MASSIVIZING COMPUTER SYSTEMS – 21st CENTURY ICT INFRASTRUCTURE

VU ON DIGITAL TWINS TO IMPROVE THE PERFORMANCE AND TECHNICAL SUSTAINABILITY OF DATACENTERS IN THE CONTINUUM

@Large Research
Massivizing Computer Systems

http://atlarge.science

bit.ly/VUDigitalTwin24

Digital Twin for ICT infrastructure = ModSim platform + monitoring and datagen + ODA + key scenarios + goal-oriented steering of RW infra.

This project has received funding from the European Union's Horizon Research and Innovation Actions under Grant Agreement № 101093202.

VRIJE UNIVERSITEIT AMSTERDAM

Contributi

Many thanks to our collaborators, international working groups, authors of all images included here. Also thanks to Adolfy Hoisie and all ModSim organizers.

GRAPH MASSIVIZER

CLOUD STARS

FUTURE 6G NETWORK SERVICES

Rijksoverheid

NWO

@Alosup

Prof.dr.ir. Alexandru IOSUP

SUSTAINABILITY IN 1 MINUTE

SUSTAINABLE, HIGH-END, HIGH-PERFORMANCE WRITING

Mabie Todd
"LongShort"
1911

Sailor
"Mini" 1st gen.
1960

Pilot Elite
95S (short)
2021

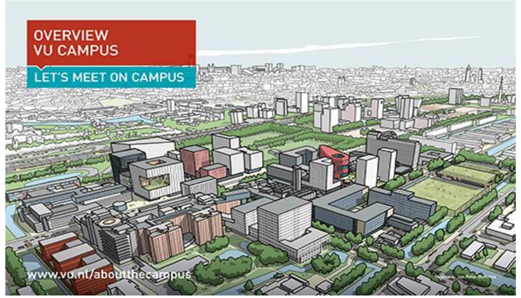
Source: fountainpens4drawing.nl

US IN 1+ MINUTES

WE’RE MASSIVIZING COMPUTER SYSTEMS!

VU AMSTERDAM < SCHIPHOL < THE NETHERLANDS < EUROPE

Amsterdam
founded 10th century
pop: 850,000



VU
founded 1880
pop: >25,000
(~ 5,000 staff)

The Netherlands

pop: 16.5 M

Europe

UNIVERSITY AMSTERDAM

http://atlarge.science

CURRENT TEAM

Professor

Assistant Prof.

Teacher

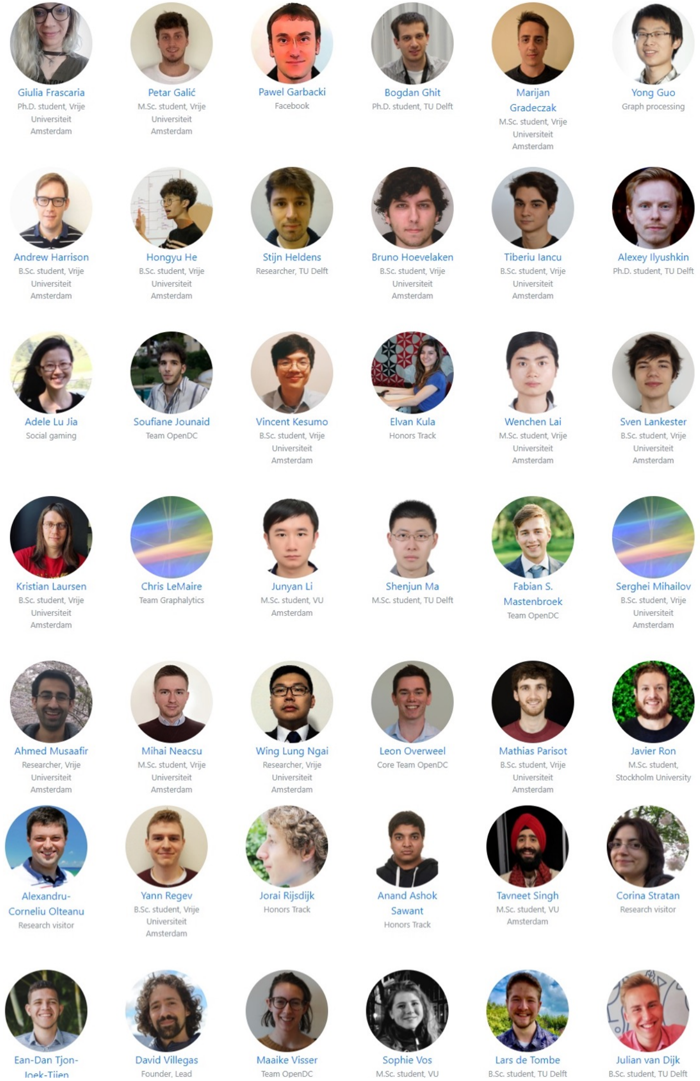
Visitor/P.-doc

Ph.D. student

Early Scientist MSc, BSc HP

ALUMNI

VISITORS



WE ARE A FRIENDLY, DIVERSE, LARGE GROUP, OF DIFFERENT RACES AND ETHNICITIES, GENDERS AND SEXUAL ORIENTATION, AND VIEWS OF CULTURE, POLITICS, AND RELIGION. YOU ARE WELCOME TO JOIN!

WHO AM I?
PROF. DR. IR. CAV. ALEXANDRU IOSUP

- Education, my courses:
  - Honours Programme, Computer Org. (BSc)
  - Distributed Systems (MSc)
- Research, 15+ years in DistribSys:
  - Massivizing Computer Systems (ecosystems)
  - Chair NL IPN SIG Future Computer Systems and Networki
  - Chair SPEC RG Cloud Group
- About me:
  - I like to help… I train people in need
  - VU University Research Chair + Group Chair
  - NL ICT Researcher of the Year
  - NL Higher-Education Teacher of the Year
  - NL Young Royal Academy of Arts & Sciences
  - Knighted in 2020

VU

Massivizing Computer Systems

Research Group Bio

https://atlarge-research.com





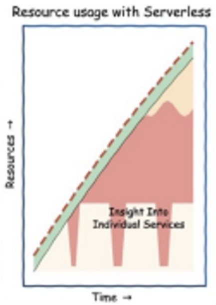
Alexandru IOSUP | Tiziano DE MATTEIS | Daniele BONETTA | Jesse DONKERVLIET | Matthijs JANSEN

Group Bio: CompSys | Massivizing Computer Systems group | National/EU projects, incl. NL Groeifondsprogramma 6G Future Network Services (€315M, 7 years), NL NWO OffSense, EU Horizon Graph-Massivizer (€5M), EU MCSA-RISE CLOUDSTARS

Serverless ± virtualized cloud environments: RM&S for workflow and serverless ops, back-end services, big data and graphs, aggregate and disaggregate resources, VM/contrainerization and JIT compiling, performance, availability, energy. Understand and experiment / analyze / benchmark ecosystems, design new parts, improve existing parts, share FAIR tools + Memex-like data.

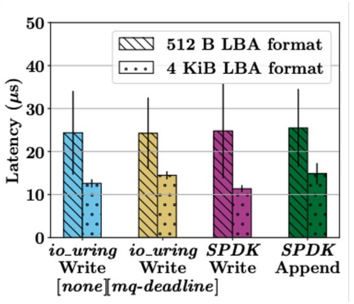
Digital Twin: ICT infrastructure/datacenter simulation, Operational Data Analytics, DC cockpit, VR/XR ops.

Relevant prior work (selection of tools and activities):

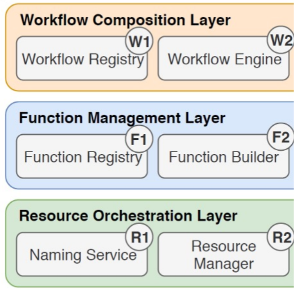


Serverless history + vision

OpenDC simulator



K8s/Storage benchmark



Continuum + Reference Archi

Benchmark

Cloud Group chair

Contact: info@atlarge-research.com




SINCE LAST YEAR – RE-BOOTED THE COMPSYS NL COMMUNITY…

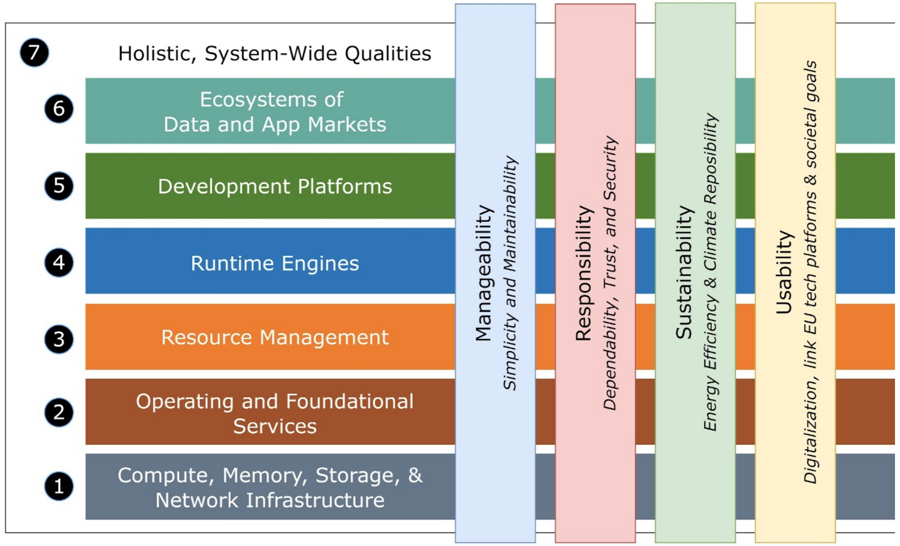
ICT-RESEARCH PLATFORM NETHERLANDS

SIG FCSN + Manifesto on Computer Systems and Networking Research

Clear vision for the field in the NL, 2021-2035

Signed

50 PIs / Leads
07 universities
05 relevant societal stakeholders



Available

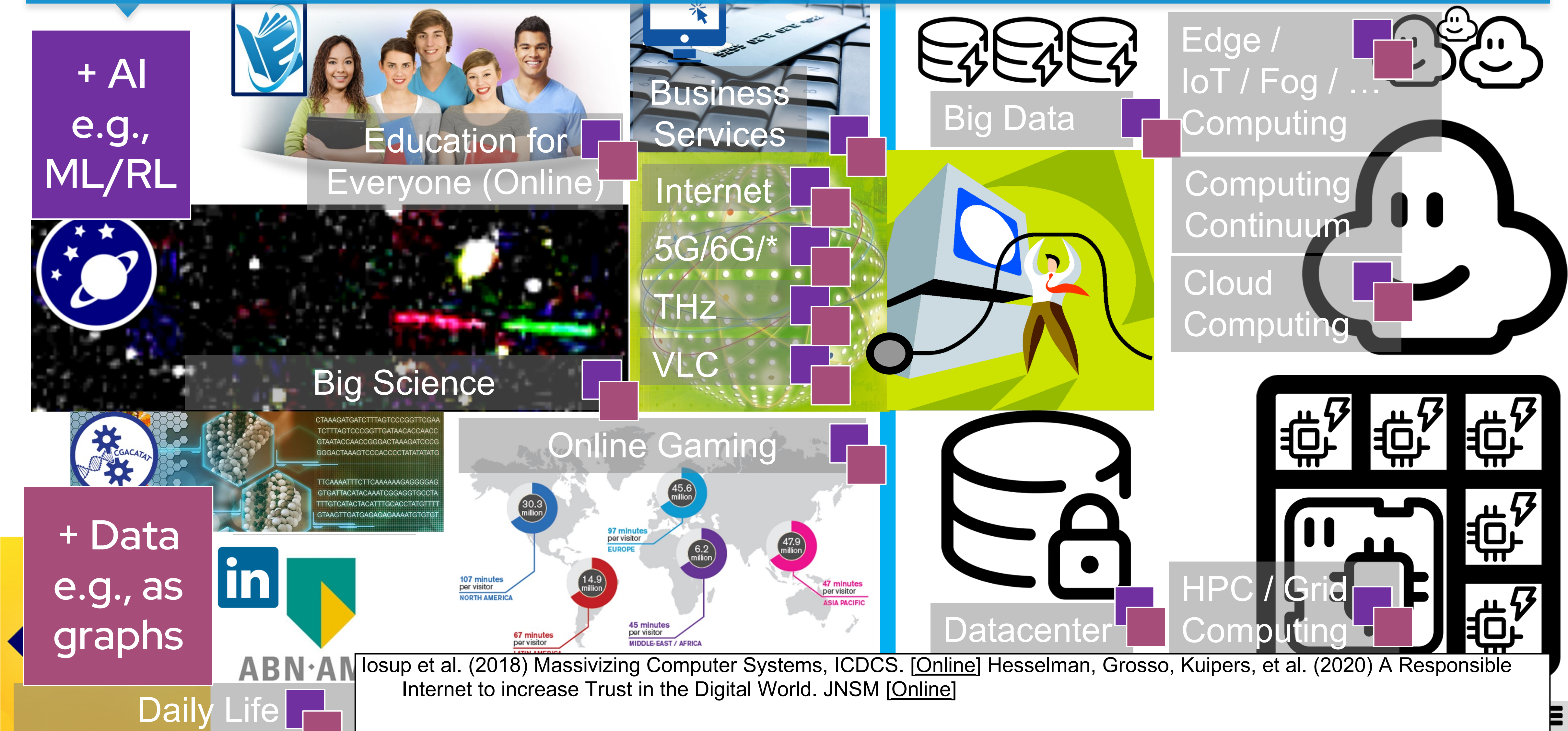
Full version https://arxiv.org/pdf/2206.03259
Who's Who in CompSysNL? https://bit.ly/CompSysNLWhosWho

VU VRIJE UNIVERSITEIT AMSTERDAM

THIS IS THE GOLDEN AGE OF COMPUTER ECOSYSTEMS

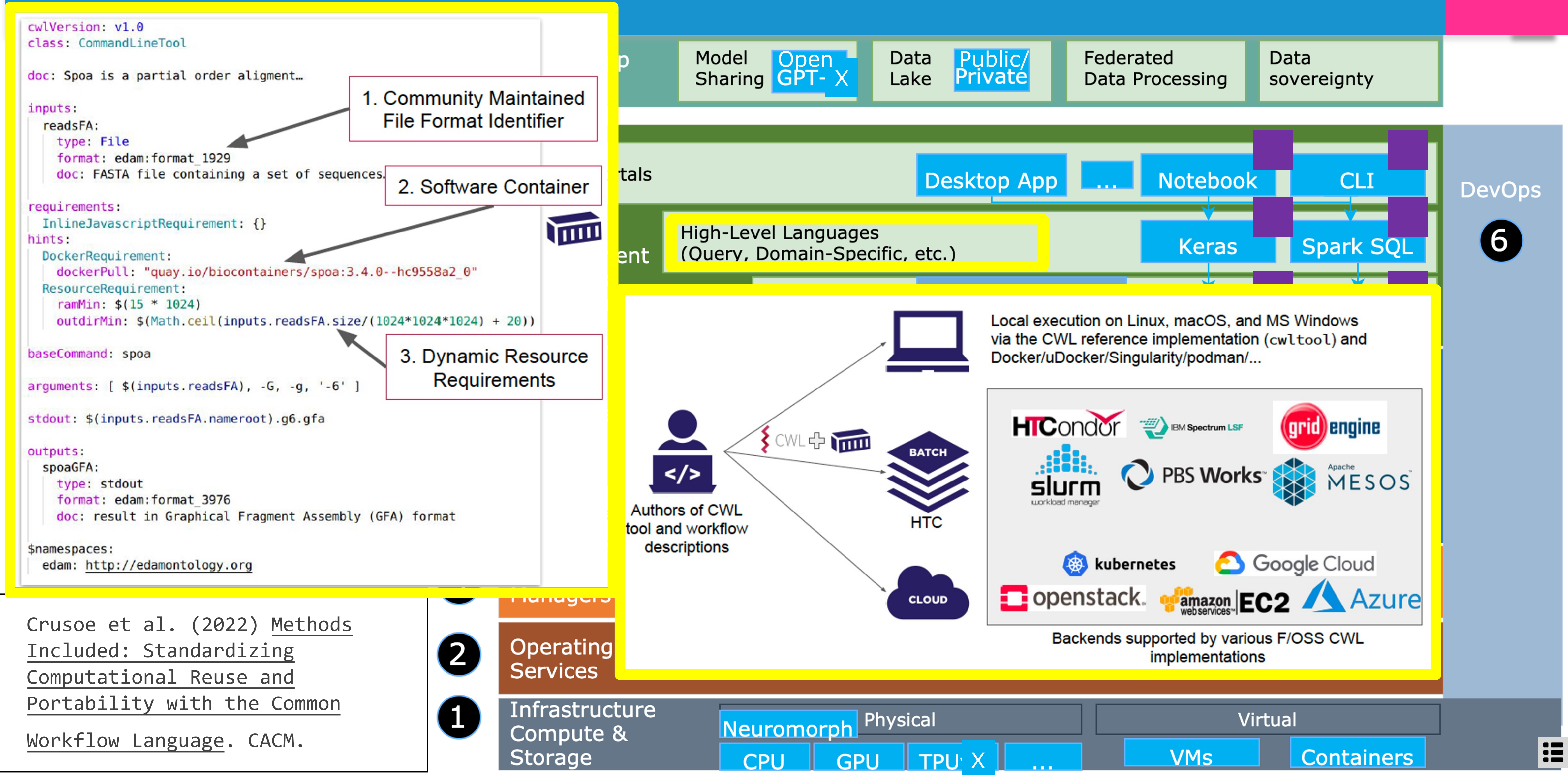
1

GENERALITY OF MASSIVE COMPUTER ECOSYSTEMS



Iosup et al. (2018) Massivizing Computer Systems, ICDCS. [Online] Hesselman, Grosso, Kuipers, et al. (2020) A Responsible Internet to increase Trust in the Digital World. JNSM [Online]

