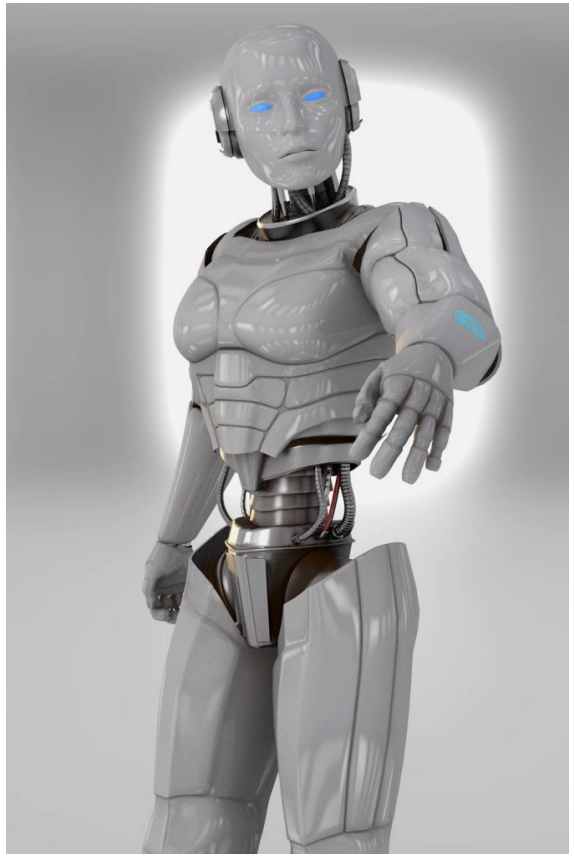
Cloud-Edge Continuum

Antonio Brogi

Department of Computer Science
University of Pisa

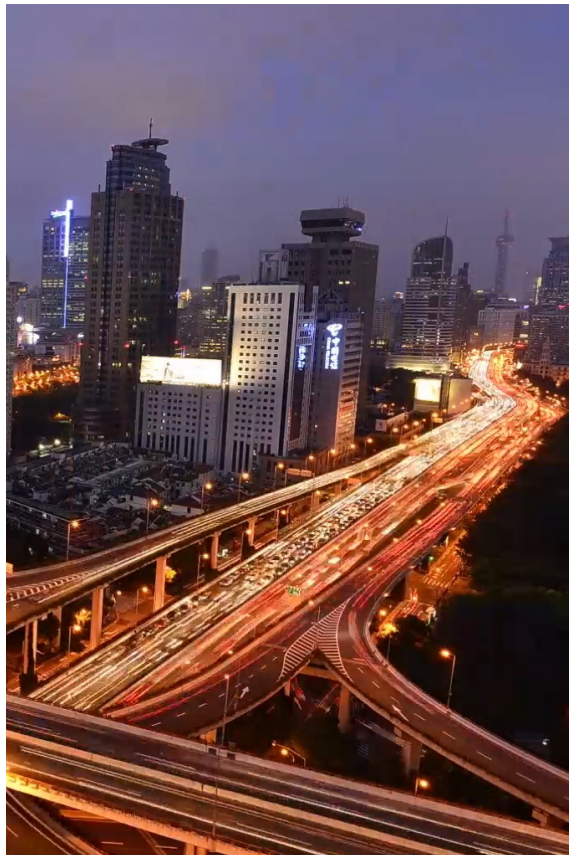
Cloud-Edge Continuum?

Pervasive IoT applications



Embedded AI

Energy production plants



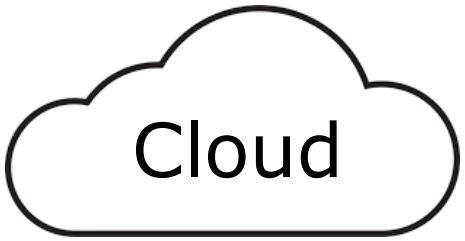
Smart Cities

IoT applications

sense | process | actuate

where?

