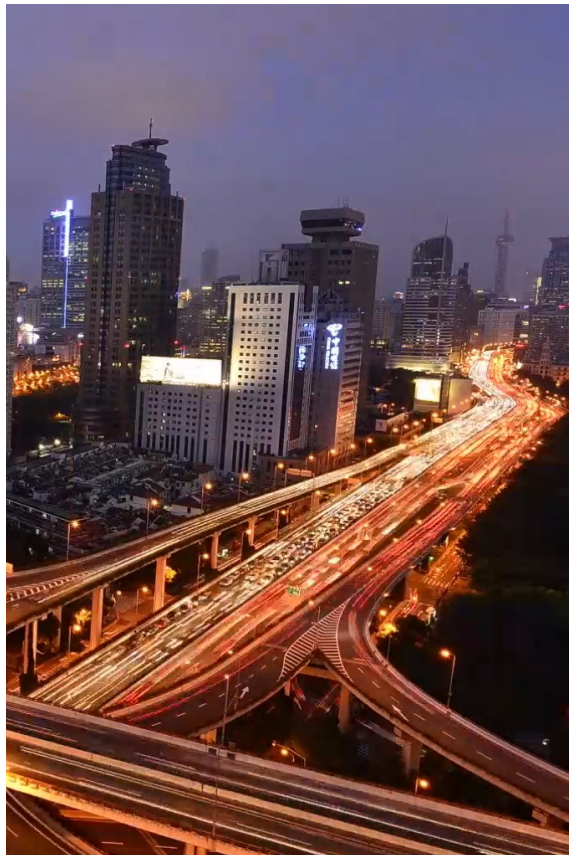


Pervasive IoT applications



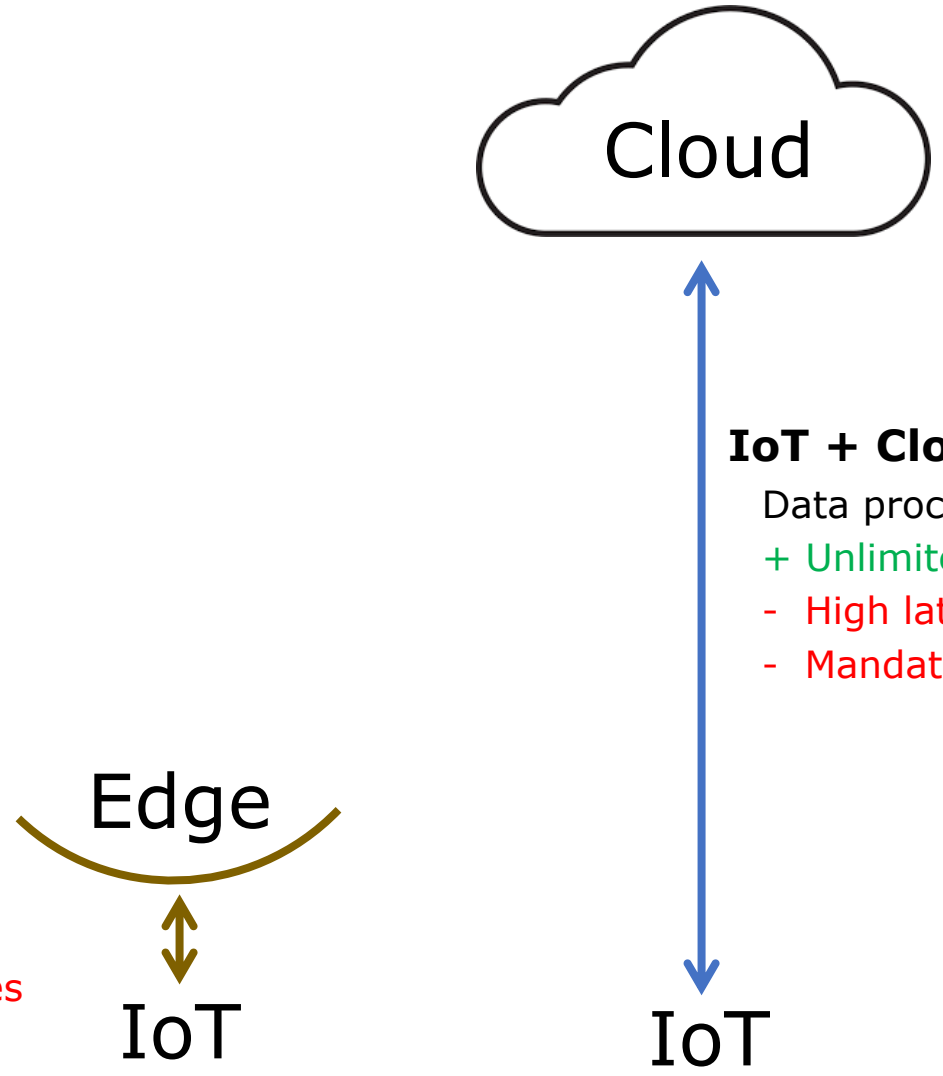
Embedded AI

Energy production plants



Smart Cities

Traditional deployment models



IoT + Edge

Data processed at the Edge

\+ Low latencies

\- Limited processing and storage capabilities

IoT + Cloud

Data processed in the Cloud

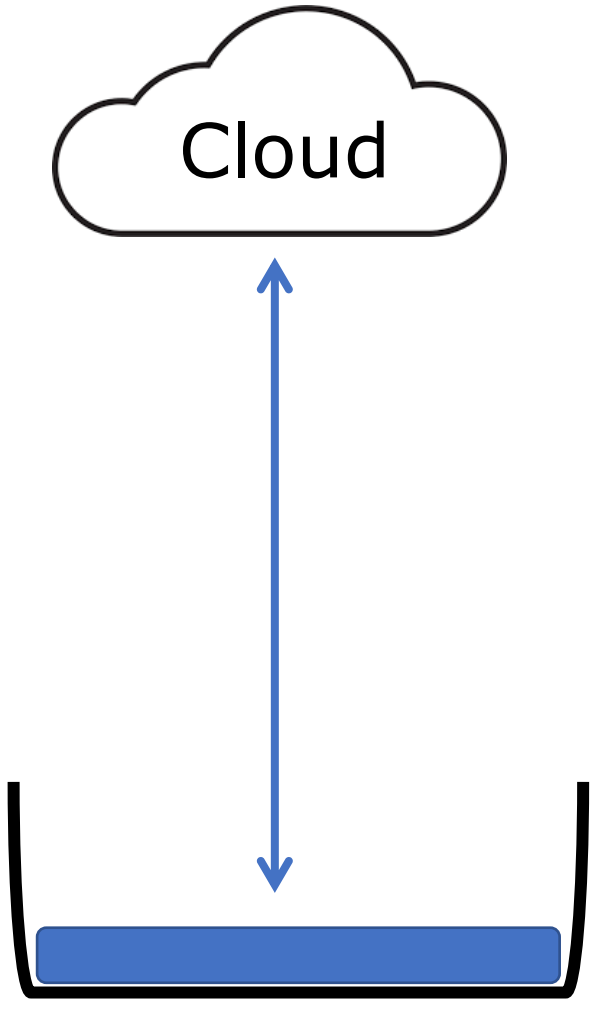
\+ Unlimited processing and storage capabilties