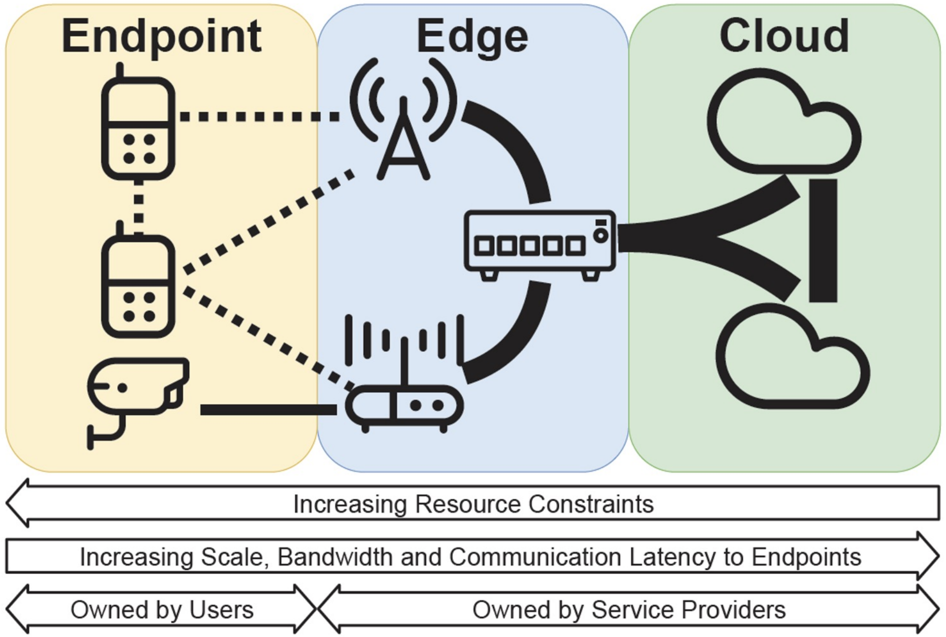
USING REPRODUCIBLE, COMPLEX WORKFLOWS …



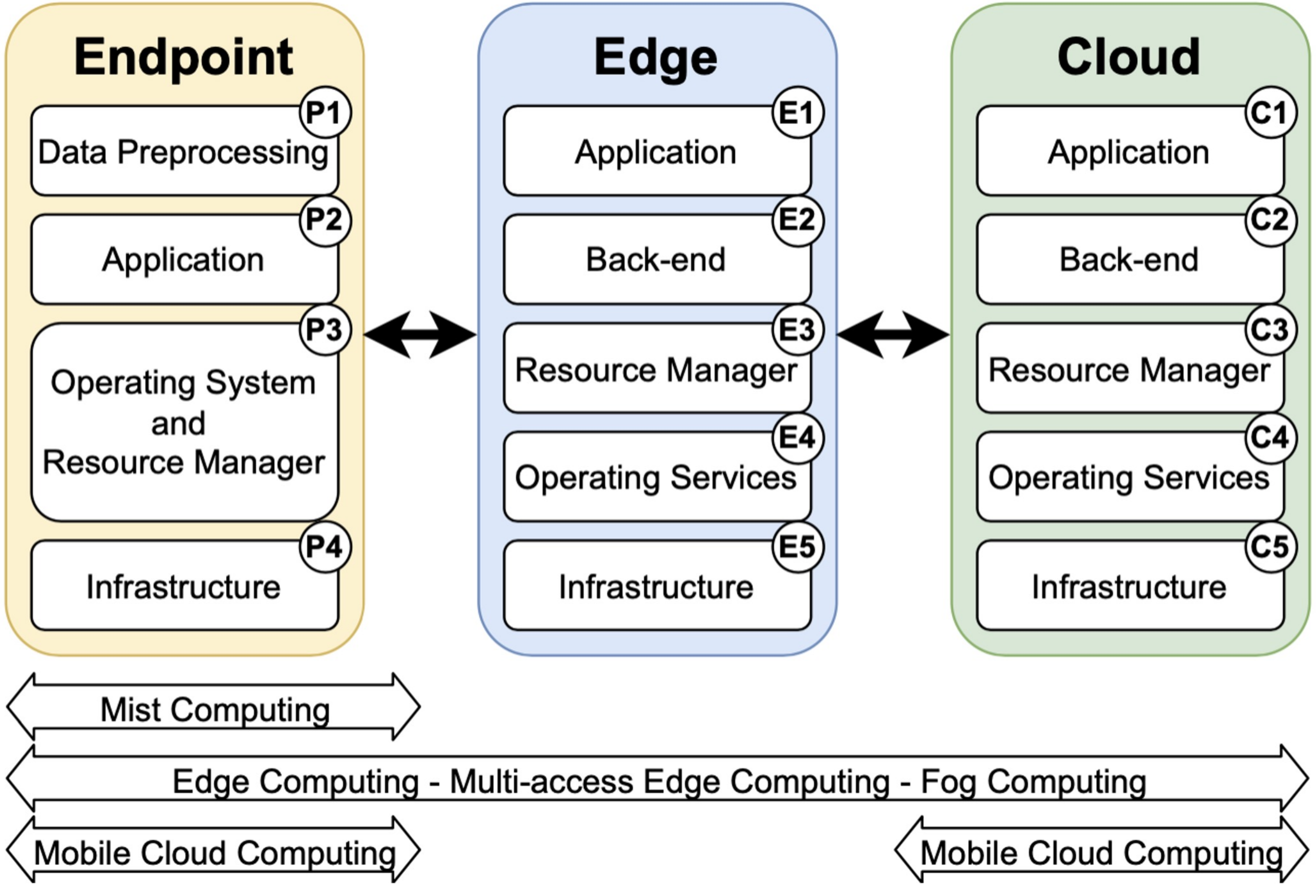
Crusoe et al. (2022) Methods Included: Standardizing Computational Reuse and Portability with the Common Workflow Language. CACM.

... RUNNING ON CONTINUUM RESOURCES & SERVICES

ISSUES: COMPLEXITY, NON-TECHNICAL

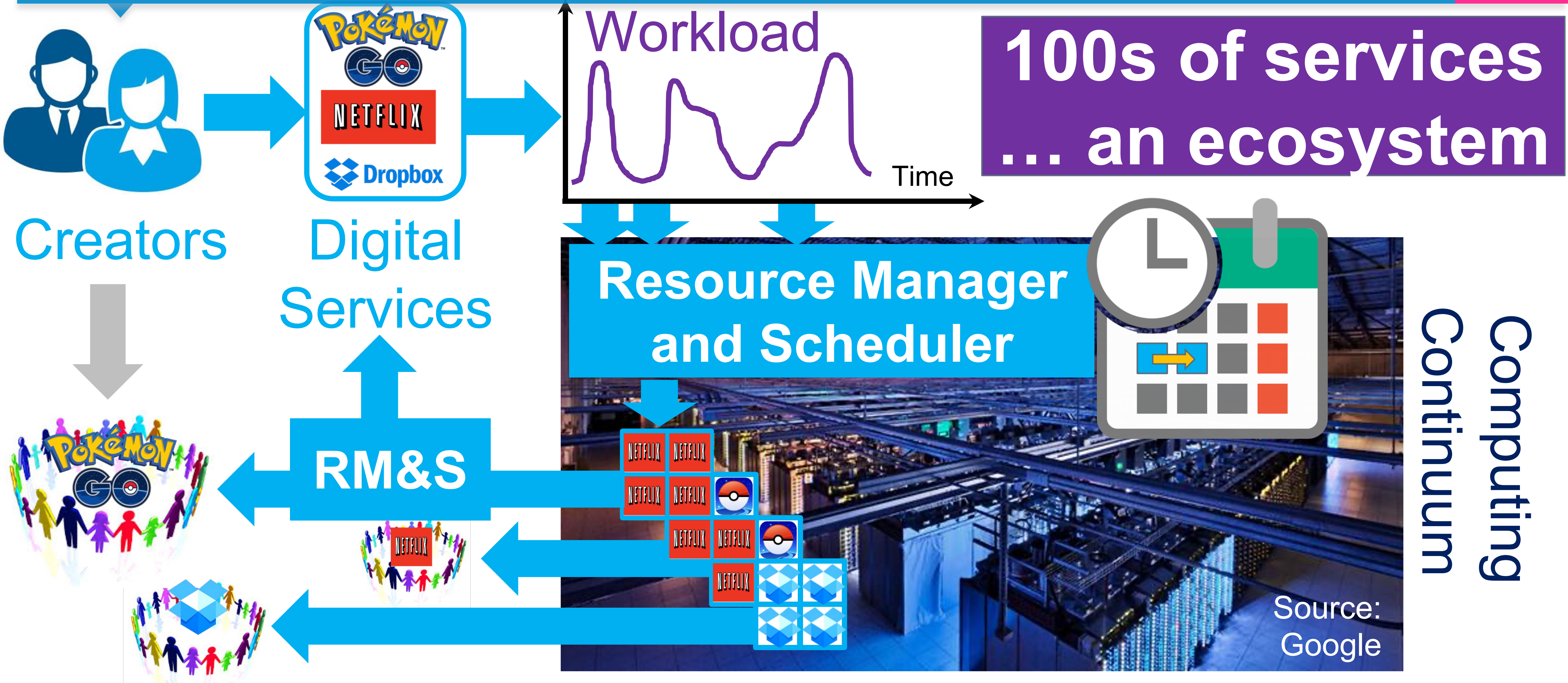


Trivedi, Wang, Bal, Iosup (2021) Sharing and Caring of Data at the Edge. HotEdge.



Jansen, Al-Dulaimy, Papadopoulos, Trivedi, Iosup (2023) The SPEC-RG Reference Architecture for the Edge Continuum. CCGRID. Open access: https://doi.org/10.48550/ARXIV.2207.04159

… IN A SMARTLY ORCHESTRATED ICT ECOSYSTEM …



Extreme Automation, Performance, Dependability, Sustainability

…DELIVERING SERVERLESS COMPUTING PROPERTIES

^

bit.ly/MassivizingServerless22

Serverless computing =
extreme automation
+ fine-grained, utilization-based billing

Application Type

42% Core functionality
39% Utility functionality
16% Scientific workload

Programming Languages

42% JavaScript
42% Python
12% Java

Simon Eismann, Joel Scheuner, Erwin Van Eyk, Maximilian Schwinger, Johannes Grohmann, Nikolas Herbst, Cristina L. Abad, Alexandru Iosup (2022) The State of Serverless Applications: Collection, Characterization, and Community Consensus. IEEE Trans. Software Eng. 48(10)

Dispelling the confusion around serverless computing by capturing its essential and conceptual characteristics.

BY SAMUEL KOUNEV, NIKOLAS HERBST, CRISTINA L. ABAD, ALEXANDRU IOSUP, IAN FOSTER, PRASHANT SHENOY, OMER RANA, AND ANDREW A. CHIEN

Serverless Computing: What It Is, and What It Is Not?

Market analysts are agreed that serverless computing has strong market potential, with projected compound annual growth rates (CAGRs) varying between 21% and 28% through 2028[4,25,33,35,49] and a projected market value of $36.8 billion[49] by that time. Early adopters are attracted by expected cost reductions (47%), reduced operation effort (34%), and scalability (34%).[17] In research, the number of peer-reviewed publications connected to serverless computing has risen steadily since 2017.[46] In industry, the term is heavily used in cloud provider advertisements and even in the naming of specific products or services.

Yet despite this enthusiasm, there exists no common and precise understanding of what serverless is (and of what it is not). Indeed, existing definitions of serverless computing are largely inconsistent and unspecific, which leads to confusion in the use of not only this term but also related terms such as cloud computing, cloud-native, Container-as-a-Service (CaaS), Platform-as-a-Service (PaaS), Function-

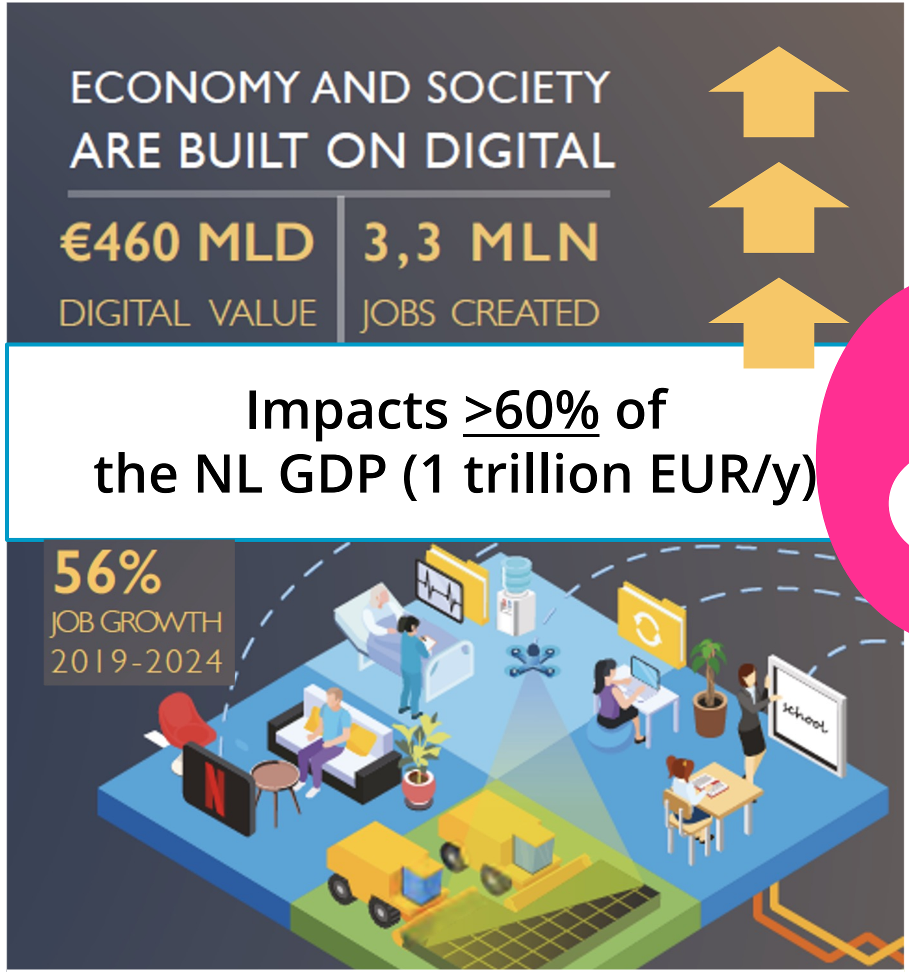
Kounev, Herbst, Abad, Iosup, Foster, Shenoy, Rana, Chien (2023) Serverless Computing: What It Is, and What It Is Not? CACM. Sep 2023 issue.

BUT WE CANNOT TAKE THIS TECHNOLOGY FOR GRANTED

2

(A few facets of sustainability)

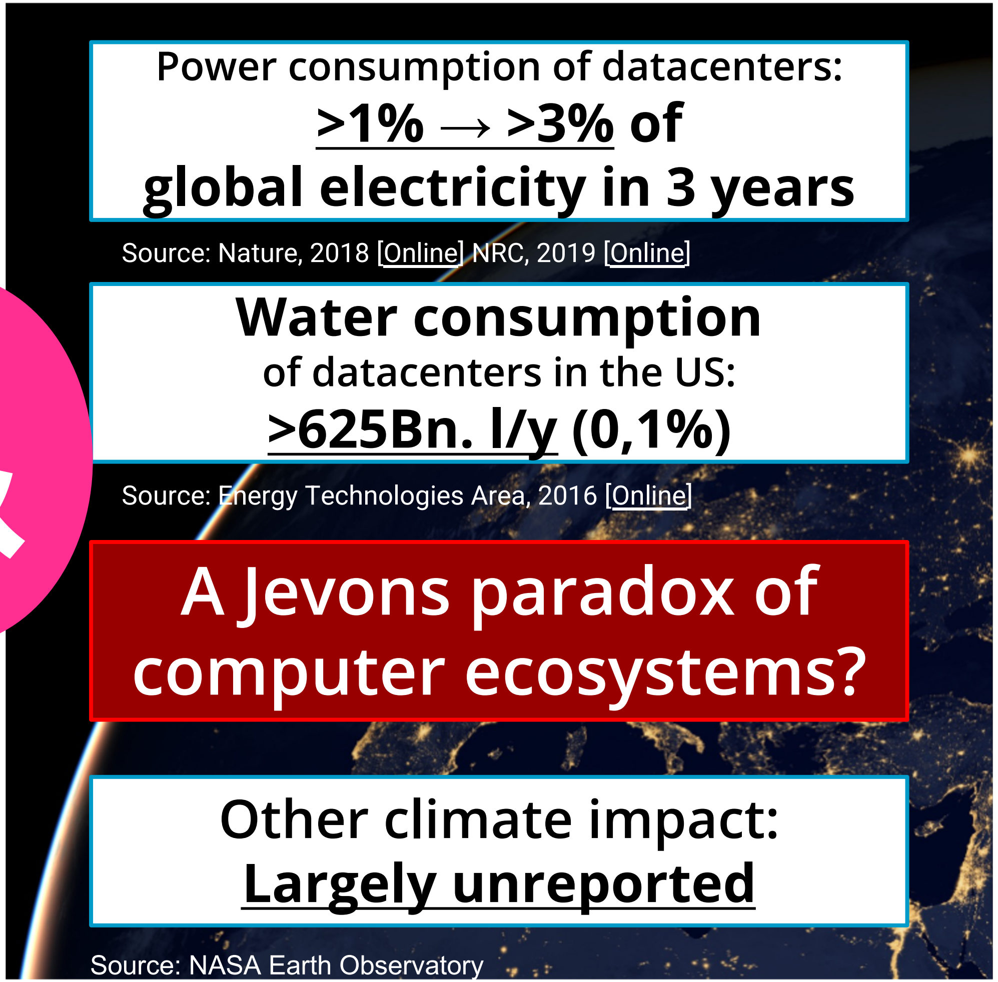
ECONOMIC AND CLIMATE SUSTAINABILITY OF MASSIVE COMPUTER ECOSYSTEMS



Impacts <u>>60%</u> of
the NL GDP (1 trillion EUR/y)

Sources: Iosup et al., Massivizing Computer Systems, ICDCS 2018 [Online] / Dutch Data Center Association, 2020 [Online] / Growth: NL Gov't, Flexera, Binx 2020. Gartner 2019. IA 2017.

&



Power consumption of datacenters:
<u>>1% → >3%</u> of
global electricity in 3 years

Source: Nature, 2018 [Online] NRC, 2019 [Online]

Water consumption
of datacenters in the US:
<u>>625Bn. l/y</u> (0,1%)

Source: Energy Technologies Area, 2016 [Online]

A Jevons paradox of computer ecosystems?