Traditional deployment models



IoT + Cloud

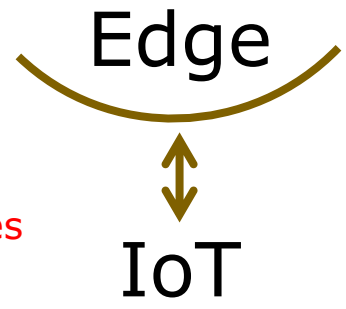
Data processed in the Cloud

\+ Unlimited processing and storage capabilties

\- Very many data transferred

\- High latencies

\- Mandatory Internet connectivity



IoT + Edge

Data processed at the Edge

\+ Low latencies

\- Limited processing and storage capabilities

IoT

So much data, really?

www.leverege.com/calculator

1	Moisture Sensor	SIZE:	8	bytes	1min	⊗
1	pH Sensor	SIZE:	8	bytes	1min	⊗
1	Thermometer	SIZE:	8	bytes	1h	⊗
1	Water Quality Monitor	SIZE:	56	bytes	1h	⊗
1	Videocamera	SIZE:	56250	bytes	1s	⊗

Resources Used

Data Transfer
145800.74
MB / month

146 GB/month

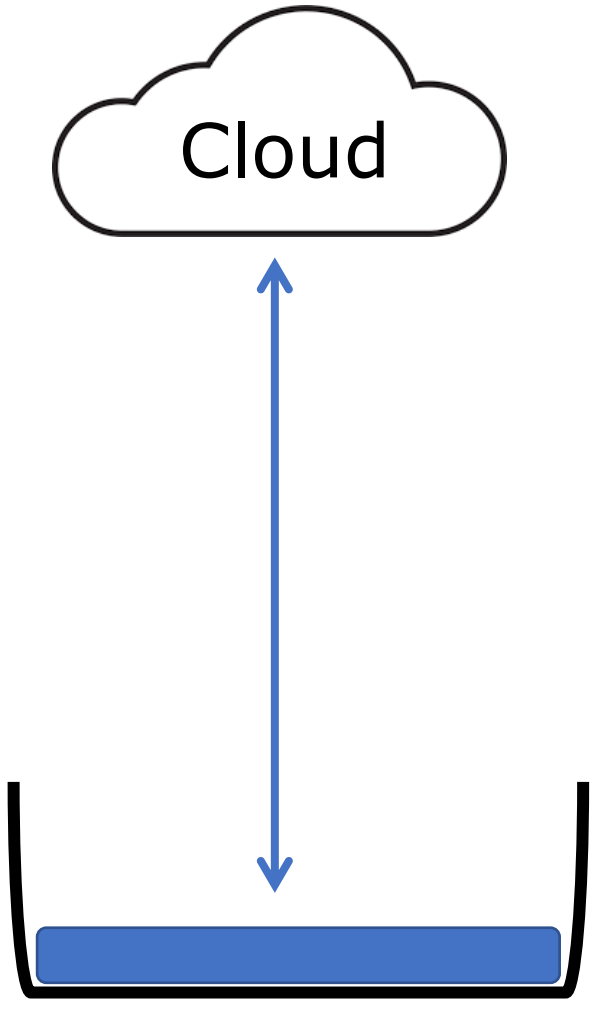
Cloud Storage
437402.21
MB / month

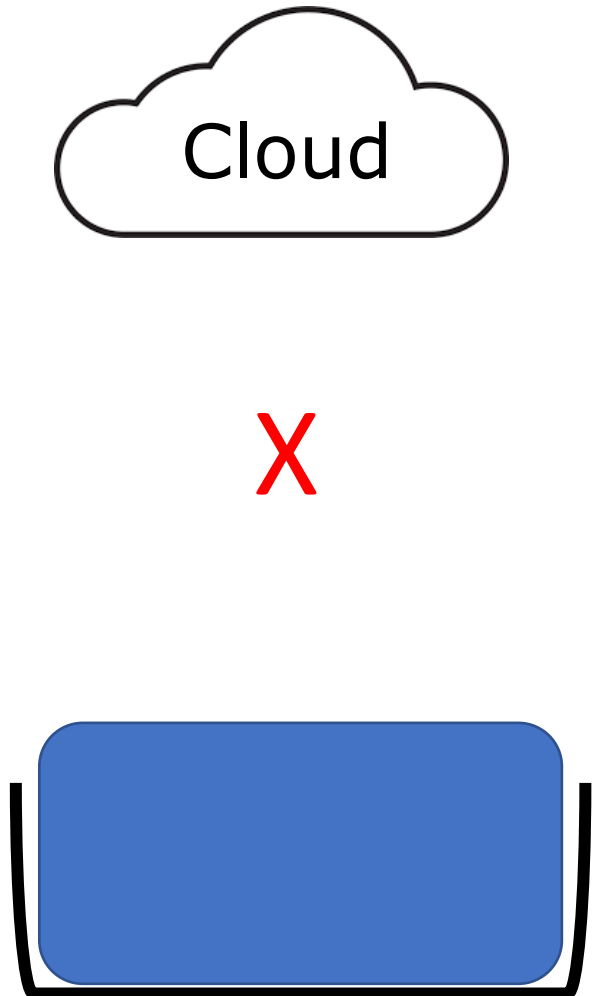
437 GB/month
5.2 TB/year

Updates
~2.68 M
per month

Mandatory Internet connectivity

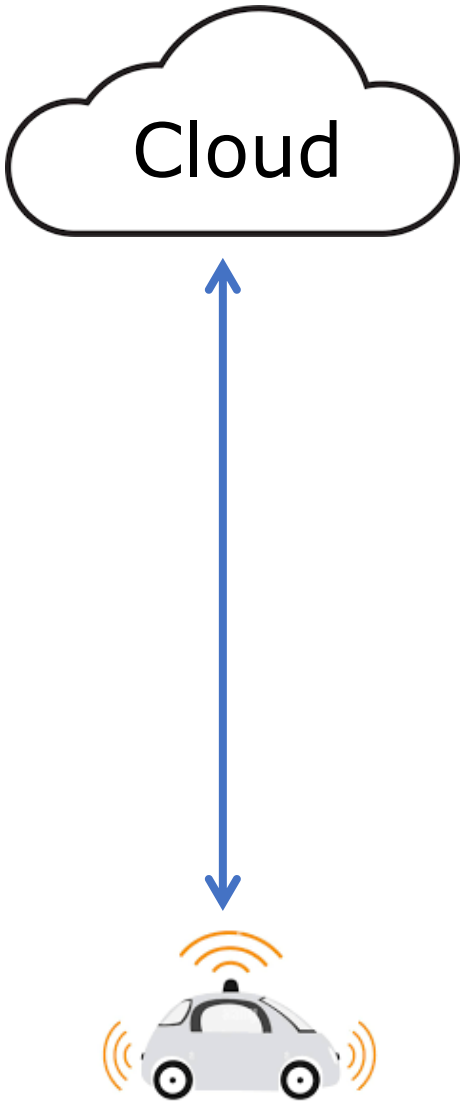
e.g. water flooding management must work in critical situations