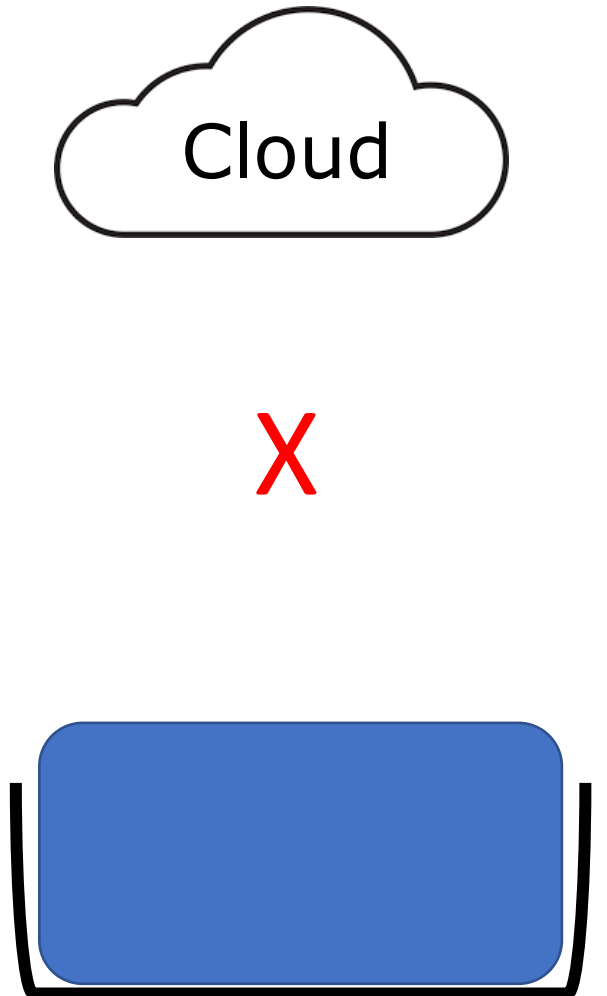
\- High latencies

\- Mandatory Internet connectivity

Mandatory Internet connectivity

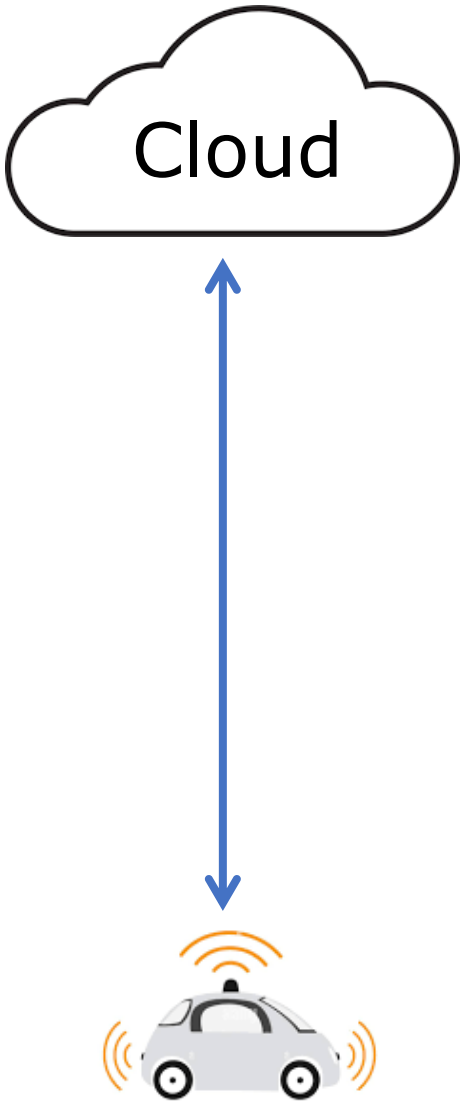
e.g. water flooding management must