Other climate impact:
<u>Largely unreported</u>

Source: NASA Earth Observatory

SUSTAINABILITY MUST ADDRESS COMPLEXITY GROWTH

COMPLEX, DISTRIBUTED ECOSYSTEMS DO NOT ACT LIKE REGULAR COMPUTER SYSTEMS

Ecosystems don't have easily…

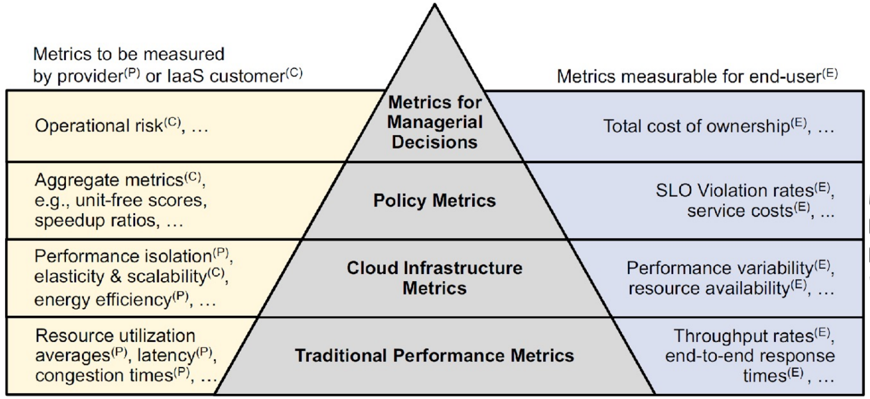
- Simplicity
- Maintainability
- Responsibility
- Sustainability
- Usability

which includes:

- Synchronization
- Consistency, consensus
- Performance
- Scalability, elasticity
- Availability, reliability
- Energy-efficiency

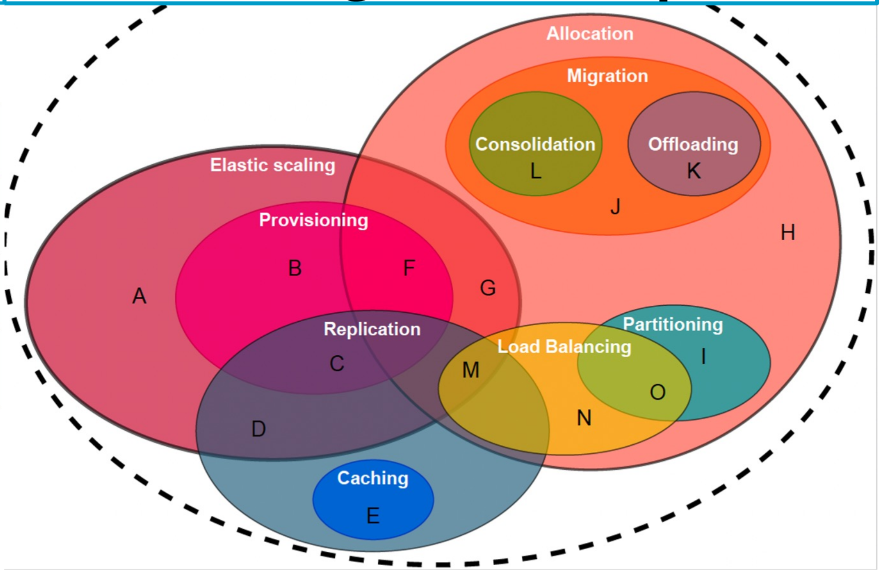
Iosup, Kuipers, Trivedi, et al. (2022) Future Computer Systems and Networking Research in the Netherlands: A Manifesto. CORR

Operational goals are becoming more complex



N. Herbst, E. Van Eyk, A. Iosup, et al. (2018) Quantifying Cloud Performance and Dependability: Taxonomy, Metric Design, and Emerging Challenges. TOMPECS 3(4).

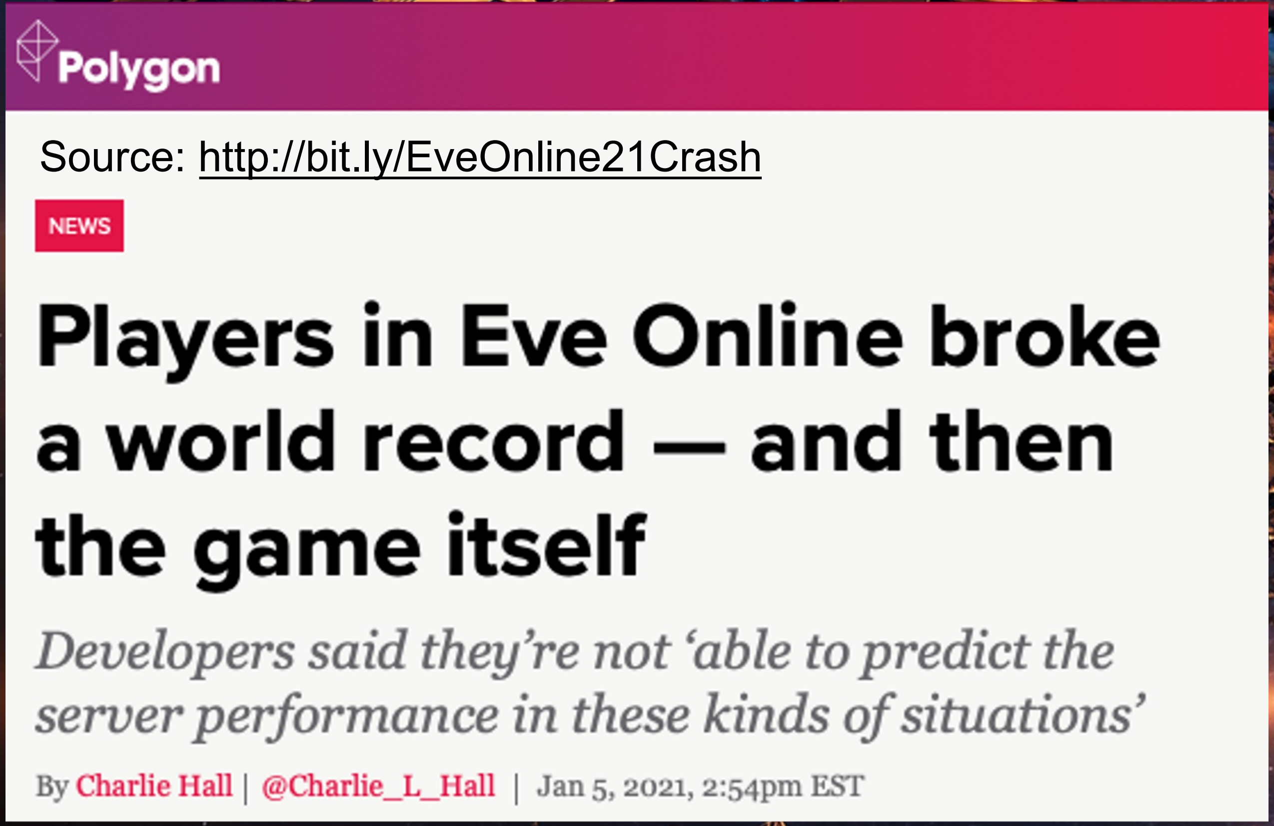
Operational techniques are becoming more complex



Stijn Meijerink, Erwin van Eyk, Alexandru Iosup (2021) Multivocal Survey of Operational Techniques for Serverless Computing. White Paper.

PHENOMENON: PERFORMANCE IN CLOUD SERVICES

PERFORMANCE SUSTAINABILITY OF MASSIVE COMPUTER ECOSYSTEMS



Polygon

Source: http://bit.ly/EveOnline21Crash

NEWS

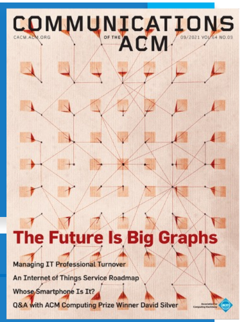
Players in Eve Online broke a world record — and then the game itself

Developers said they're not 'able to predict the server performance in these kinds of situations'

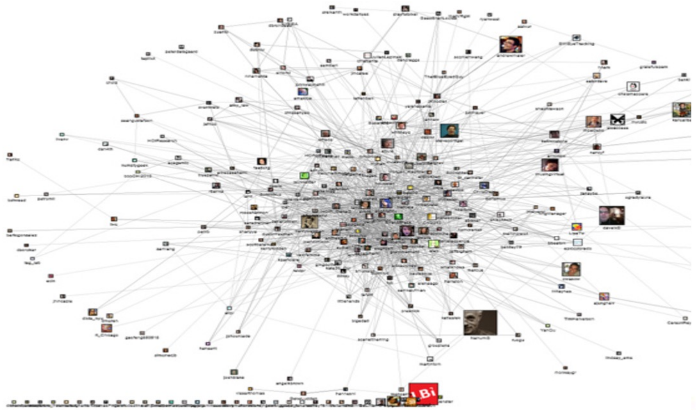
By Charlie Hall | @Charlie_L_Hall | Jan 5, 2021, 2:54pm EST

Source: Razorien/CCP Games

GRAPH DATA EVERYWHERE, UNPRECEDENTED SCALE

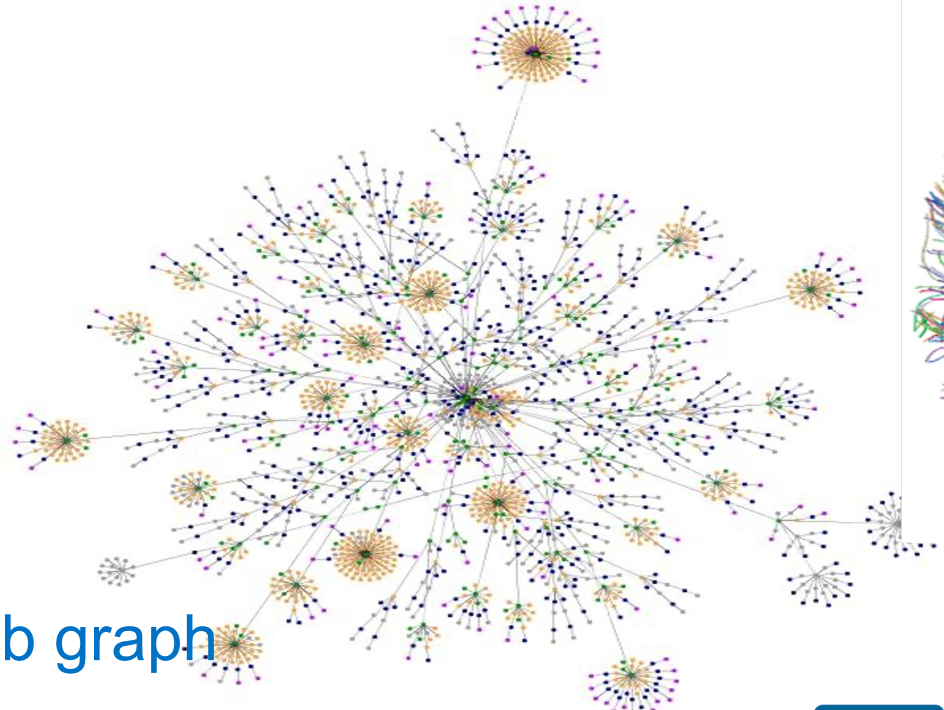


NEED TO MASSIVIZE GRAPH PROCESSING



Social / gaming network

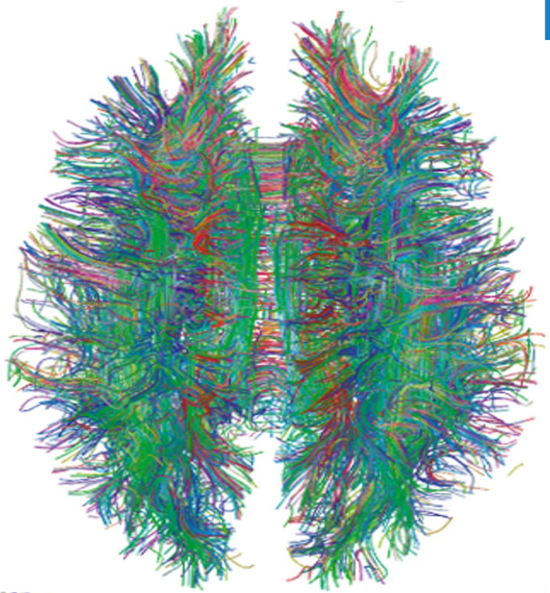
~1 billion vertices
~100 billion connections



Web graph

~50 billion pages
~1 trillion hyperlinks

Linked in

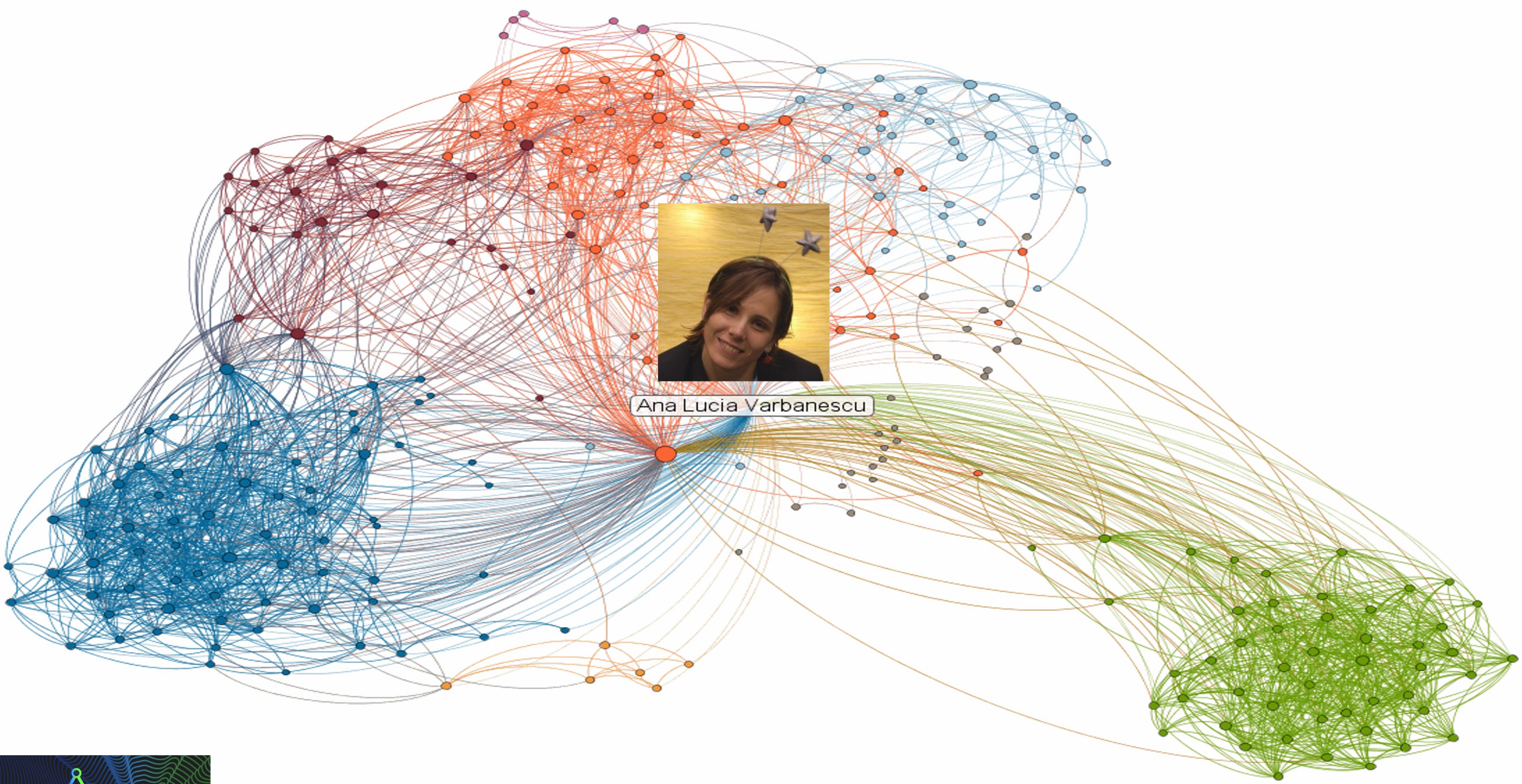


Brain network

~100 billion neurons
~100 trillion connections

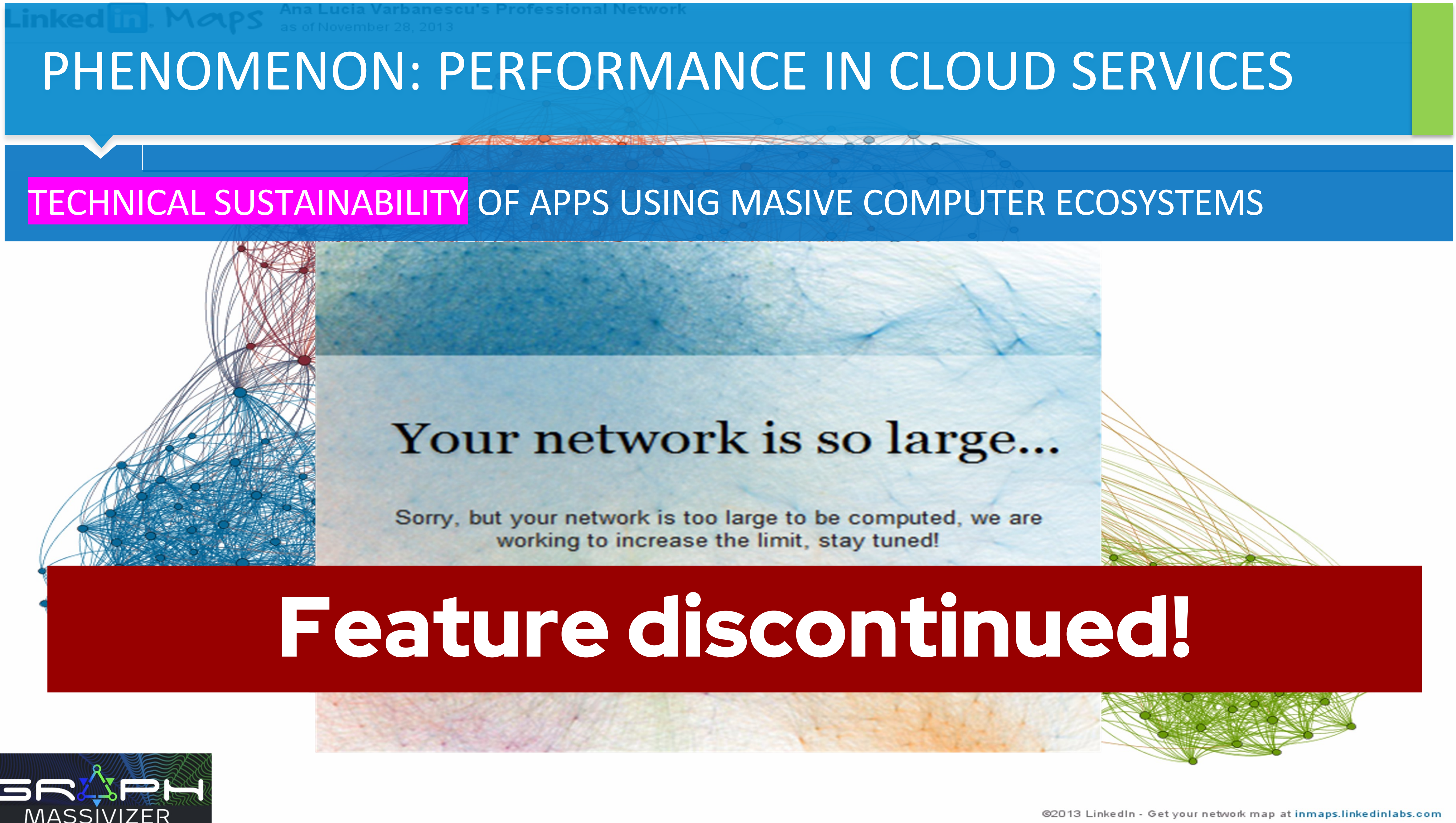
Sakr, Bonifati, Voigt, Iosup, et al. (2021) The Future Is Big Graphs! CACM.

LinkedIn Maps **Ana Lucia Varbanescu's Professional Network**
as of November 28, 2013



PHENOMENON: PERFORMANCE IN CLOUD SERVICES

TECHNICAL SUSTAINABILITY OF APPS USING MASIVE COMPUTER ECOSYSTEMS



Feature discontinued!