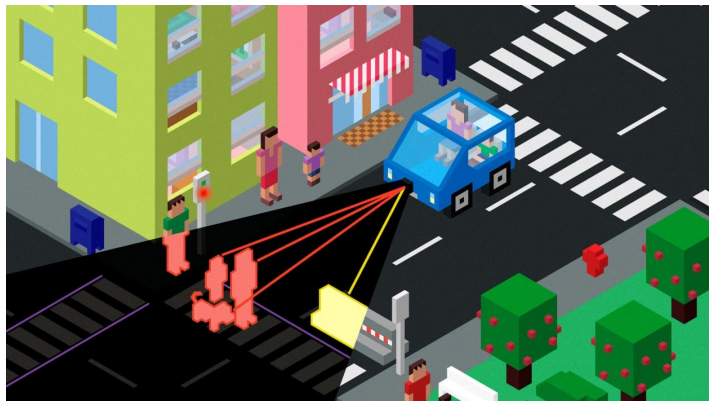


High latencies

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

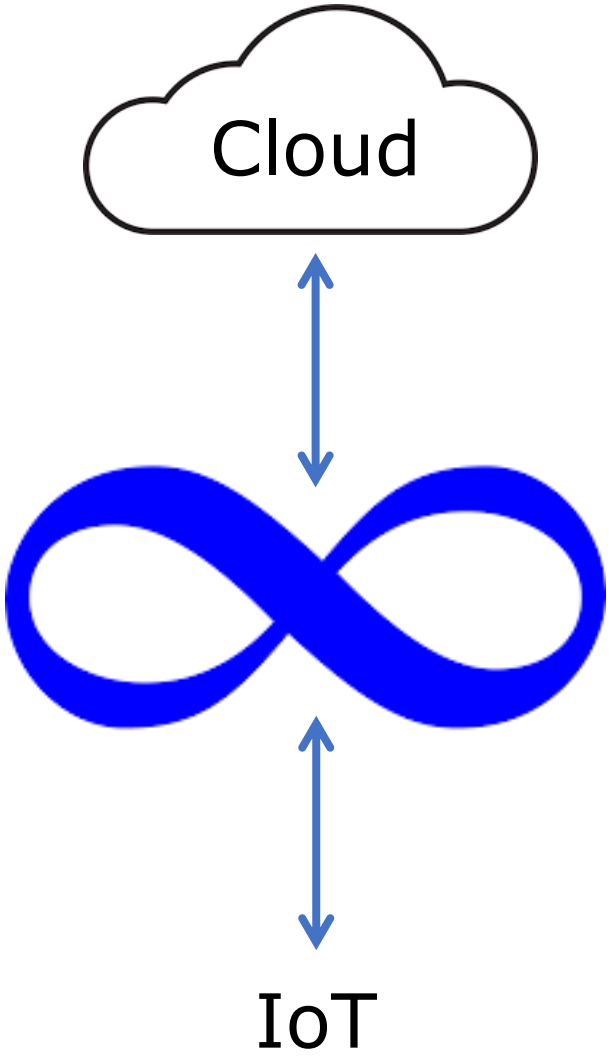


[ethics issues



too]

Cloud-Edge Continuum

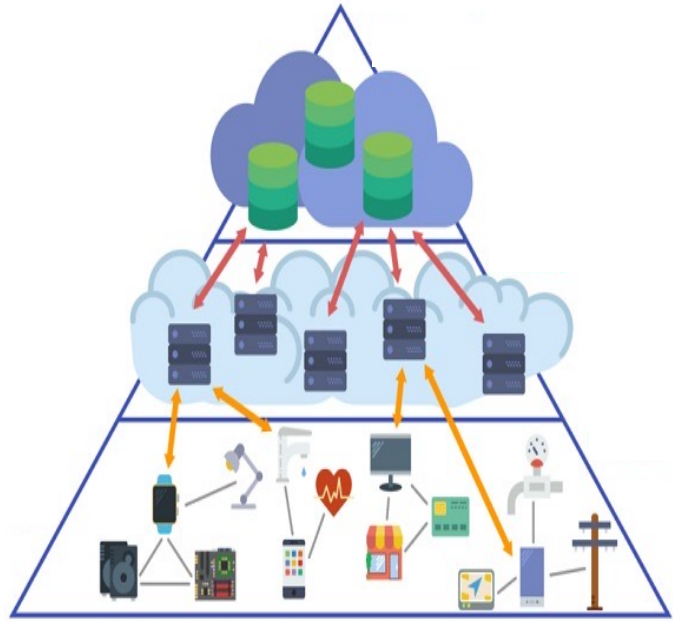


Extending the Cloud towards the IoT

with a **distributed, heterogenous infrastructure**

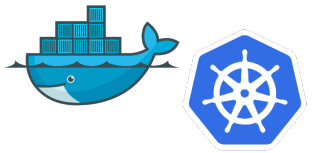
to get the "best of both worlds"

- computing power ✓
- connectivity ✓
- latency ✓



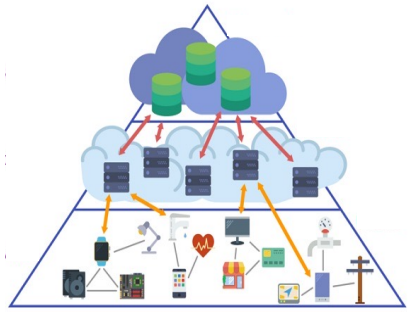
https://www.youtube.com/watch?v=ICQ0AAYO0mQ