work in critical situations



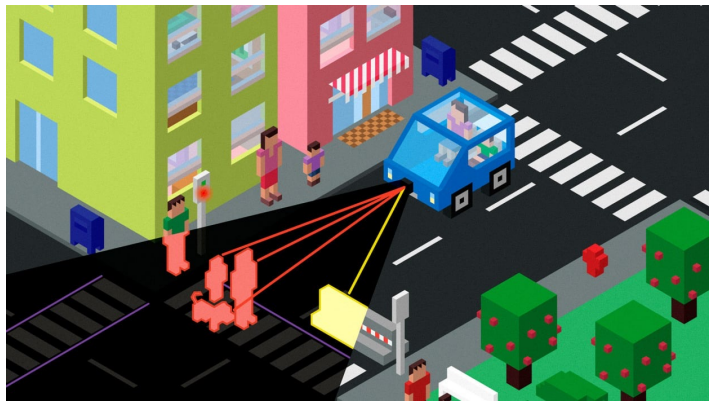


High latencies

e.g. self-driving cars need to stop promptly

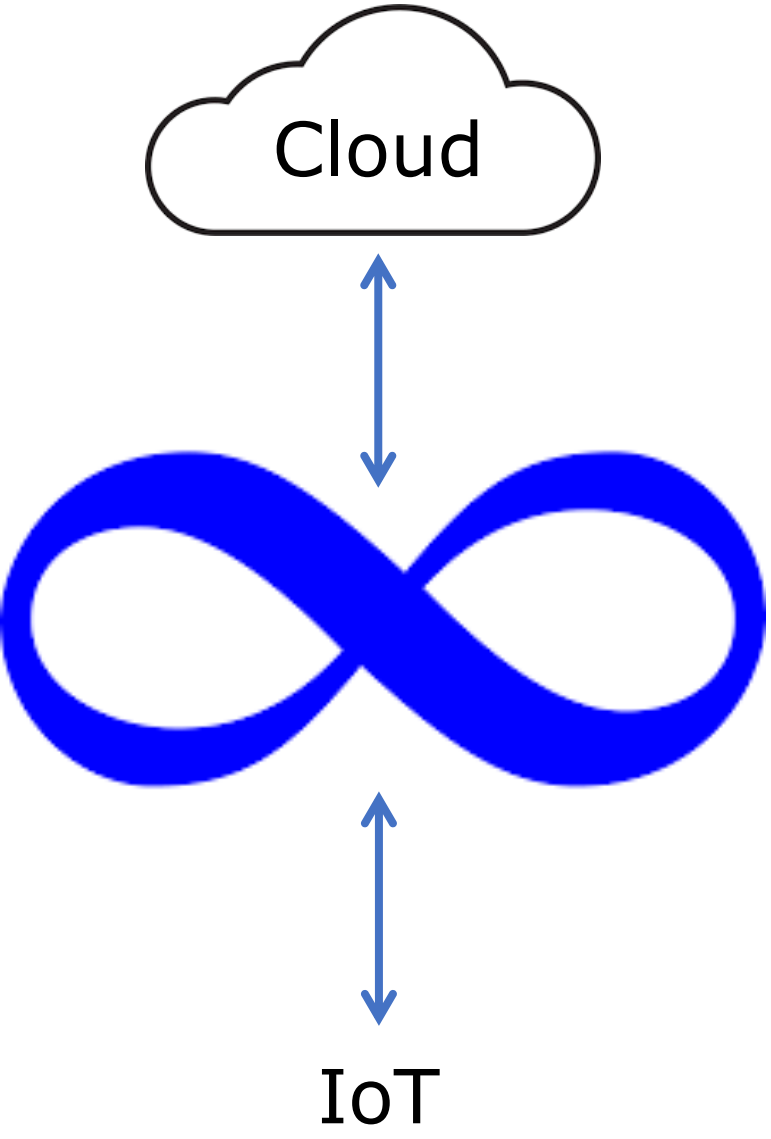


[ethics issues



too]