GRAPH MASSIVIZER

OVER-ARCHING GOAL:

WE'RE BUILDING 21st CENTURY ICT INFRASTRUCTURE

3

BUILDING ICT INFRASTRUCTURE FOR THE 21ST CENTURY

THIS TALK IN A NUTSHELL

Technology **not ready, many issues**

A Why does this happen?

B What to do about it?

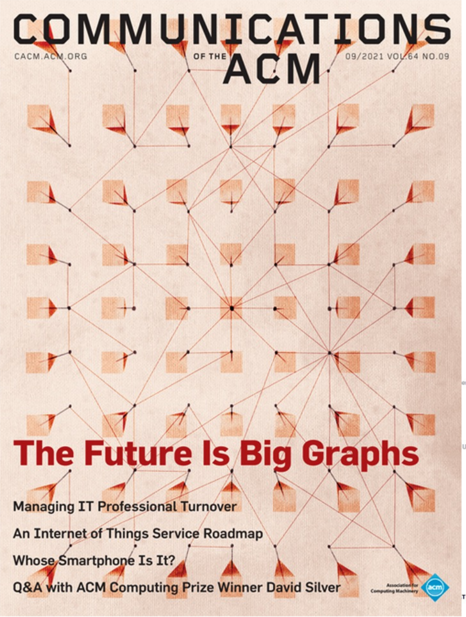
In modern computer systems, issues are often linked

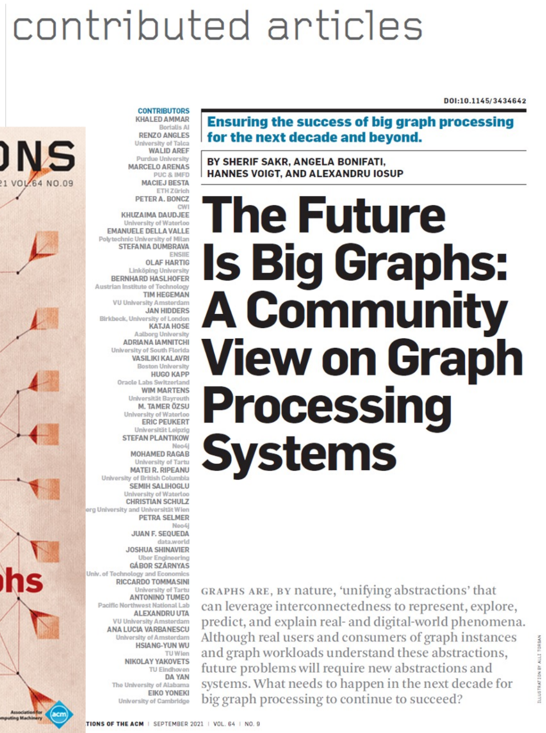
Source: Alexandru's personal library.

GRAPH MASSIVIZER

Big Graph Processing: Used in AI/ML, FinTech, ICT Infra., Industry 4.0, Energy Mgmt.*, etc.

Vision: Massivizing computer systems approaches are key to enable big graph ecosystems




contributed articles

DOI:10.1145/3434642

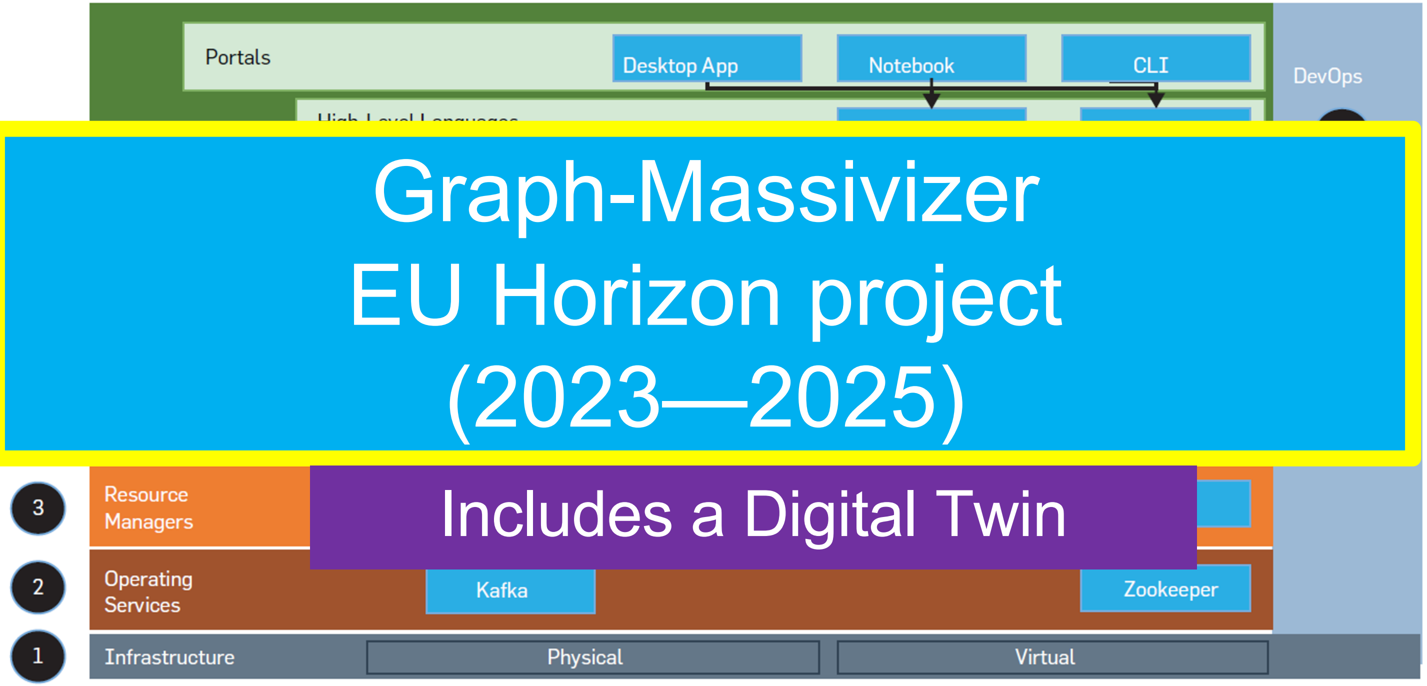
Ensuring the success of big graph processing for the next decade and beyond.

BY SHERIF SAKR, ANGELA BONIFATI, HANNES VOIGT, AND ALEXANDRU IOSUP

The Future Is Big Graphs: A Community View on Graph Processing Systems

GRAPHS ARE, BY nature, 'unifying abstractions' that can leverage interconnectedness to represent, explore, predict, and explain real- and digital-world phenomena. Although real users and consumers of graph instances and graph workloads understand these abstractions, future problems will require new abstractions and systems. What needs to happen in the next decade for big graph processing to continue to succeed?

CACM Cover/Featured article, Sep 2021



Graph-Massivizer
EU Horizon project
(2023—2025)

Includes a Digital Twin

(*) Digital twin for datacenters, with partners including CINECA

Sakr, Bonifati, Voigt, Iosup, et al. (2021) The Future Is Big Graphs! CACM.

https://graph-massivizer.eu/

6G FUTURE NETWORK SERVICES

Trustworthiness

Digital autonomy

Sustainable

Earning power

6G FUTURE NETWORK SERVICES

Includes a Digital Twin

2024—2030
€ 315 M
75+ partners

PL1: Intelligent components

TU/e Eindhoven University of Technology, NXP, AMPLEON, TU Delft, Ericsson, TNO innovation for life, CITC, ItoM, imec, ANTENNA COMPANY, ALTUM RF, Signify, vodafone, Ziggo, robin radar systems, University of Twente, ANTENNEX, kpn, CHALMERS University of Technology, VTEC Lasers & Sensors, ASTRON Netherlands Institute for Radio Astronomy, Viasat, sabic, AIRCISION empowering connectivity, PRODRIVE TECHNOLOGIES

PL2: Intelligent networks

TU Delft, TNO innovation for life, SURF, NVIDIA, NOKIA, TU/e Eindhoven University of Technology, Solvinity Secure Managed IT Services, Ericsson, kpn, Universiteit van Amsterdam, University of Twente, VU Vrije Universiteit Amsterdam, vodafone, Ziggo, amsix, Viasat, almende, is-wireless

PL3: Leading applications

Drone Delivery Services, PHILIPS, Vialis, Gomibo, CORDIS SUITE, T Mobile, Future Mobility Network, ASML, Ericsson, alliander, TNO innovation for life, Gemeente Amsterdam, Port of Rotterdam, COMFOREST, kpn, TU/e Eindhoven University of Technology, provincie groningen, provinsje fryslân provincie fryslân, TenneT, robin radar systems, provincie Drenthe, Schiphol, TU Delft, Gemeente Rotterdam, PWXR

PL4: Strengthen ecosystem

TU Delft, Do IoT Fieldlab, dutch digital delta, BTG, human capital agenda, University of Twente, ECP Platform voor de InformatieSamenleving, 5Groningen, Rijksinspectie Digitale Infrastructuur, SURF, kpn, Ministerie van Economische Zaken en Klimaat, ttt, T Mobile, NOKIA, TNO innovation for life, KORE, oost nl, LIBERTY GLOBAL, Ericsson, Hanzehogeschool Groningen University of Applied Sciences

©JOCHEM GALAMA

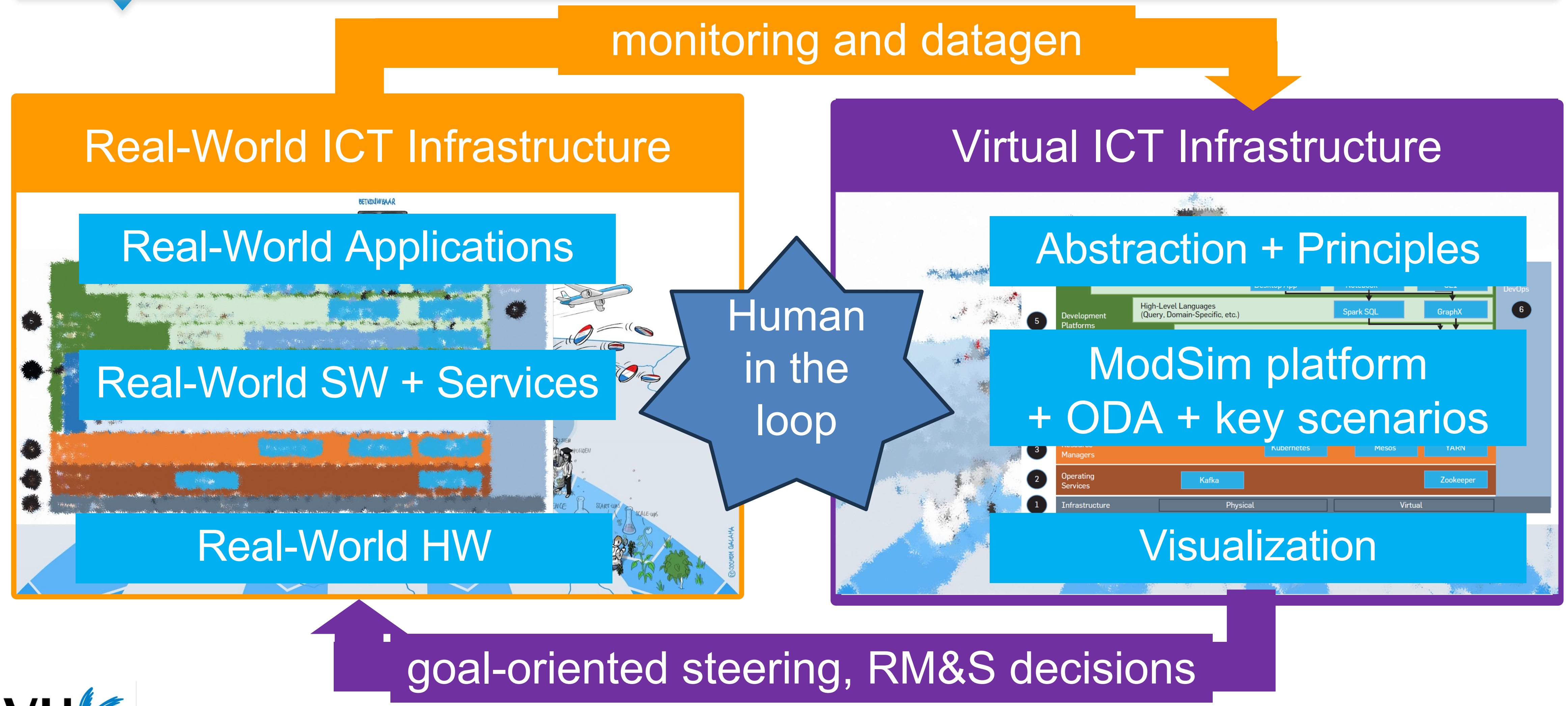
IN THIS TALK:

WE’RE BUILDING A DIGITIAL TWIN TO ADDRESS LONG-TERM NEEDS

3’

(What we need in CompSys infrastructure)

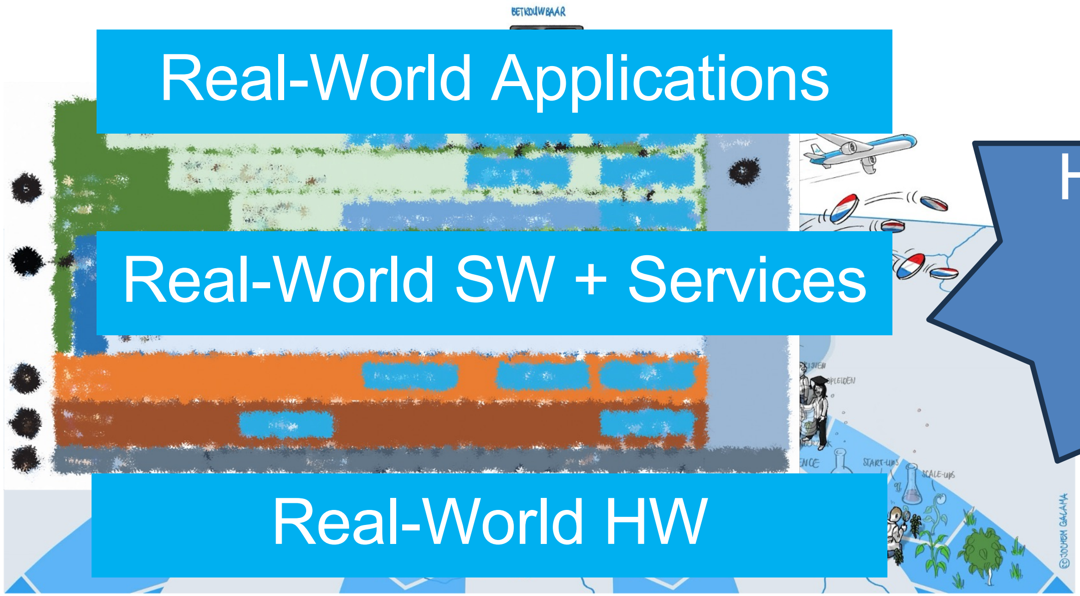
WHAT IS A DIGITAL TWIN FOR ICT INFRASTRUCTURE?



WHAT IS A DIGITAL TWIN FOR ICT INFRASTRUCTURE?

X1 monitoring and datagen

A Real-World ICT Infrastructure



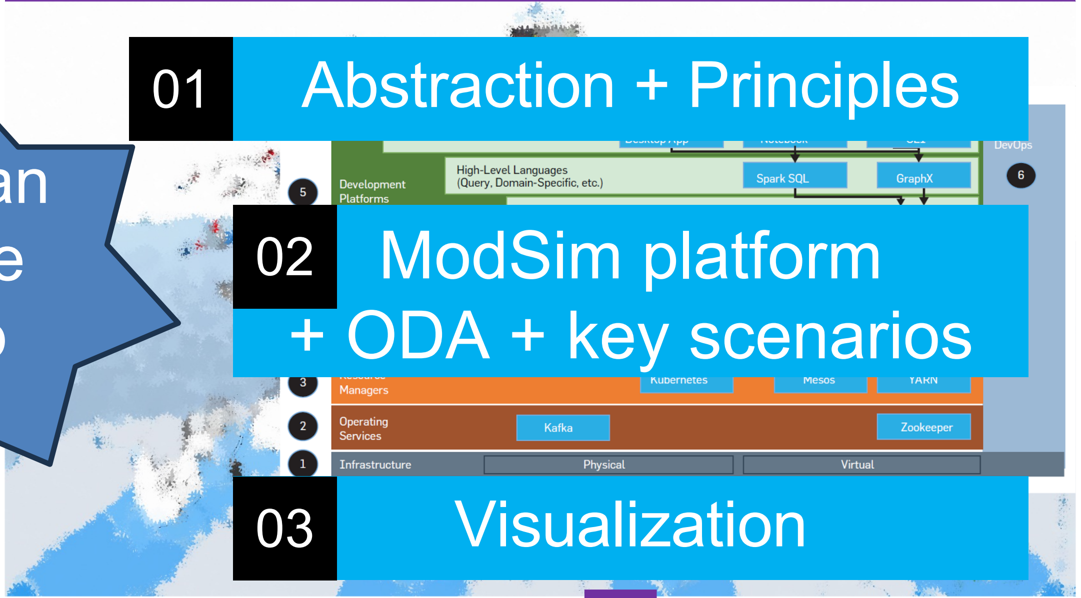
Real-World Applications

Real-World SW + Services

Real-World HW

Human in the loop

B Virtual ICT Infrastructure



01 Abstraction + Principles

02 ModSim platform + ODA + key scenarios

03 Visualization

goal-oriented steering, RM&S decisions X2

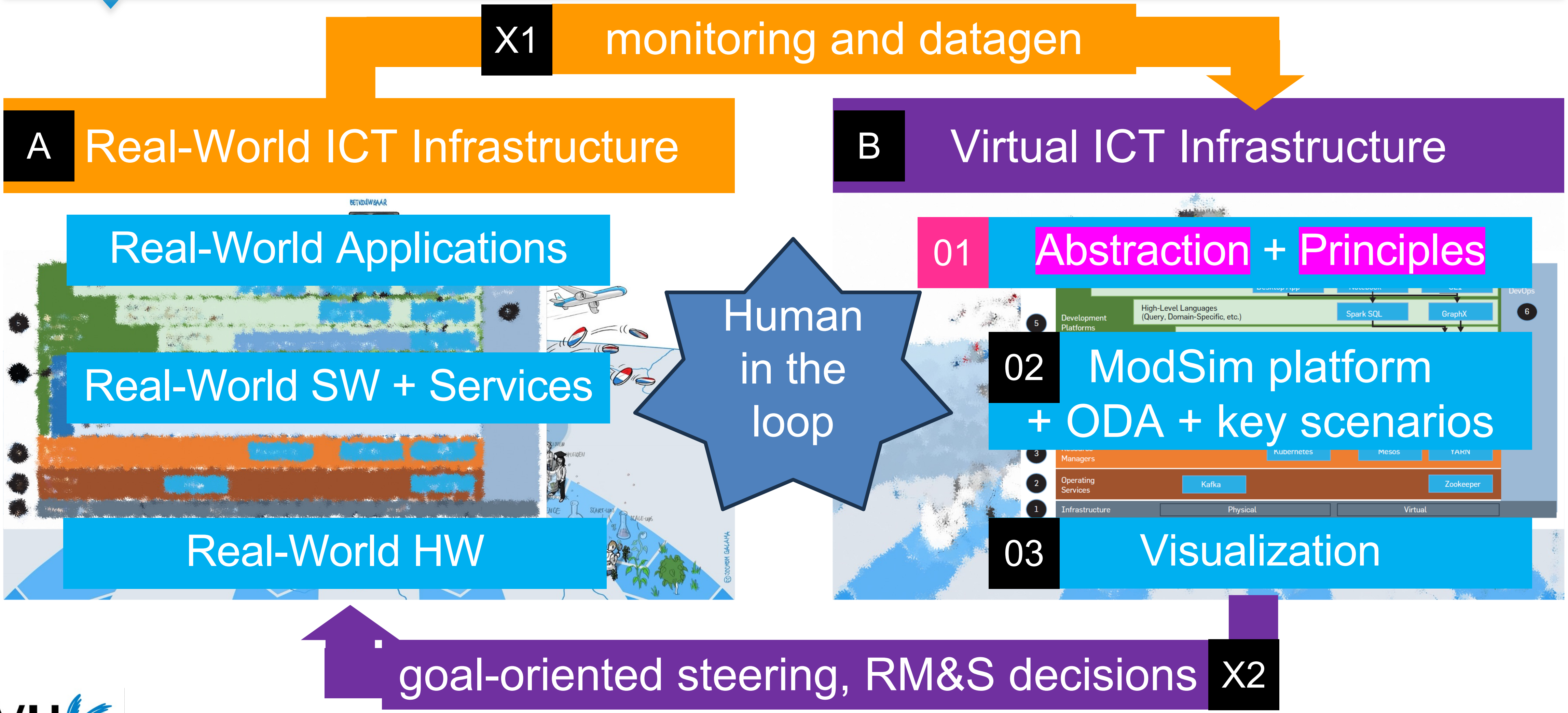
VU

AGENDA

IN THIS TALK:

1. Golden Age of Computer Ecosystems
2. But we cannot take this technology for granted
3. We can try to understand and improve things for targeted applications – with Digital Twins (DT)
4. We defined a high-level reference architecture for DT
5. Let’s detail the main components (not comprehensively, not necessarily in order, and time-allowing)
6. Let’s conclude

WHAT IS A DIGITAL TWIN FOR ICT INFRASTRUCTURE?