Next-gen applications



Containerised,

microservice-based applications



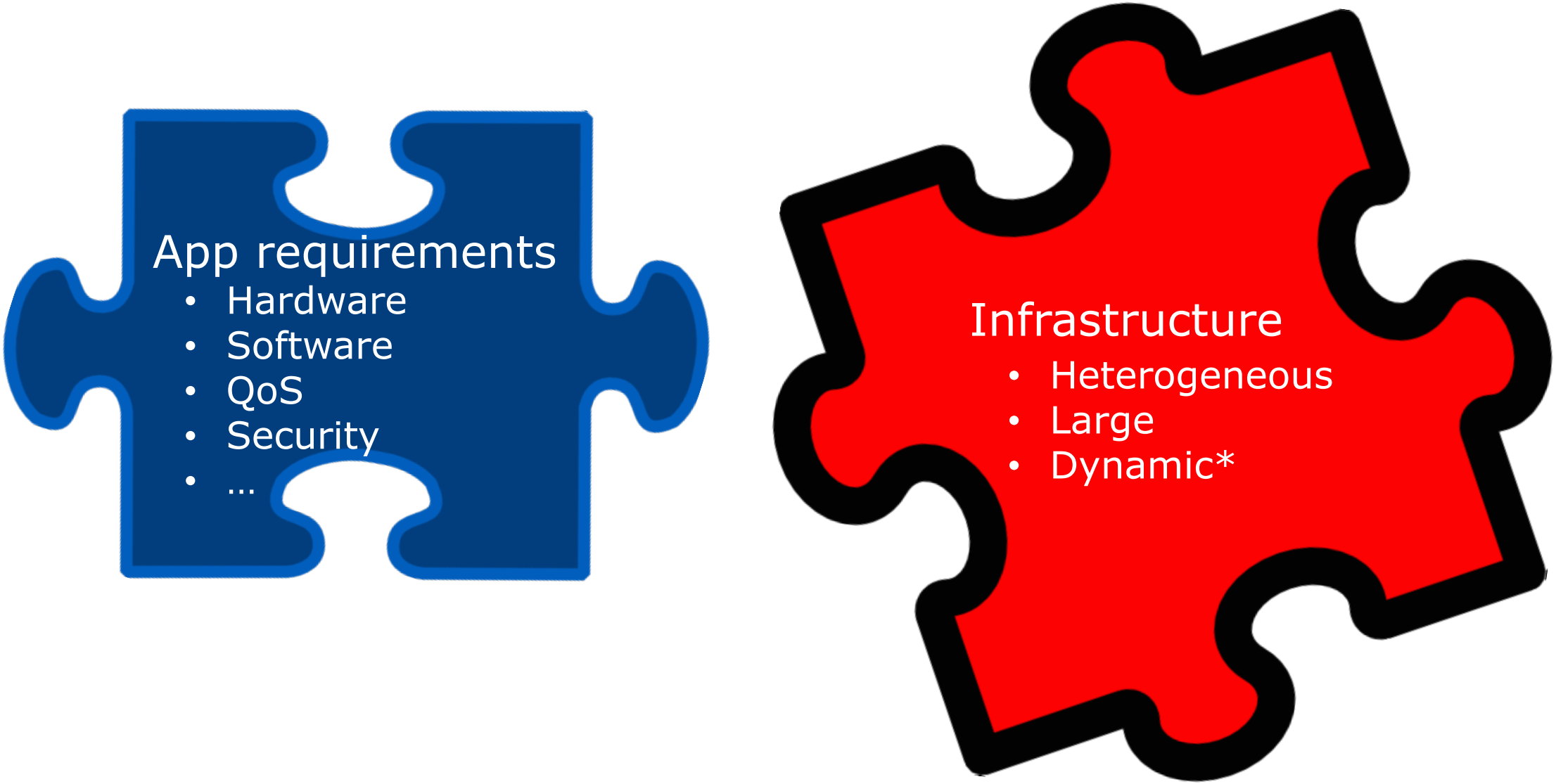
deployed on

a **continuous Cloud-Edge infrastructure**

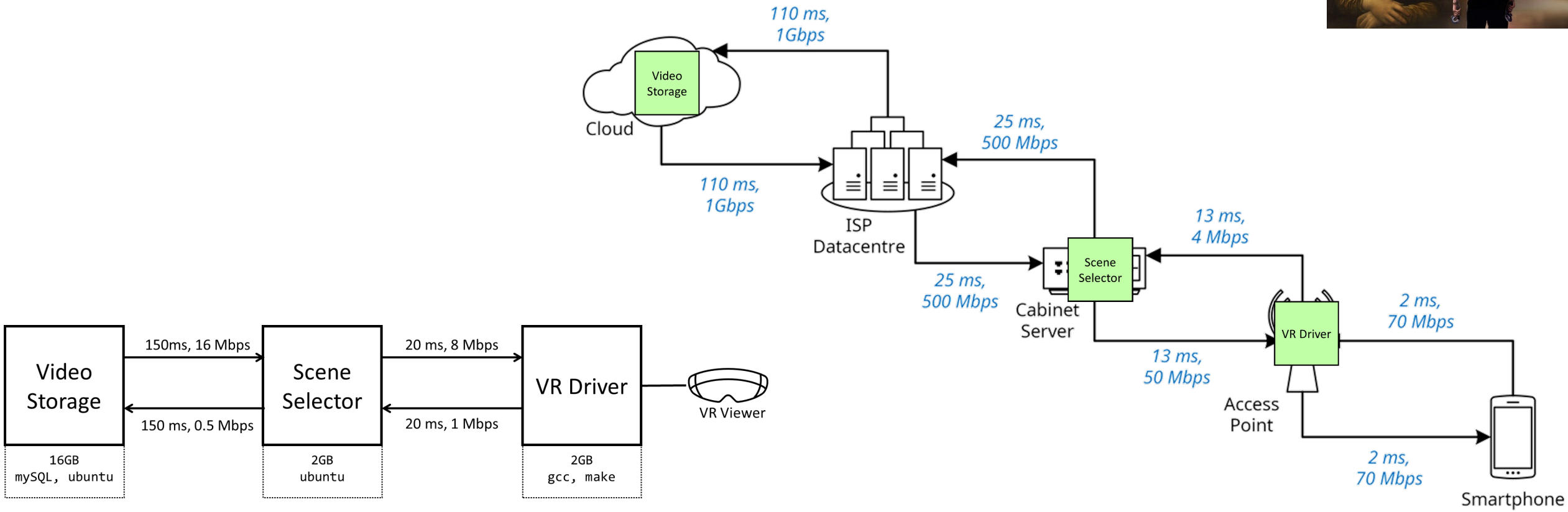
Cloud-Edge Continuum?

Application management