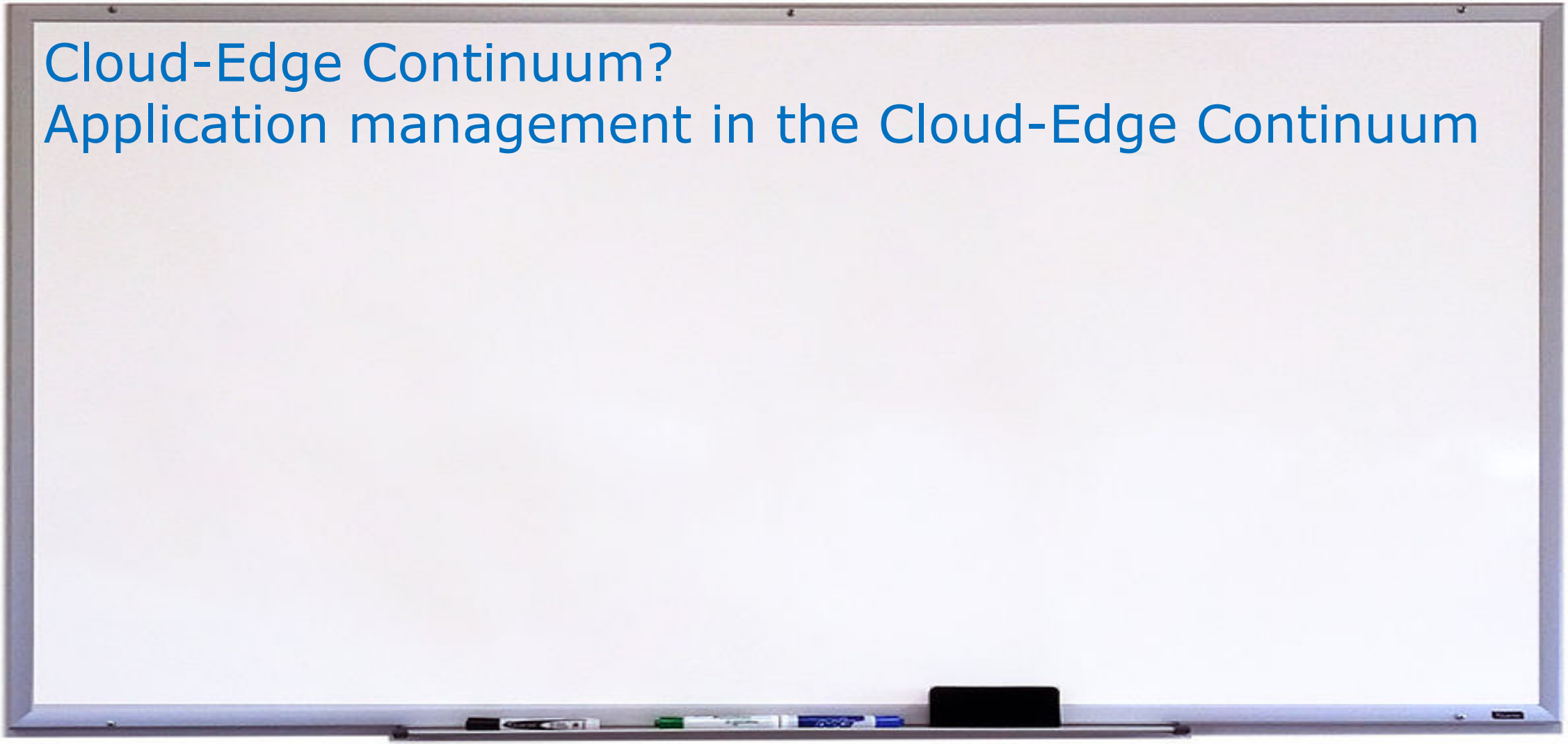
Cloud-Edge Continuum



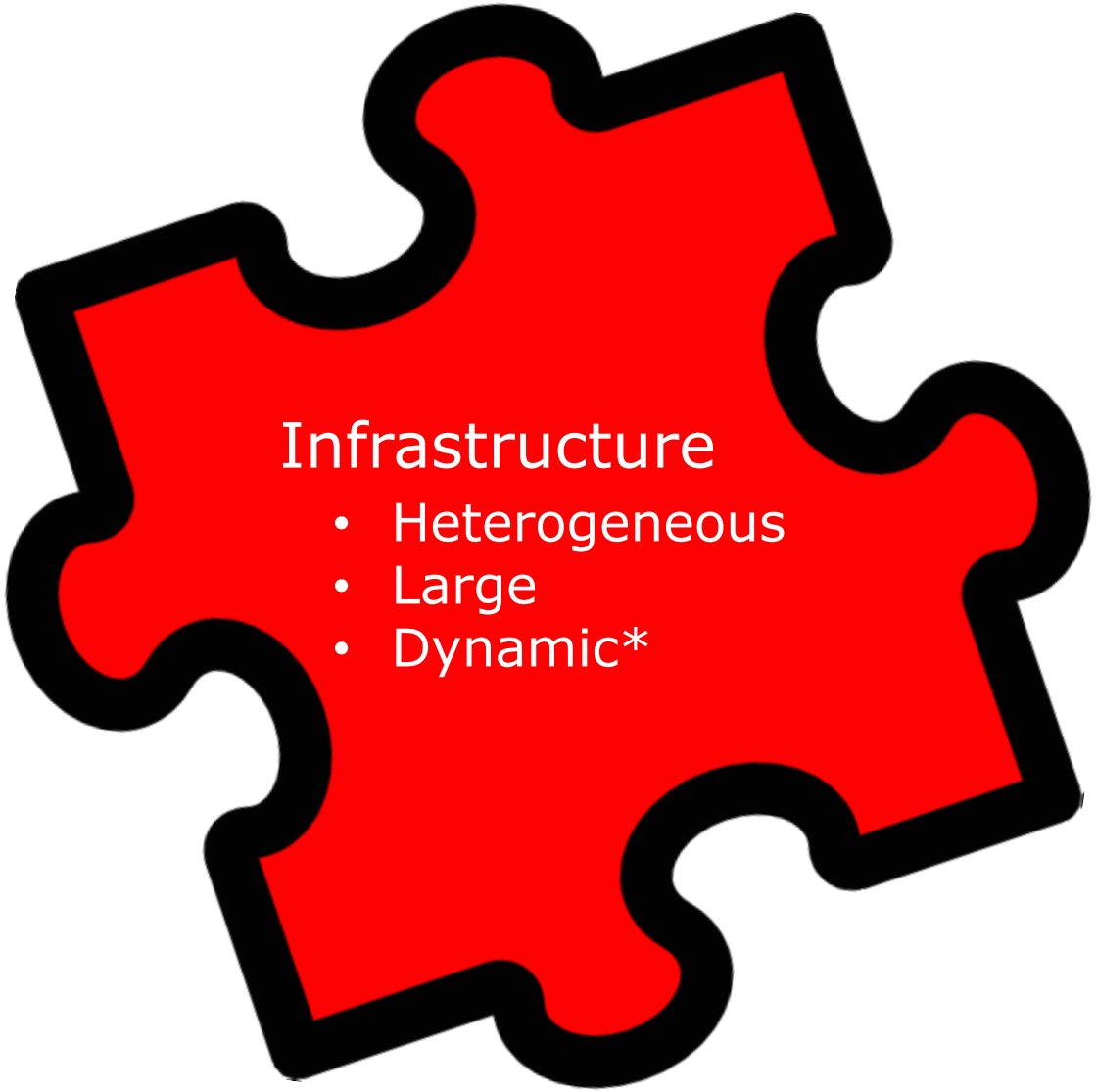
Extending the Cloud towards the IoT
with a **distributed, heterogenous infrastructure**
to better support **latency-sensitive** applications