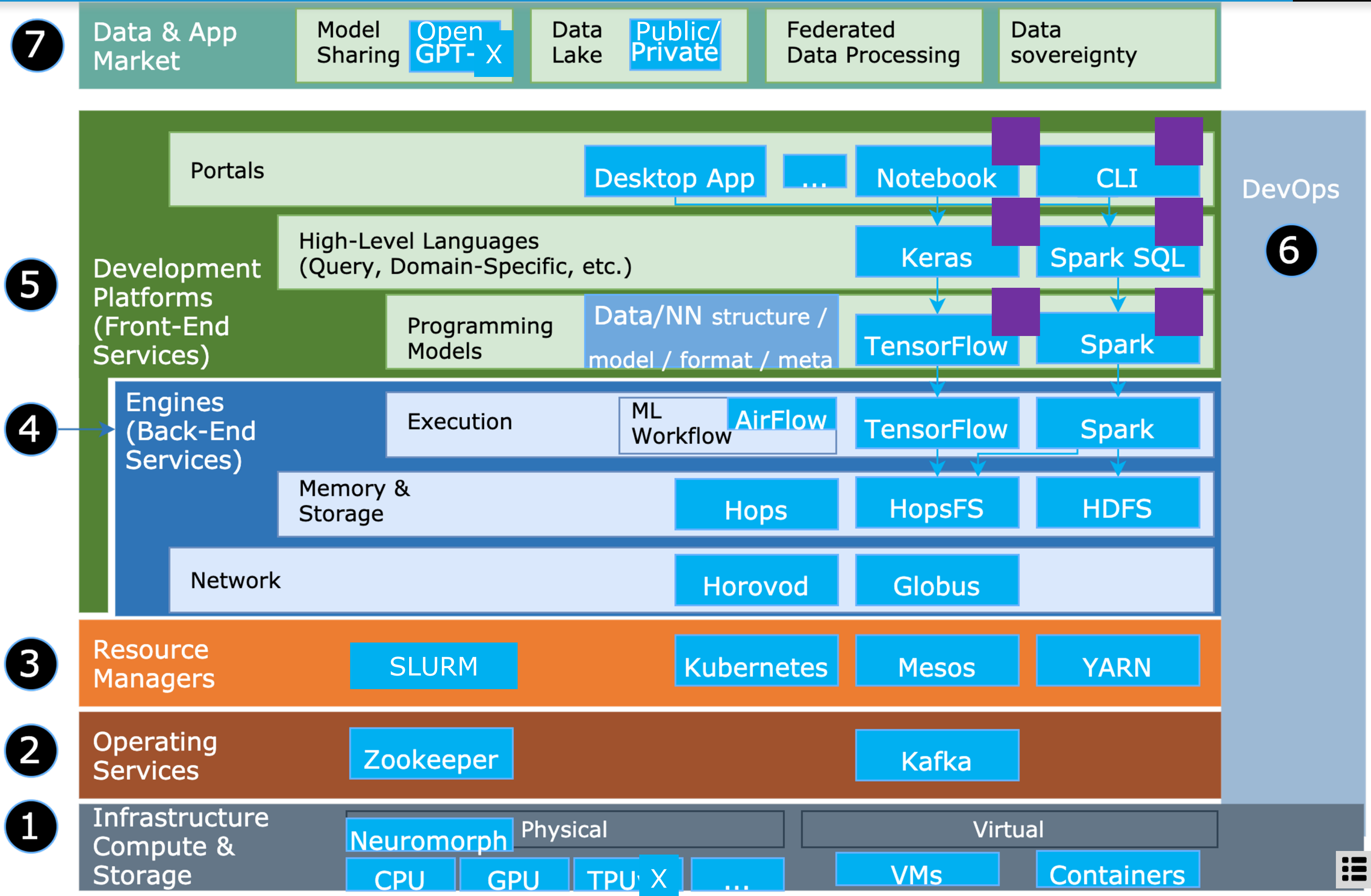


ONE SYSTEM MODEL: FITS AI/ML, BIG DATA, SCIENTIFIC, ENGINEERING, BUSINESS CRITICAL, ONLINE GAMING, OTHER APPS

B 01

ML app is a small part, but now we can do complex workflows

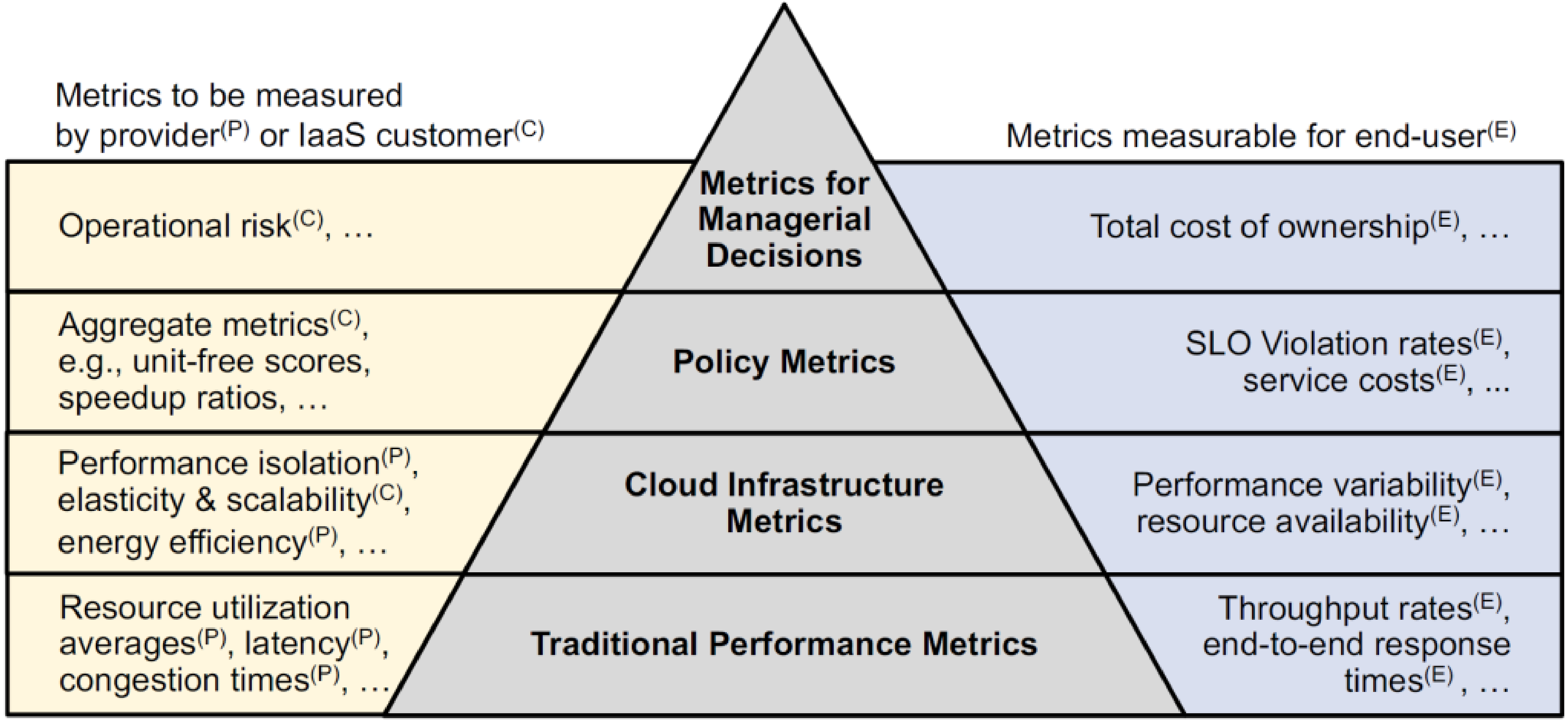
Rest is systems, HW+SW, including HPC



Adapted from:

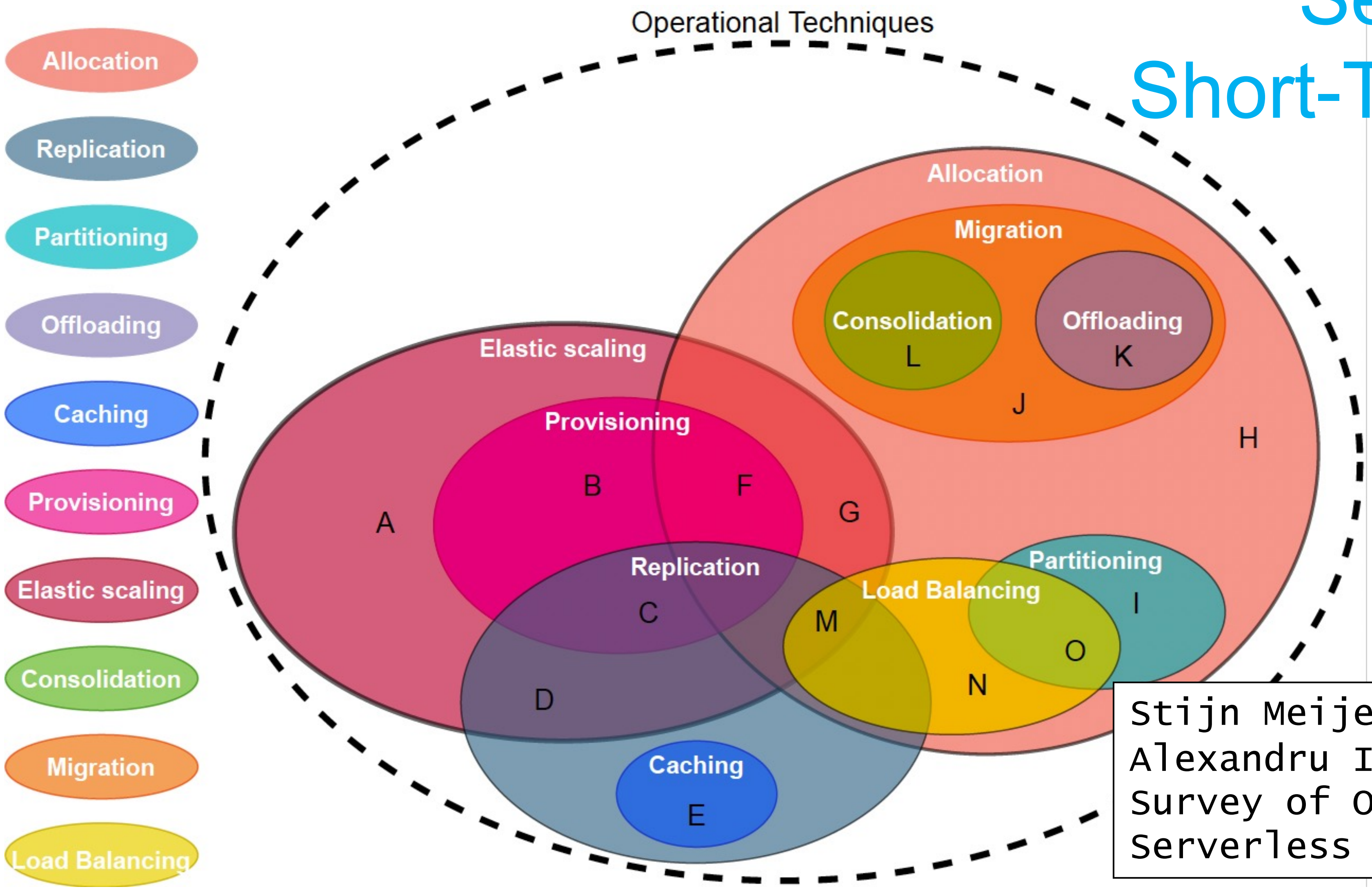
Sakr, Bonifati, Voigt, Iosup, et al. (2021) The Future Is Big Graphs! CACM.

Different Performance Goals for Different Stakeholders Framework for (New) Operational Metrics



N. Herbst, A. Bauer, S. Kounev, G. Oikonomou, E. Van Eyk, G. Kousiouris, A. Evangelinou, R. Krebs, T. Brecht, C. L. Abad, A. Iosup: Quantifying Cloud Performance and Dependability: Taxonomy, Metric Design, and Emerging Challenges. TOMPECS 3(4): 19:1-19:36 (2018)

Selected Dynamic, Short-Term Techniques



Stijn Meijerink, Erwin van Eyk, Alexandru Iosup (2021) Multivocal Survey of Operational Techniques for Serverless Computing. White Paper.

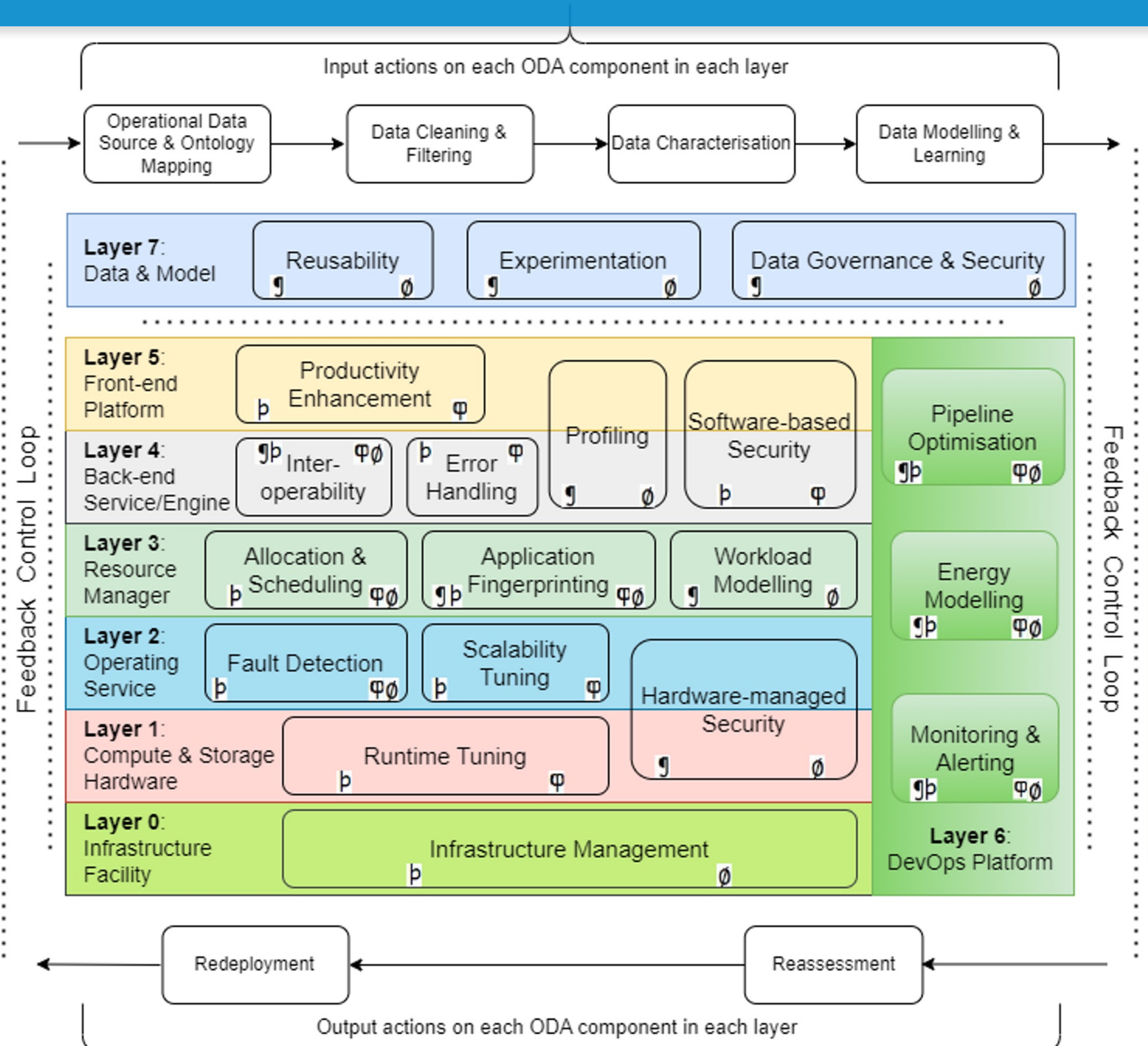
Recap →

ENABLE ODA (OPERATIONAL PRINCIPLE)

B
01

Problem:
Data often seen as overhead, loose process

Solution:
1. Link data collection to per-layer capabilities
2. Enable data science process



Legend

Data Sources:
¶: On-demand
þ: Online

ODA mode of operation:
φ: In-band
Ø: Out-of-band

Ongoing work

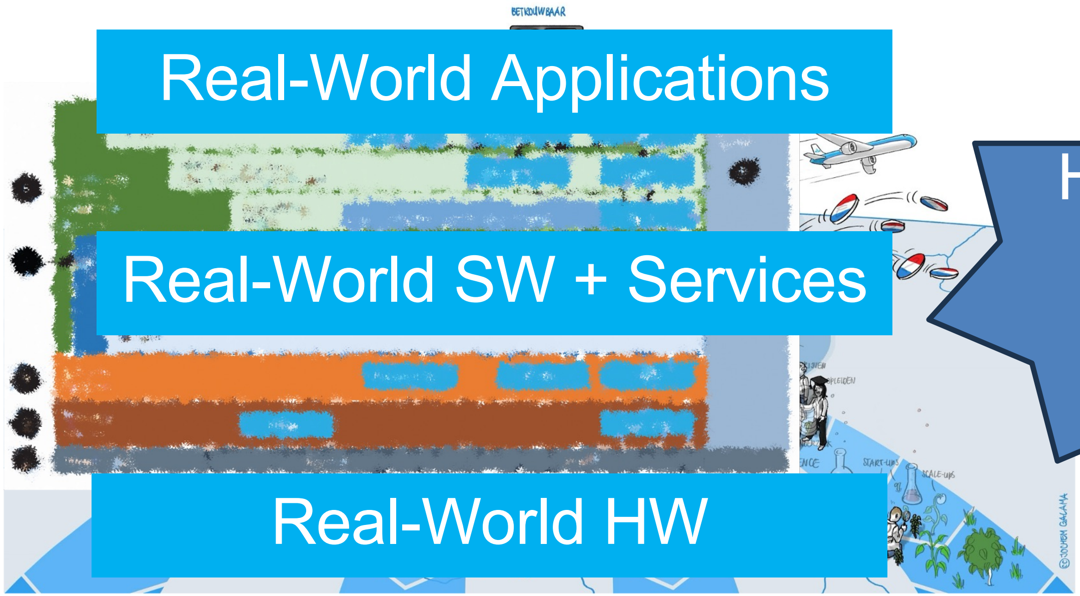
Shekhar Suman, Xiaoyu Chu, Martin Molan, Andrea Bartolini (UniBo), Iosup, et al. (2023) Ontology for HPC Infrastructure Data

WHAT IS A DIGITAL TWIN FOR ICT INFRASTRUCTURE?