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



Example: Simple VR Application



Where should we deploy our IT solution?

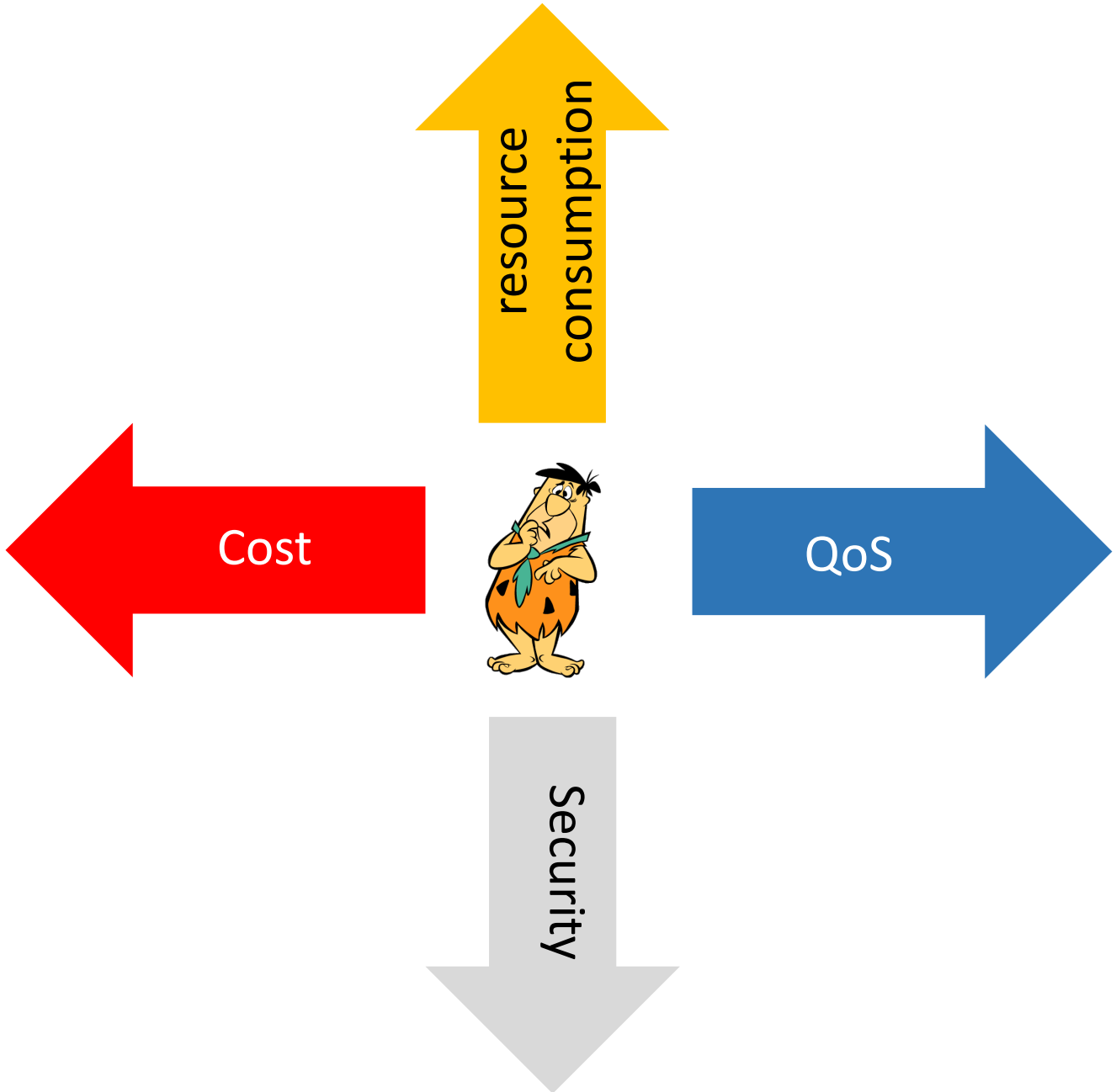
Should we get a new, more powerful server to be installed on-premise?