https://www.youtube.com/watch?v=ICQ0AAYO0mQ

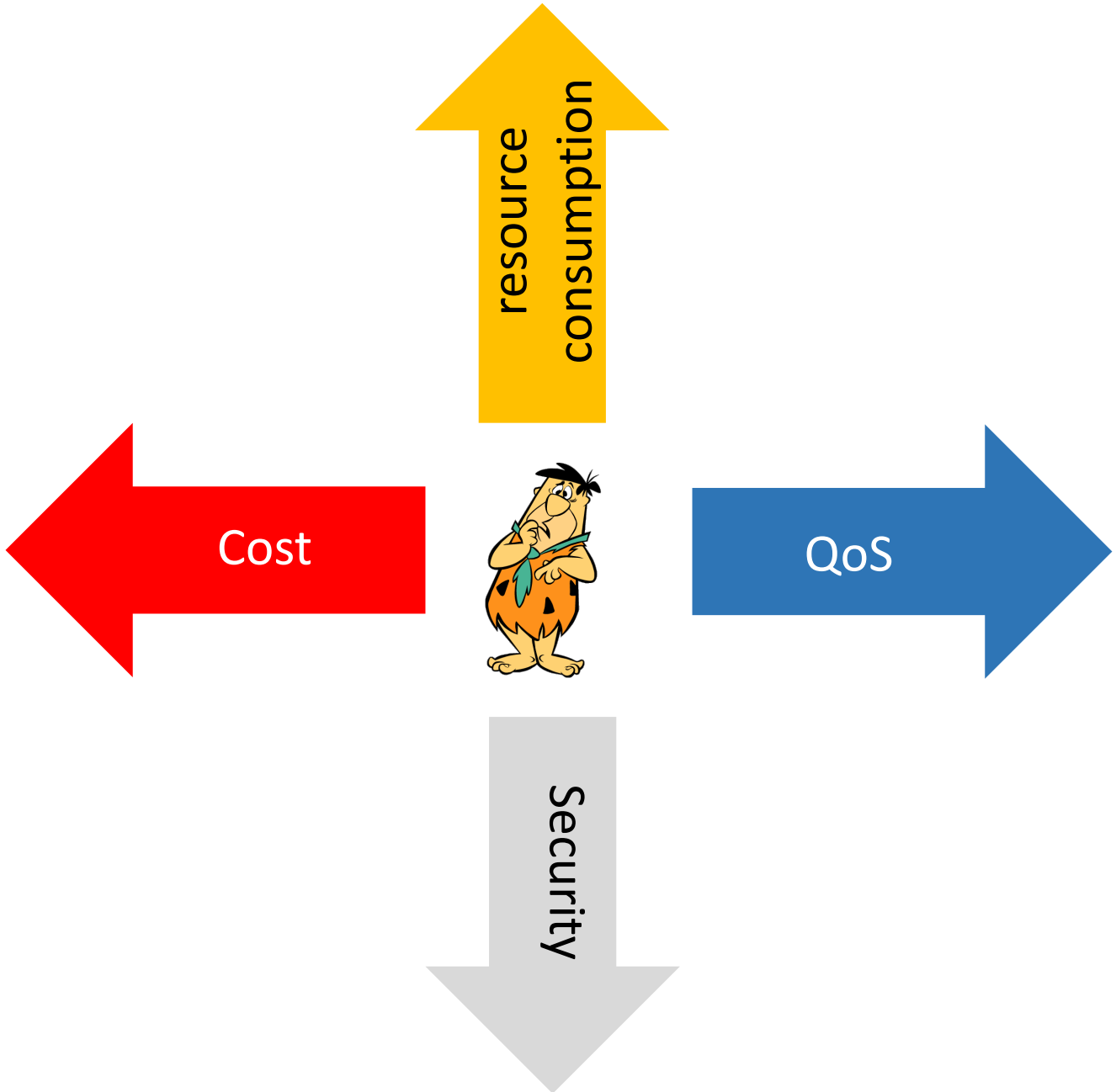
Cloud-Edge Continuum?
Application management in the Cloud-Edge Continuum



Deploying composite applications in a QoS- and context-aware manner on the Cloud-Edge Continuum is challenging …



Need **tools** helping to master orthogonal dimensions



Problem #1: How to suitably **place** a composite application on the Cloud-Edge Continuum

Different approaches

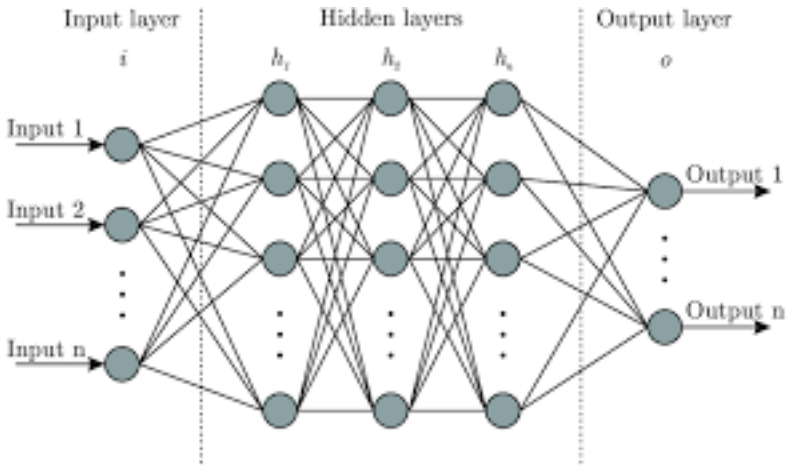
MILP

- X_{ij}^{k} = BINARY
- $\sum_{j=1}^{n} X_{0j}^{k} = \sum_{i=1}^{n} X_{i0}^{k} = 1$; k=1,2…m : Every route should start from depot and end on depot only,
- $\sum_{i=1}^{n} X_{ih}^{k} = \sum_{j=1}^{n} X_{hj}^{k} <= 1$, h=1,2,…n; k=1,2…m :every node should be selected at most once and every node served should have a in as well as out arc,
- $\sum_{k=1}^{m}\sum_{j=1}^{n} X_{hj}^{k} = \sum_{k=1}^{m}\sum_{i=1}^{n} X_{ih}^{k} = 1$; h=1,2…n : Every node should be selected at least once,
- $\sum_{j=1}^{n}\sum_{i=1}^{n} D_i X_{ij}^{k} <= Q$; k=1,2… m :Carrier can't carry more than q quantity,
- $\sum_{k=1}^{m}\sum_{j=1}^{n}\sum_{i=1}^{n} D_i X_{ij}^{k} = \sum_{i=1}^{n} D_i$; Total supply to nodes should equal to total demand,
- $X_{hj}^{k} + X_{ih}^{k} <= 1$; for i,j=1,2…n; h=1,2…n for every k=1,2…m: Every node visited should have an arc to other then its preceding node.