X1 monitoring and datagen

A Real-World ICT Infrastructure

B Virtual ICT Infrastructure

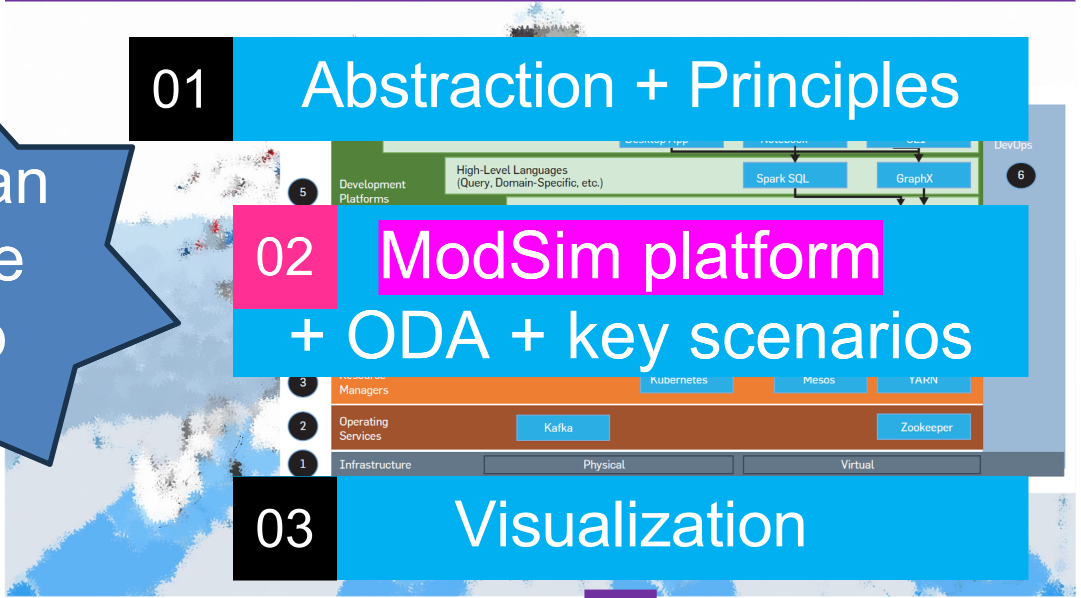


Real-World Applications

Real-World SW + Services

Real-World HW

Human in the loop



01 Abstraction + Principles

02 ModSim platform + ODA + key scenarios

03 Visualization

goal-oriented steering, RM&S decisions X2

OpenDC: SIMULATION PLATFORM FOR ICT INFRASTRUCTURE

… SHORT-TERM: X or Y? WHERE TO PUT Z? LONG-TERM: CAN WE AFFORD A? WHAT IF B HAPPENS?

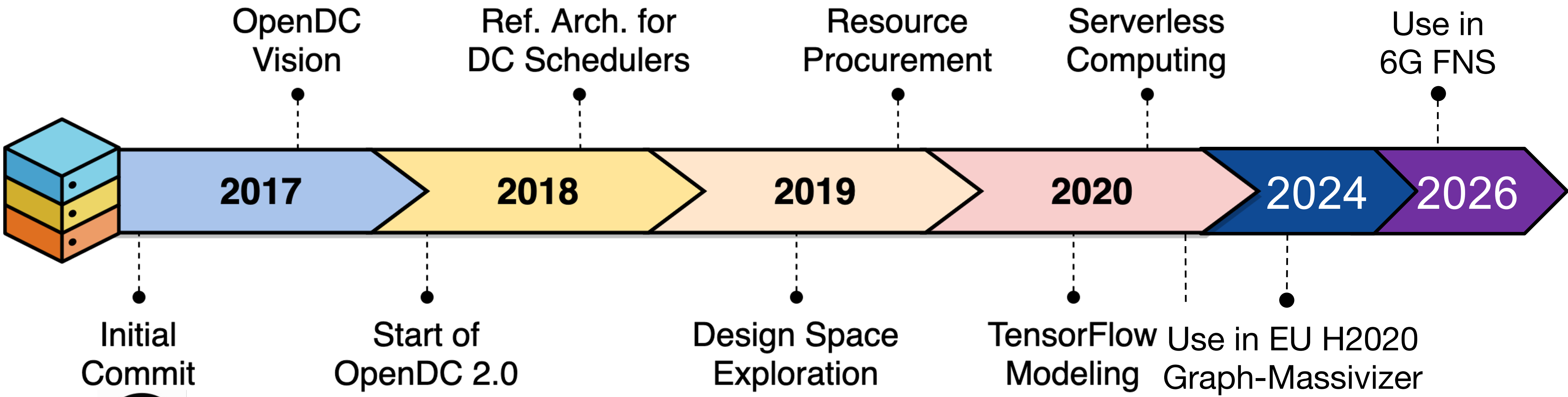
[OpenDC](opendc.org) simulator

- Short-term resource management
- Long-term decision making
- Sophisticated model → many Qs, goals
- Supports many kinds of workloads
- Supports many kinds of resources
- Validated for various scenarios
- Work with major NL operators, incl. HPC
- Used in training, education, research

and more…

Learn more: opendc.org

OpenDC 2.0



https://github.com/atlarge-research/opendc

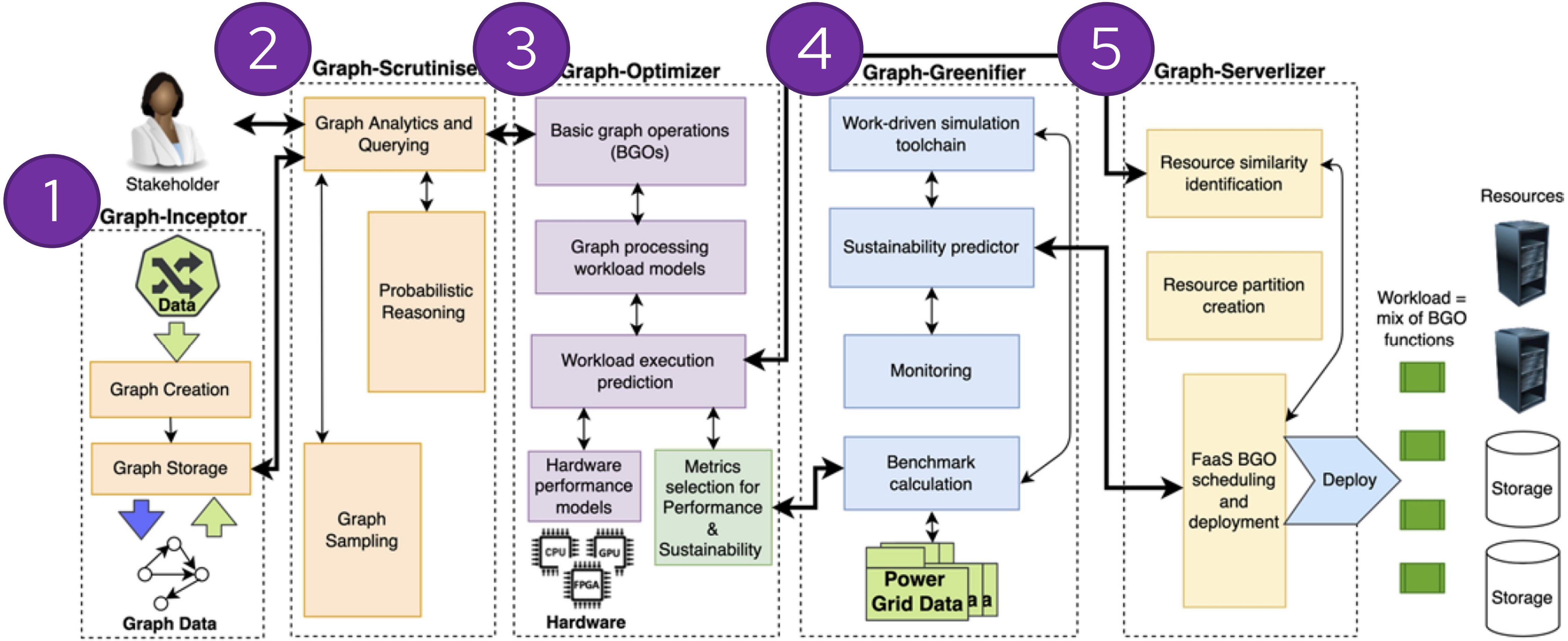
Learn more:
opendc.org

Scan me for
OpenDC paper

Fabian Mastenbroek, Georgios Andreadis, Soufiane Jounaid, Wenchen Lai, Jacob Burley, Jaro Bosch, Erwin Van Eyk, Laurens Versluis, Vincent van Beek, Alexandru Iosup (2021) OpenDC 2.0: Convenient Modeling and Simulation of Emerging Technologies in Cloud Datacenters. CCGRID 2021: 455-464. [Online]

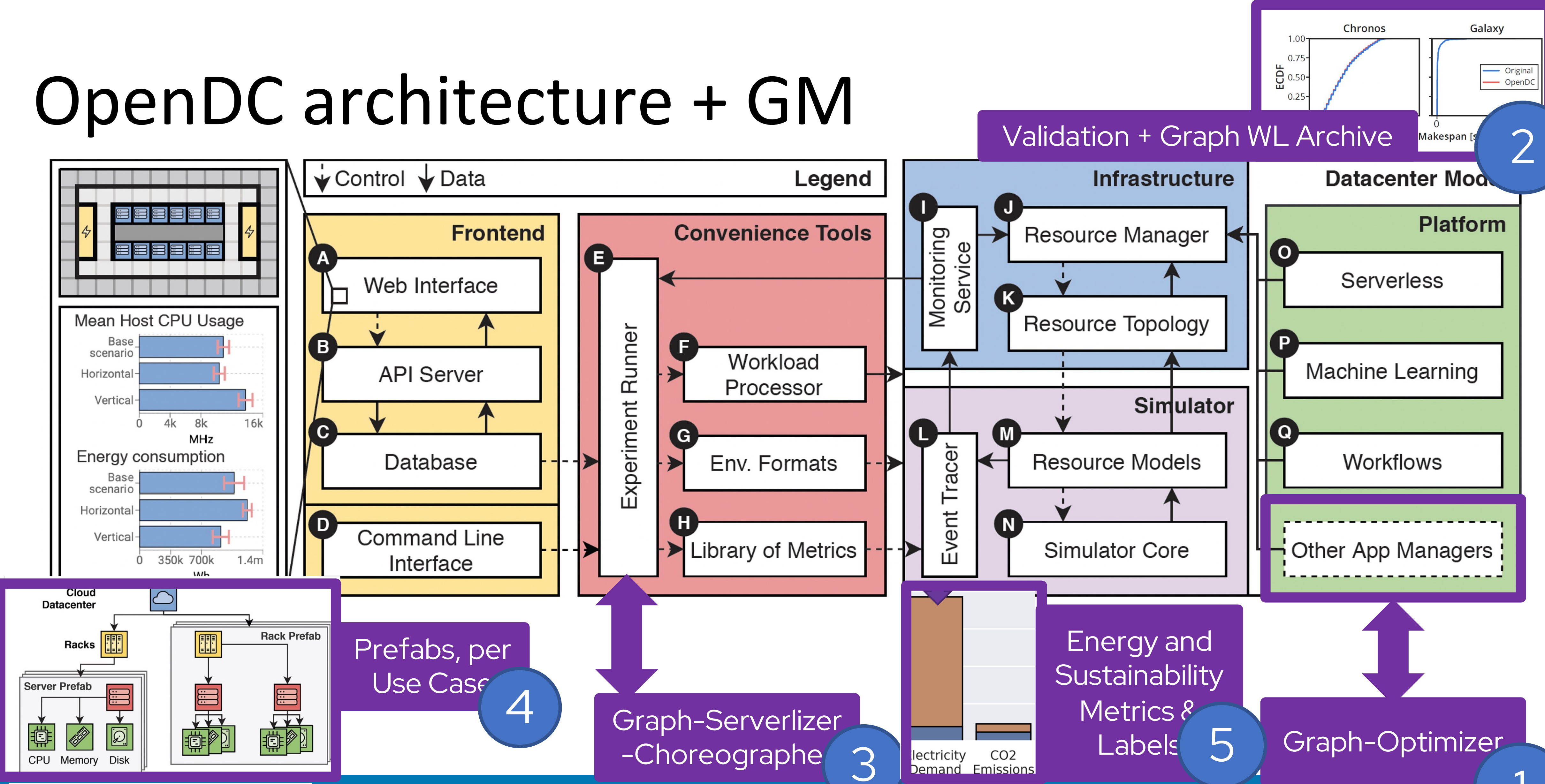
Graph-Massivizer: Technical Architecture

B
02



Prodan, Iosup, Varbanescu, et al. (2022) Towards Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe, IEEE Cloud Summit, 23-30.

OpenDC architecture + GM



Mastenbroek, Andreadis, Iosup, et al. (2021) OpenDC 2.0: Convenient Modeling and Simulation of Emerging Technologies in Cloud Datacenters. CCGRID.

CAPELIN: FAST DATA-DRIVEN DC CAPACITY PLANINNG

Capelin: Fast Data-Driven Capacity Planning for Cloud Datacenters

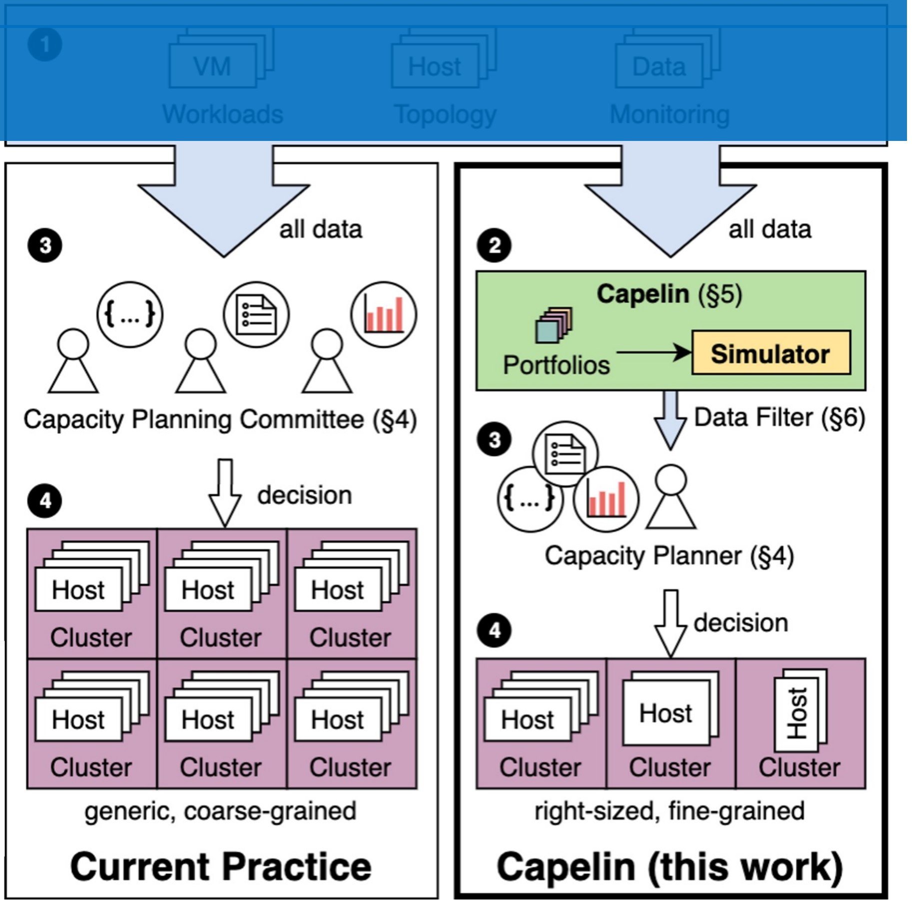
B
02

WE SURVEYED AND ANALYZED 89 USE-CASES

- High-impact problem
- Long-term effects
- Difficult to predict

Our work: Capelin + OpenDC

- Sophisticated approach
- Based on validated simulator
- Work with major NL hoster