Should we get a better connectivity for our IoT devices?

How much can we trust that infrastructure provider?

…

Need **tools** helping to master orthogonal dimensions



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

Application placement: different approaches

MILP

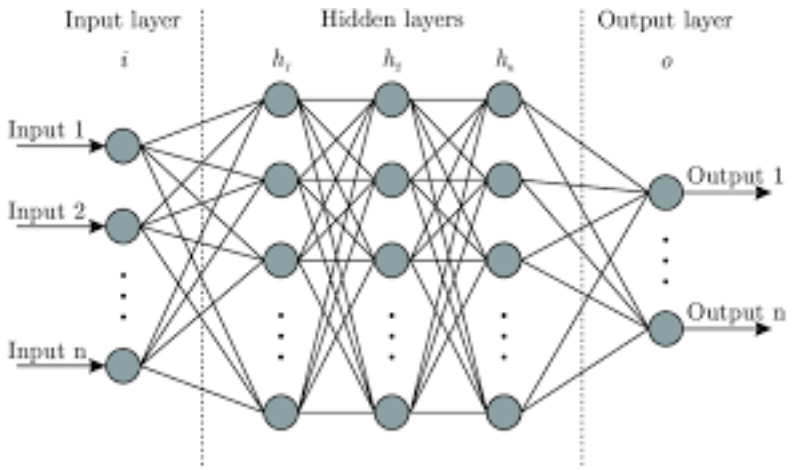
- X_{ij}^{k} = BINARY
- $\sum_{j=1}^{n} X_{0j}^{k} = \sum_{i=1}^{n} X_{i0}^{k} = 1$; k=1,2…m : Every route should start from depot and end on depot only,
- $\sum_{i=1}^{n} X_{ih}^{k} = \sum_{j=1}^{n} X_{hj}^{k} <= 1$, h=1,2,…n; k=1,2…m :every node should be selected at most once and every node served 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

Hard to code non-numerical info

Slow to run

ML



Infrastructure is very dynamic

Lack explainability

Declarative

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

Easy to read

Faster than MILP

Easy to code non-numerical info

Explainable

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

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

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

service S2 can be placed on node N2 ***and***

… ***and***

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

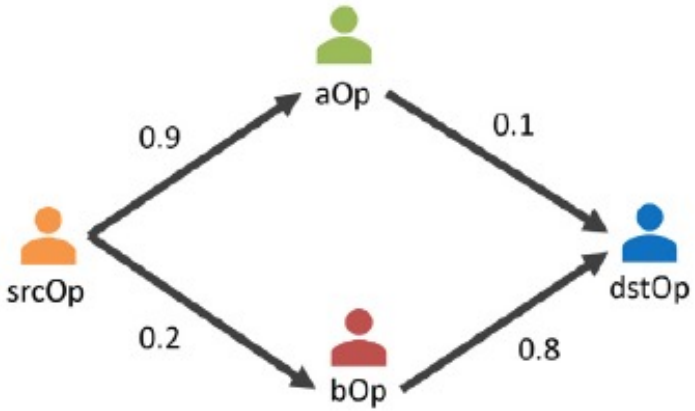
the QoS reqs of S1 … Sm are met

…

2) Let the inference engine look for it!

```
:- placement([s1,…,sm], P).
```

+ probabilities to model **infrastructure dynamicity**



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

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

Monitoring

monitoring (applications and infrastructure)