Hard to read

Slow to run

ML



Infrastructure is very dynamic

Explainability

Declarative

"service S can be placed on node N if …"

Declarative approach

1) Declare what an eligibile placement is

service S can be placed on node N ***if***

the hardware reqs of S are met by N ***and***

the IoT connection reqs of S are met by N ***and***

the software reqs of S are met by N

2) Let the inference engine look for it!

services S1 … Sm can be placed on nodes N1 … Nm ***if***

service S1 can be placed on node N1 ***and***

… ***and***

service Sm can be placed on node Nm … ***and***

the QoS reqs of S1 … Sm are met

…

Problem #2: How to suitably **manage** application deployments in the Cloud-Edge Continuum (after first deployment)

Continuous Reasoning

Exploit compositionality to differentially analyse a large-scale system:

- by mainly **focussing on the latest changes** introduced in the system, and
- by **re-using previously computed results** as much as possible

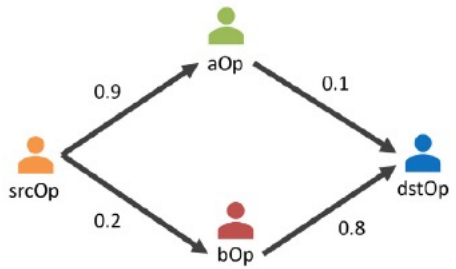
Only 43 SLoC!

(10 predicates!)

https://www.youtube.com/watch?v=F7oLVrNWADA

Other aspects

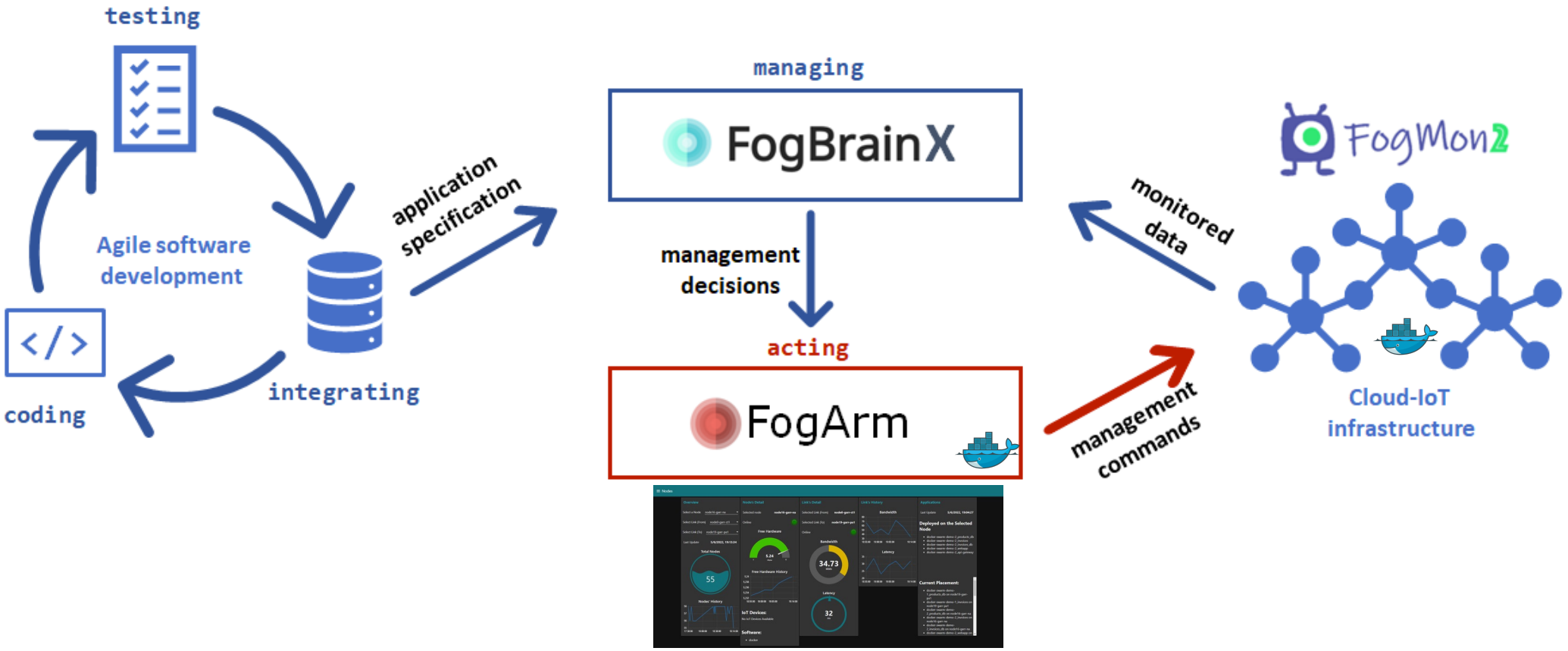
+ probabilities to model **infrastructure dynamicity**