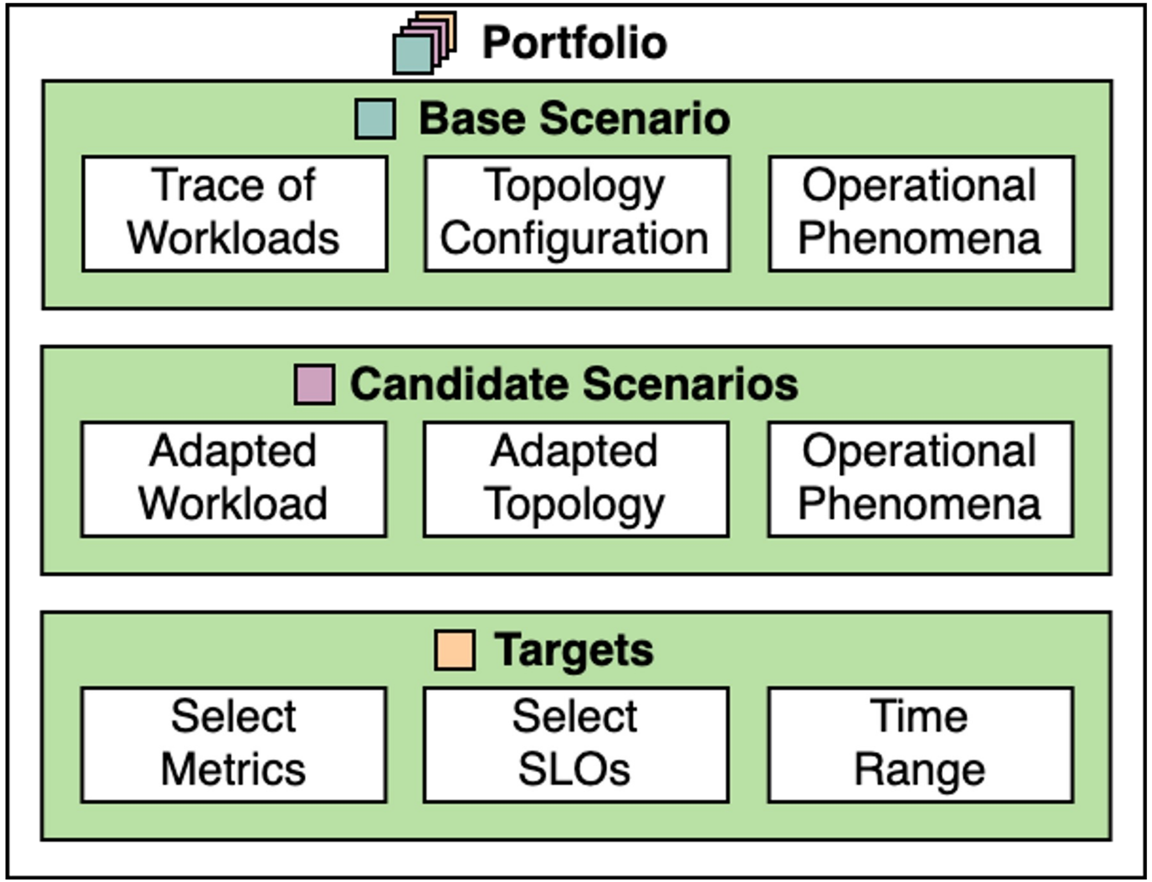
A New Abstraction: Portfolios of Scenarios

Novel features:

- Operational phenomena
- Multi-scenario

Evolution of design:

- One mandatory base scenario
- Targets on portfolio level



Slides by Georgios Andreadis,
with input from Alexandru Iosup et al.

Andreadis et al. (2022) Capelin: Data-Driven Capacity Procurement for Cloud Datacenters using Portfolios of Scenarios. IEEE TPDS 33(1).

A Multi-Metric View of Datacenter Operations

Requested CPU cycles	Failed VM time slices	Total submitted VMs
Granted CPU cycles	Mean CPU usage	Max. num. queued VMs
Overcommitted CPU cycles	Mean CPU demand	Total finished VMs
Interfered CPU cycles	Mean num. VM per host	Total failed VMs
Power consumption	Max. num. VM. per host	

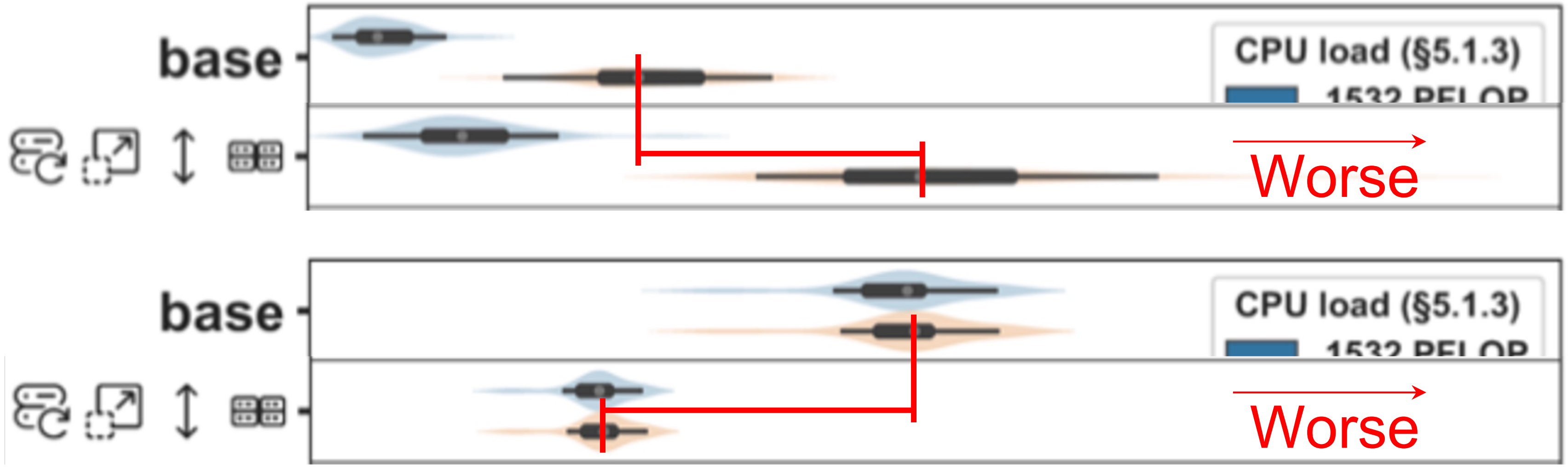
Slides by Georgios Andreadis,
with input from Alexandru Iosup et al.

Andreadis et al. (2022) Capelin: Data-Driven Capacity Procurement for Cloud Datacenters using Portfolios of Scenarios. IEEE TPDS 33(1).

Example: Horizontal vs. Vertical Scaling

CPU Overcomm.

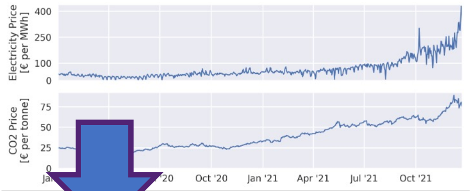
Power consumption

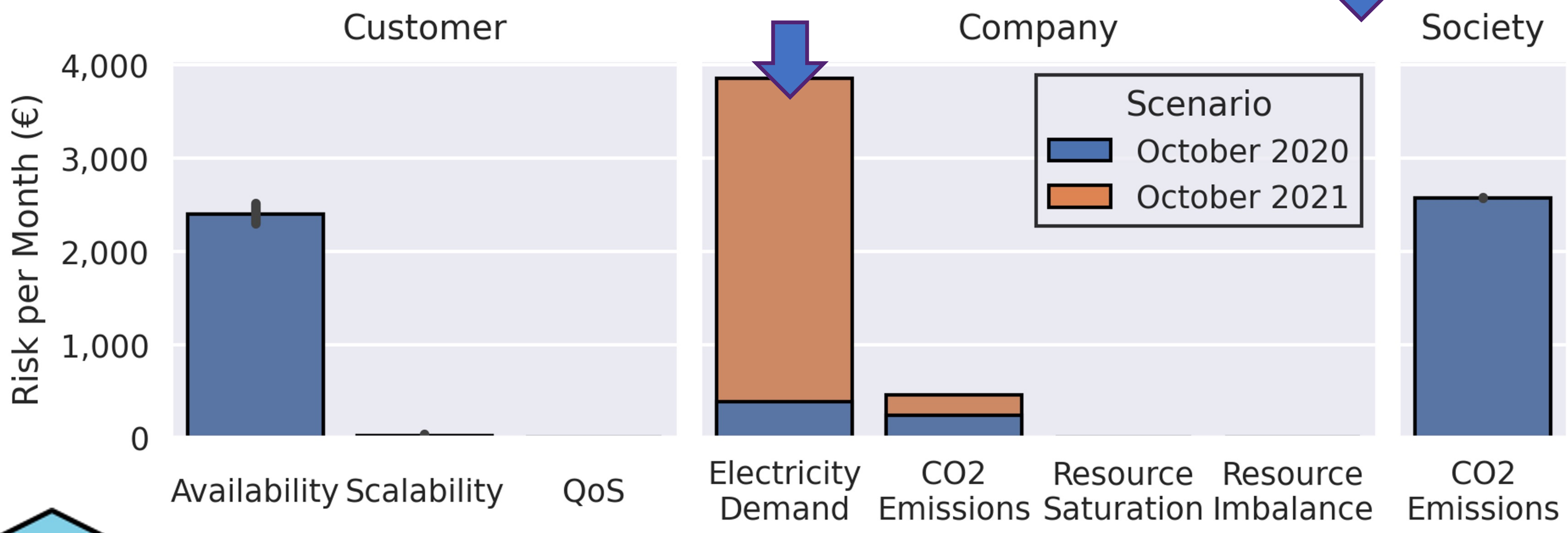


Capelin enables complex **trade-off exploration**, e.g., **power** vs. **performance**

Slides by Georgios Andreadis,
with input from Alexandru Iosup et al.

Example: Evolution of Electricity Expenses





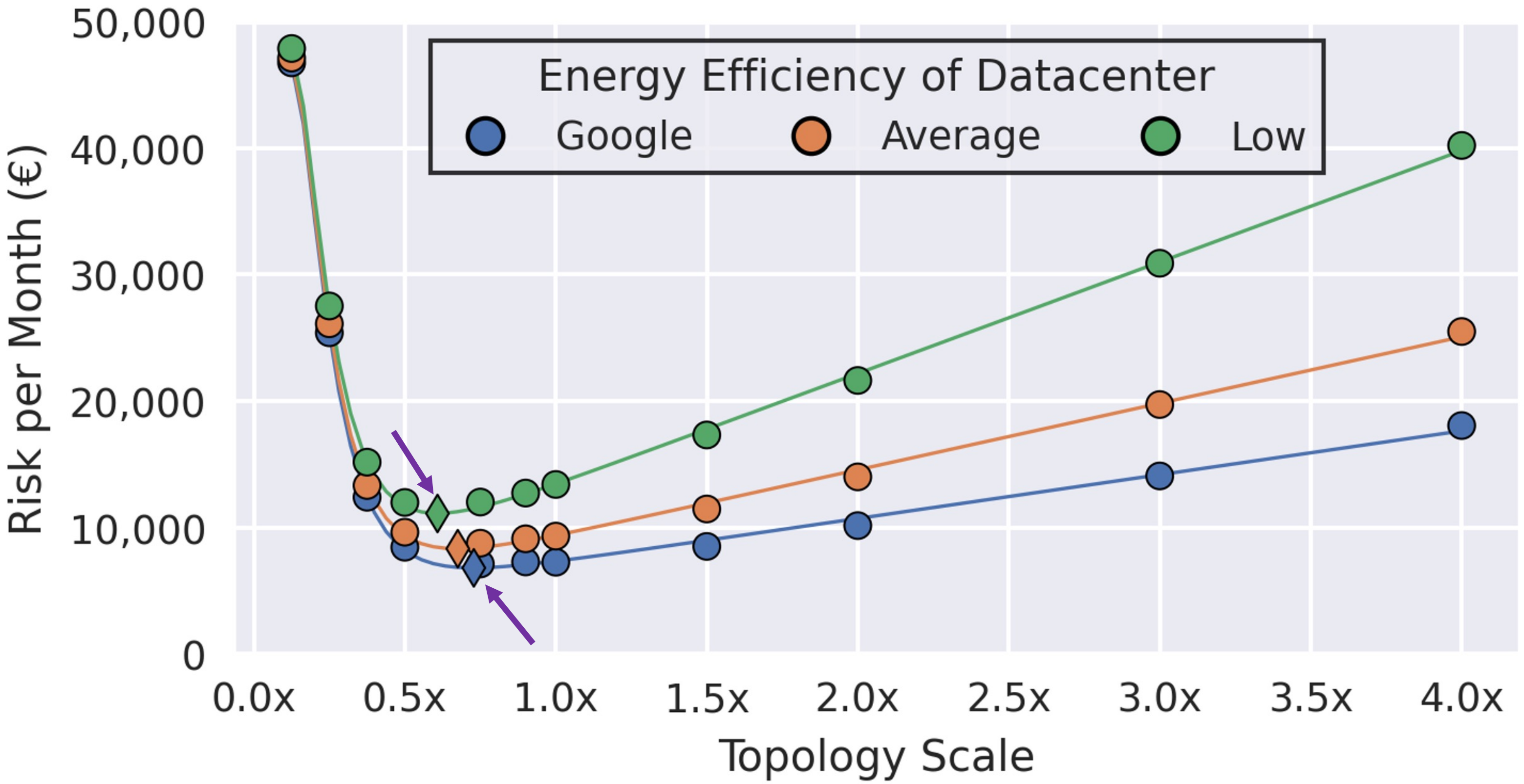
Electricity expenses become **primary risk** in datacenters

Downscaling to save electricity?

1. **Downscaling** results in lower risk amid soaring electricity prices

   but

2. Datacenter operators must **compromise** between lower **topology scale** and higher **SLA compliance**



Experiments are very expensive!
An Environmental Perspective

112

Core-hours

2.1

kWh

Delft →  → Rotterdam

In Simulation

8,600,000,000

Core-hours

116,000,000,000

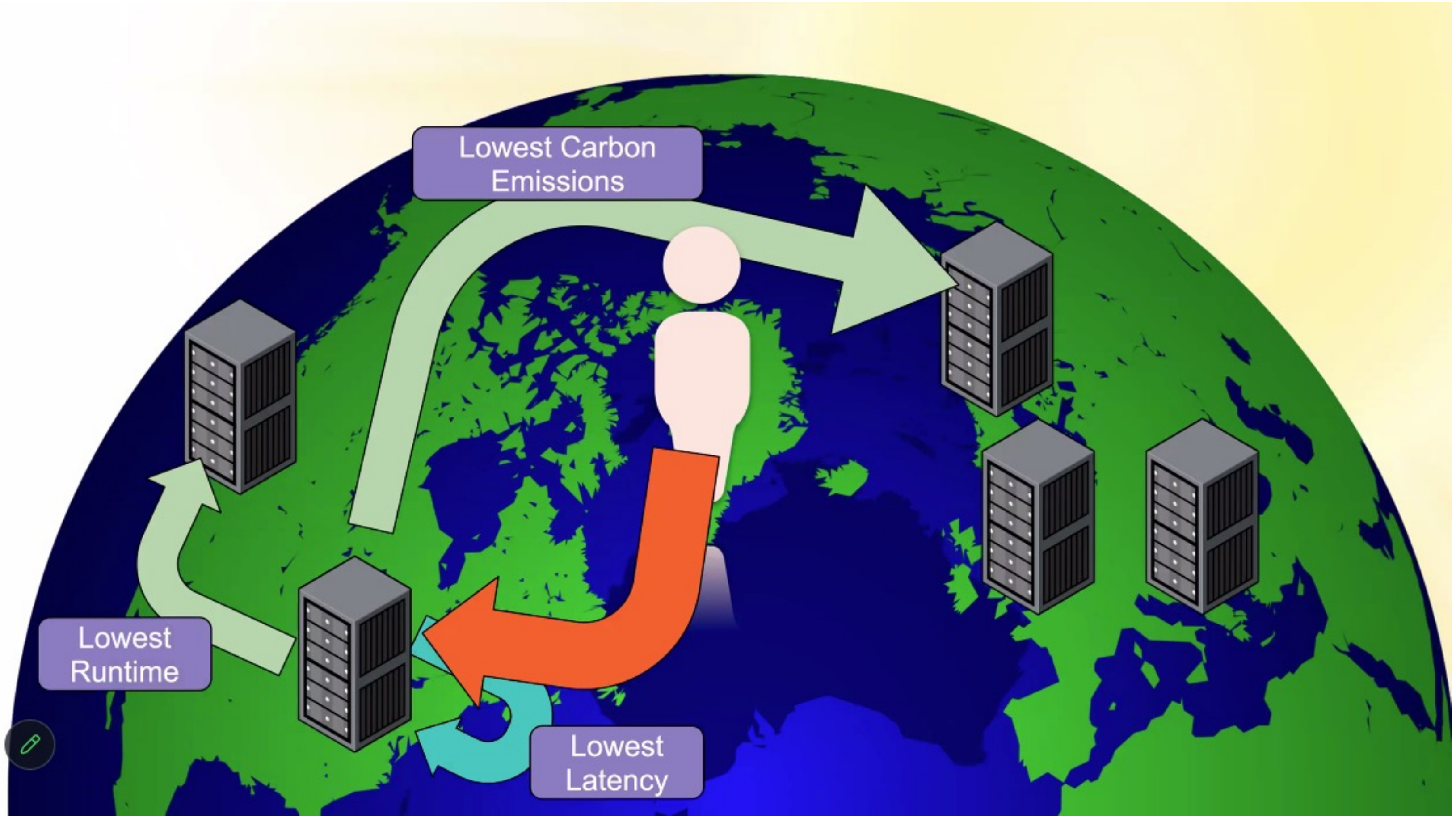
kWh

x 55 billion!