Continuous reasoning

Differential analysis

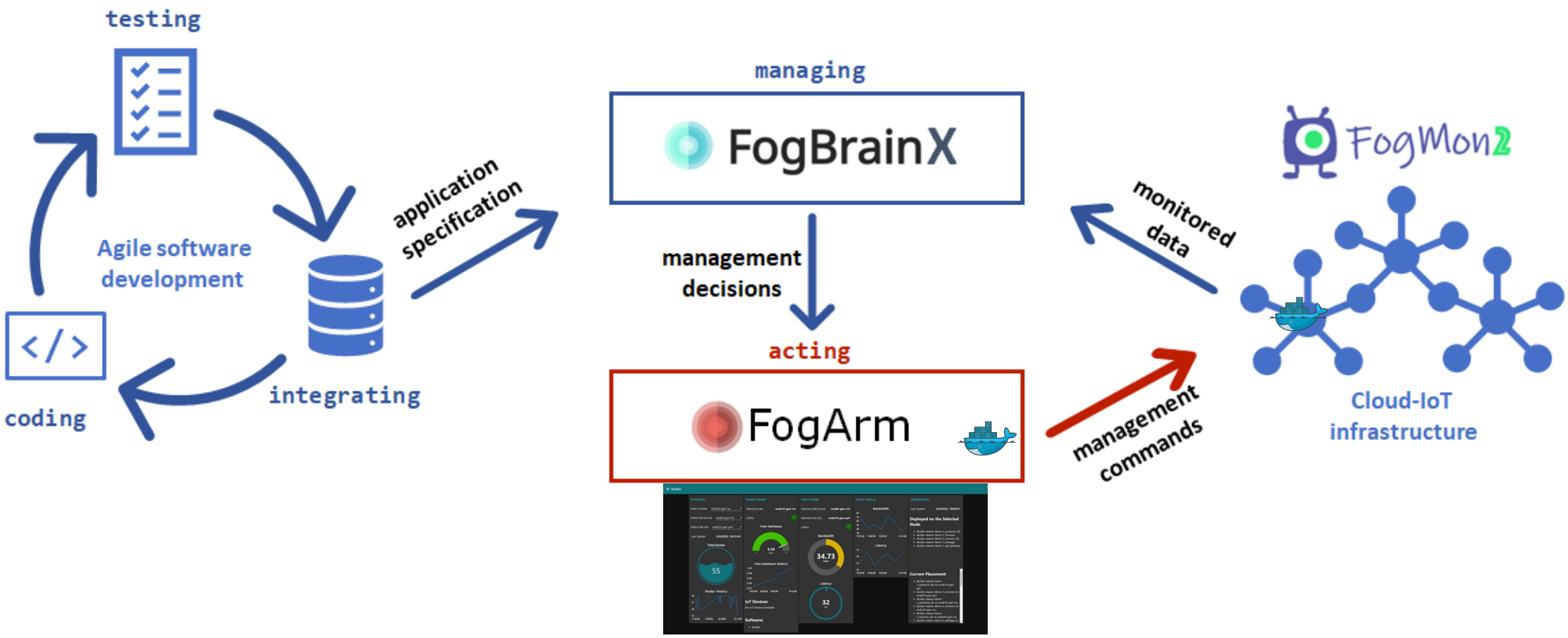
- focus on last changes

- re-use previously computed results

to scale up/down, migrate, restart application services

```
cr(P, NewP) :-
        partition(P,StableP,UnstableP),
        re-place(UnstableP,NewStableP),
        append(StableP,NewStableP,NewP).
```

Monitoring, reasoning, enacting



Cloud-Edge Continuum?

Application management

Concluding remarks

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

Many challenges:

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

- business models
  - no single infrastructure provider
  - service customers can become providers