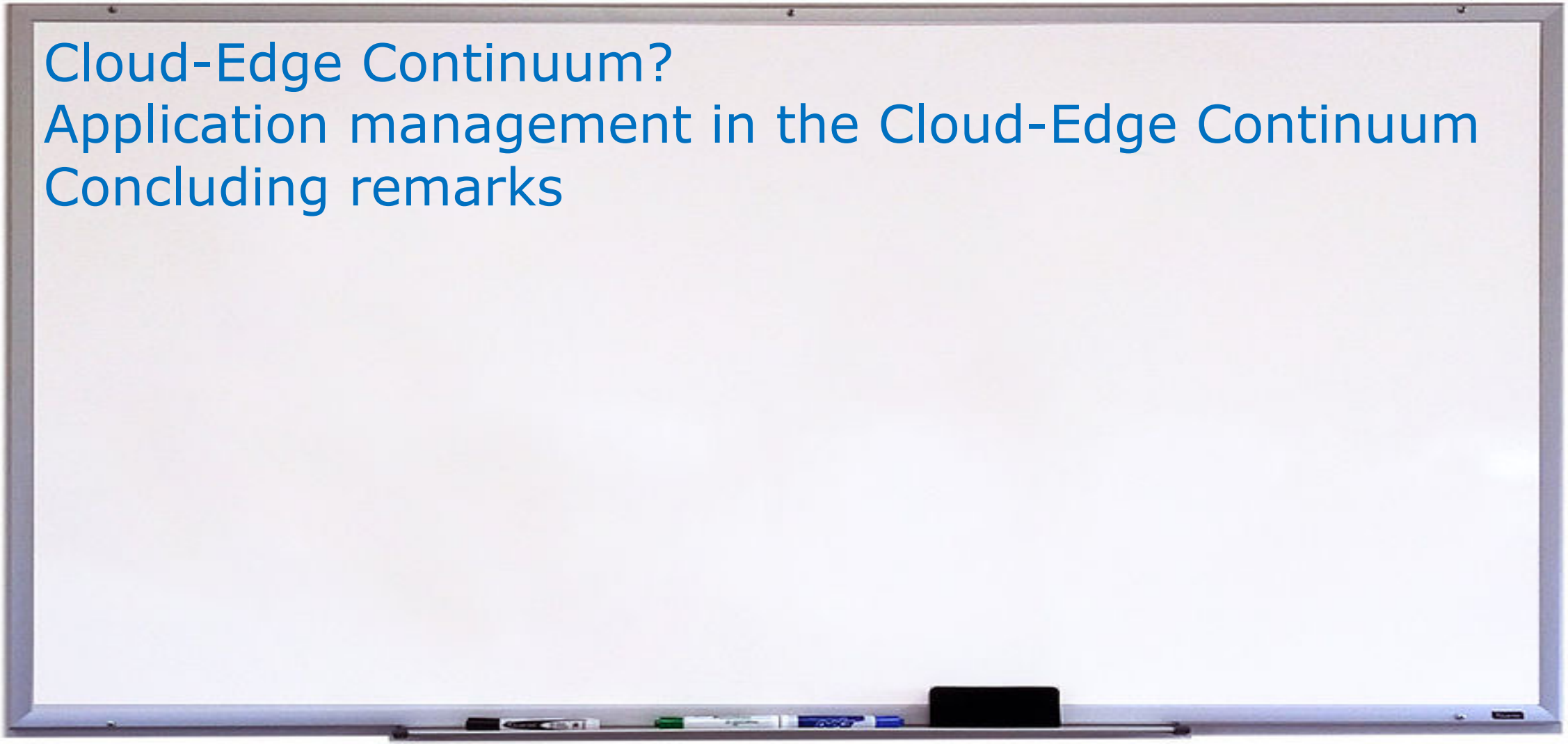


+ semirings to model (non-monotonic, conditionally transitive) **trust** relations among different stakeholders

Other aspects (cont.)

+ **infrastructure monitoring** and **management enactment**




Cloud-Edge Continuum?
Application management in the Cloud-Edge Continuum
Concluding remarks

Industrial & academic interest on the Cloud-Edge Continuum continues to grow

Many challenges:

- adaptive application deployment
- (distributed) application management
- privacy/security/trust
- fault resilience
- testbeds
- Continuum for AI and viceversa
- sustainability
- ...

- business models