In Reality

Exact numbers confidential, depend on topology

Different Performance Goals for Different Stakeholders Practical Example



Robert Cordingly, Jasleen Kaur, Divyansh Dwivedi, Wes Lloyd (2023) Towards Serverless Sky Computing: An Investigation on Global Workload Distribution to Mitigate Carbon Intensity, Network Latency, and Cost. IC2E 2023: 59-69

Slide by Robert Cordingly, presented to SPEC RG Cloud on Oct 31, 2023

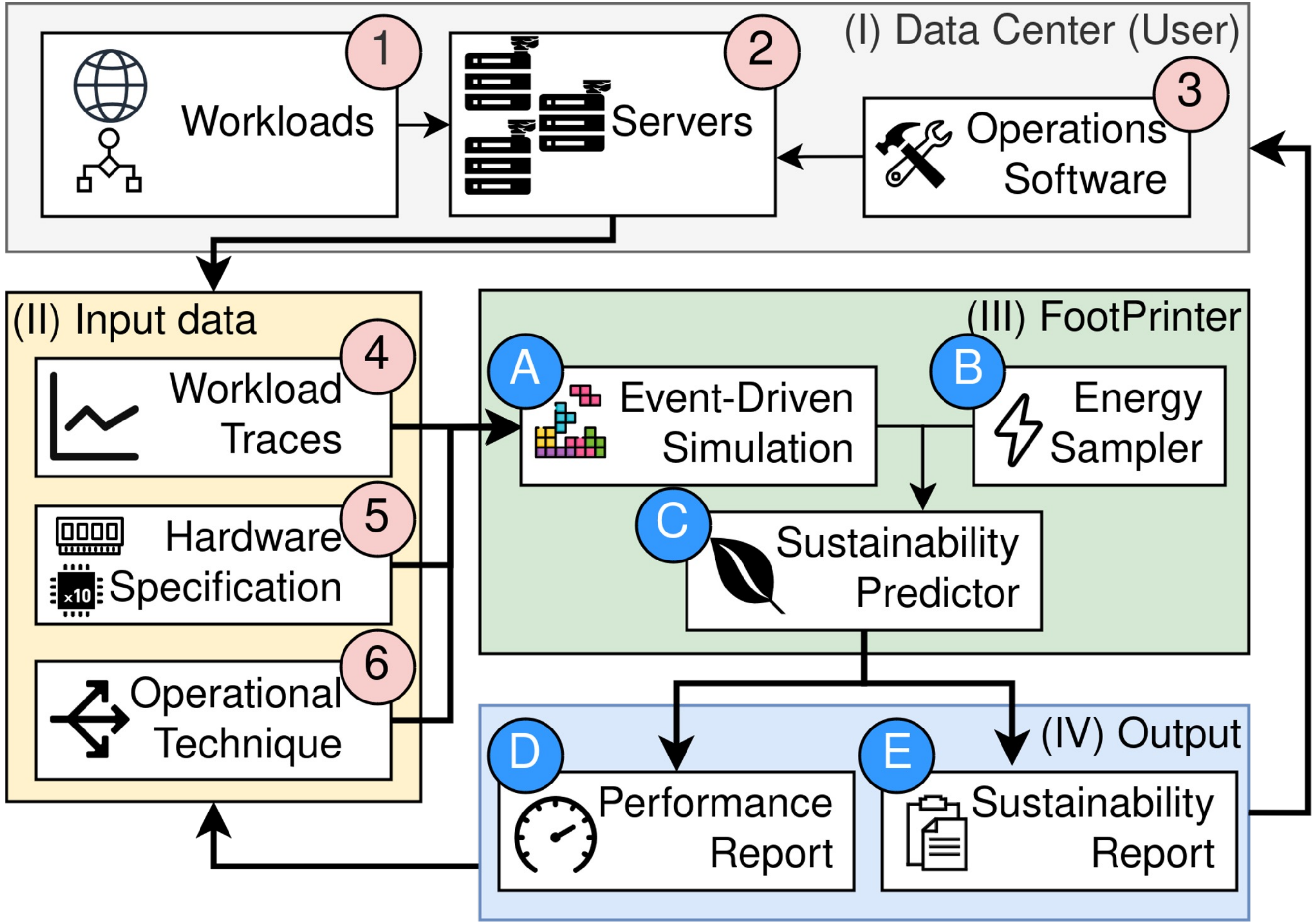
FootPrinter: Quantifying the Carbon Footprint

(I) Gather data

(II) Convert to input data

(III) Run FootPrinter

(IV) Analyze performance and sustainability reports & act



Dante Niewenhuis, Sacheendra Talluri, Alexandru Iosup, Tiziano De Matteis (2024) FootPrinter: Quantifying Data Center Carbon Footprint. HotCloudPerf.

Use Case 1: Evaluate carbon footprint

B

02

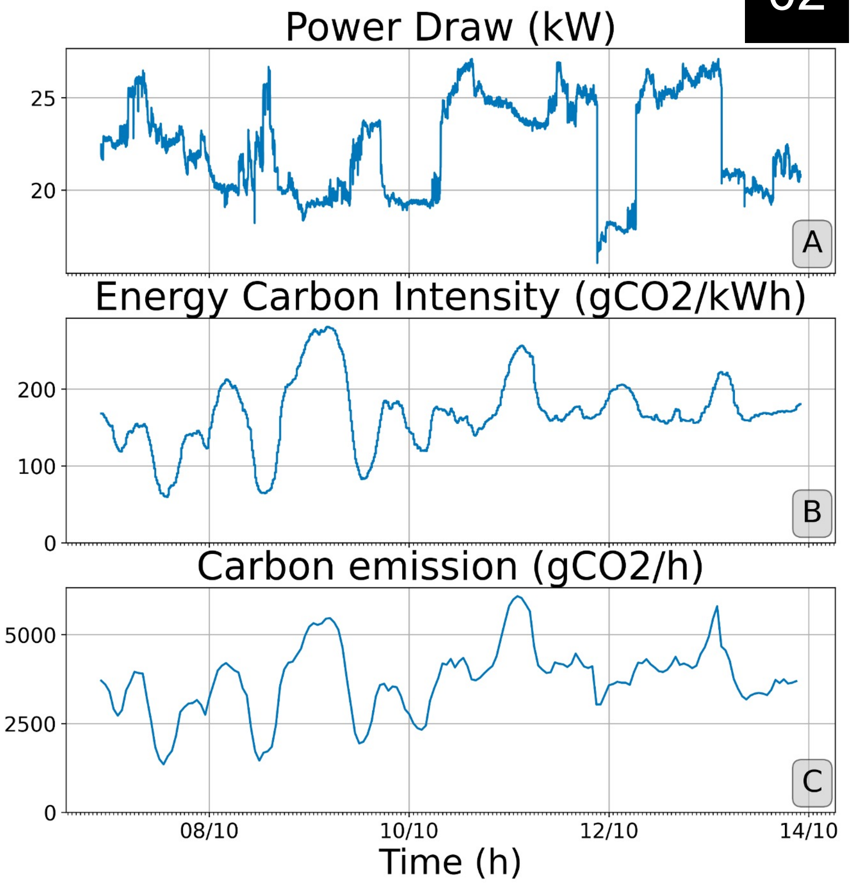
The carbon footprint is determined:

A. Determine power draw

B. Collect the carbon intensity (e.g., ENTSO-E)

C. Calculate carbon emissions

Carbon emission is primarily influenced by carbon intensity

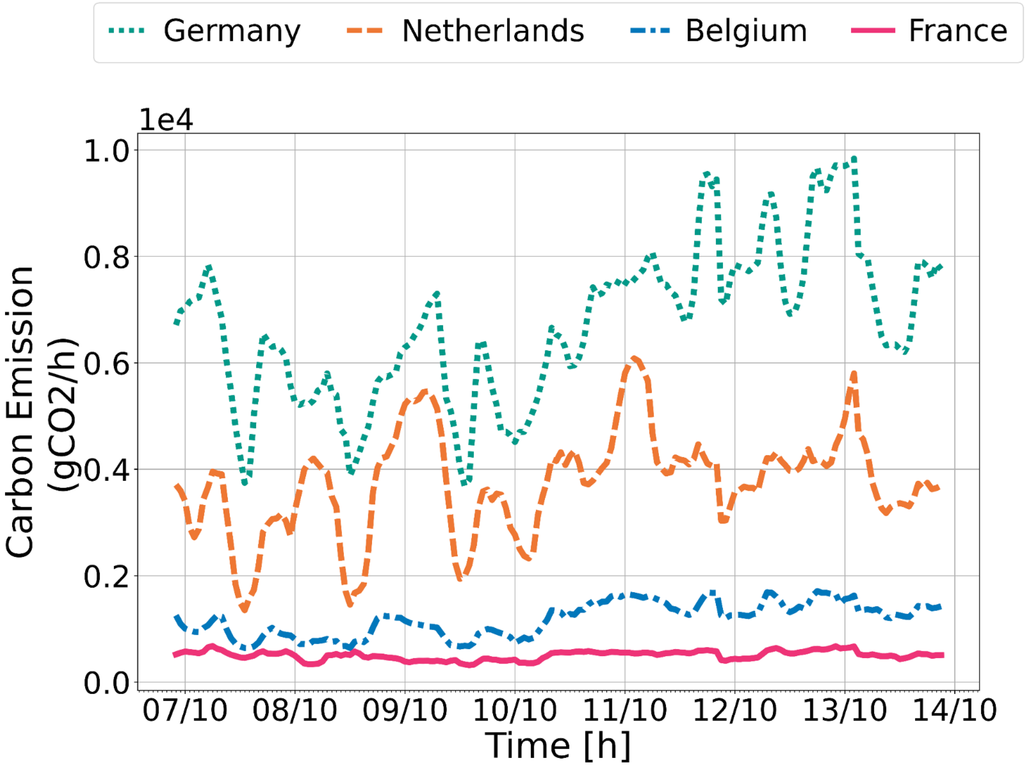


The Carbon emissions of a data center for a specific workload and RM&S policy

Use Case 2: Compare location

- Run the same workload
- Change data center location
- Compare carbon footprint

The location of a data center has significant effect on its carbon footprint



The Carbon emissions during the same workload on different locations

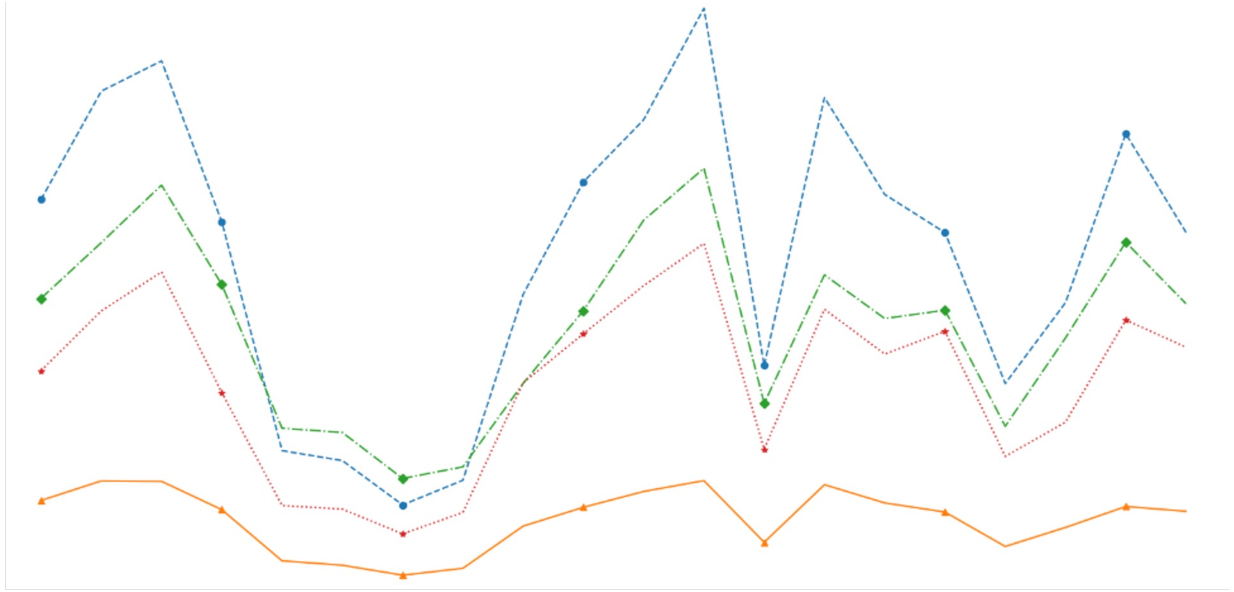
Dante Niewenhuis, Sacheendra Talluri, Alexandru Iosup, Tiziano De Matteis (2024) FootPrinter: Quantifying Data Center Carbon Footprint. HotCloudPerf.

Multi-Model Simulation and Analysis for Datacenters

B
02

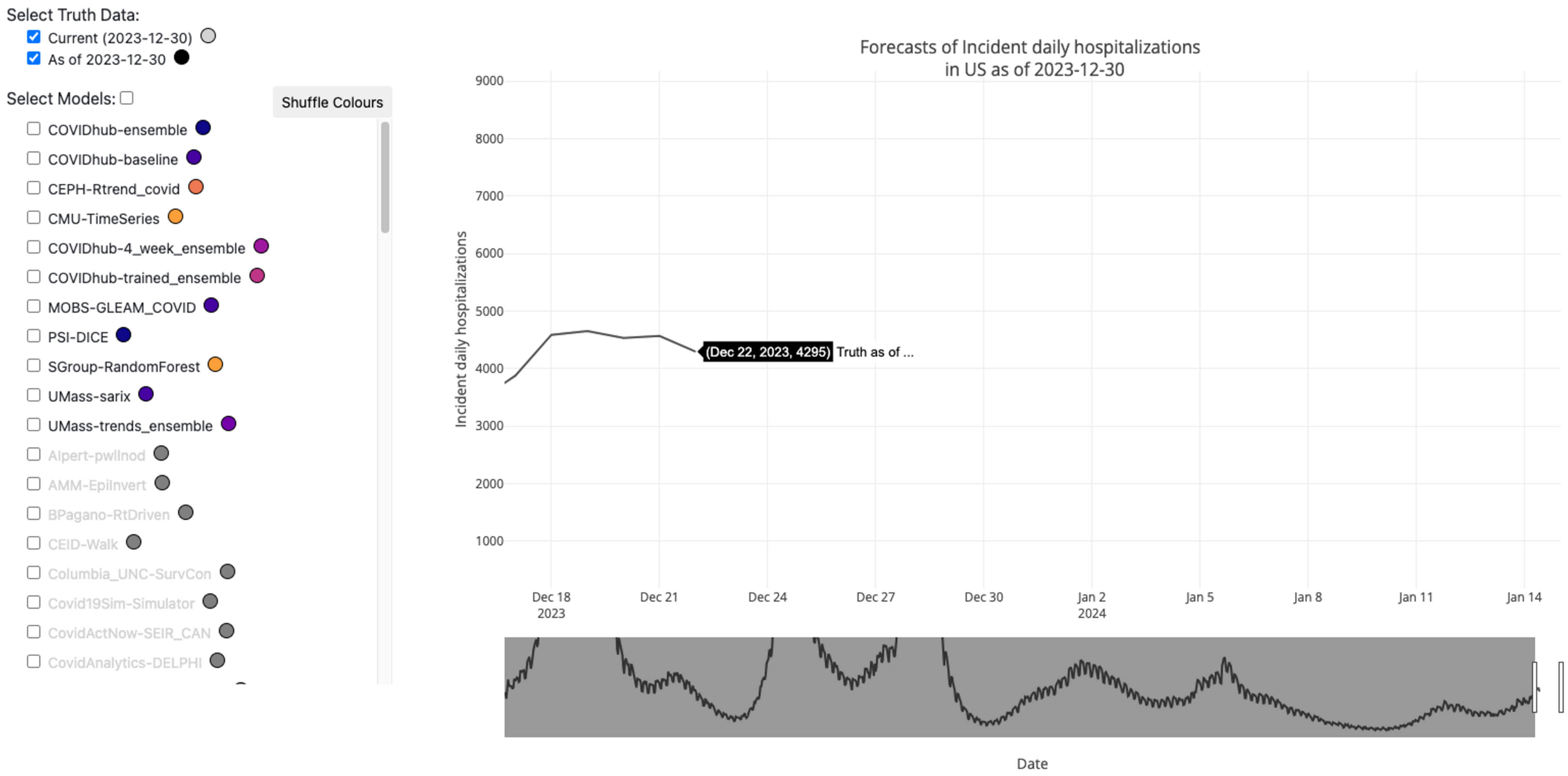
Radu Nicolae
@VU Amsterdam
@Large Research

in @rnicolae, mail@radu-nicolae.com



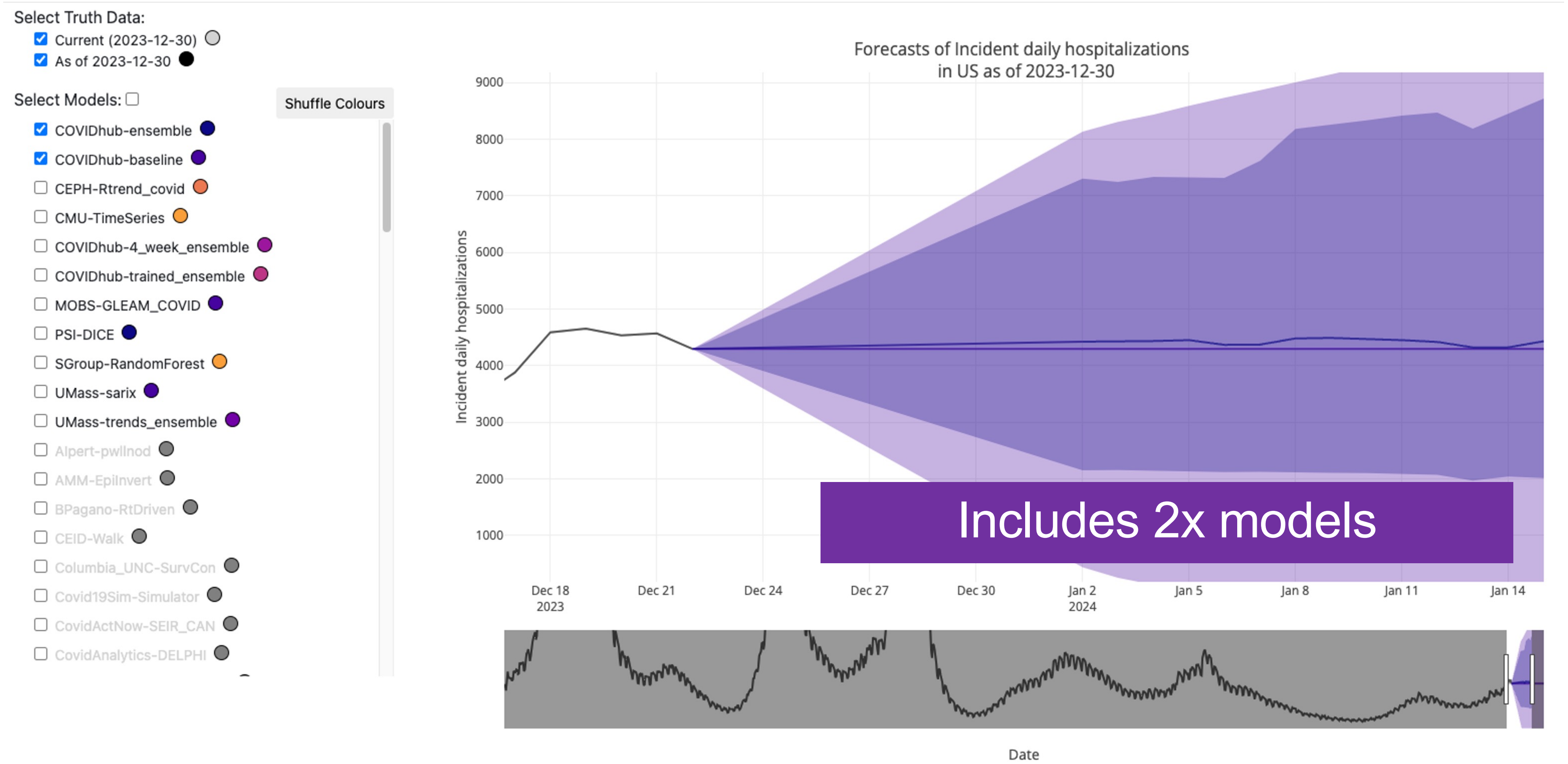
Honours Programme, Jan 24th, 2024

Multi-Model in Virology - Covid19ForecastHub



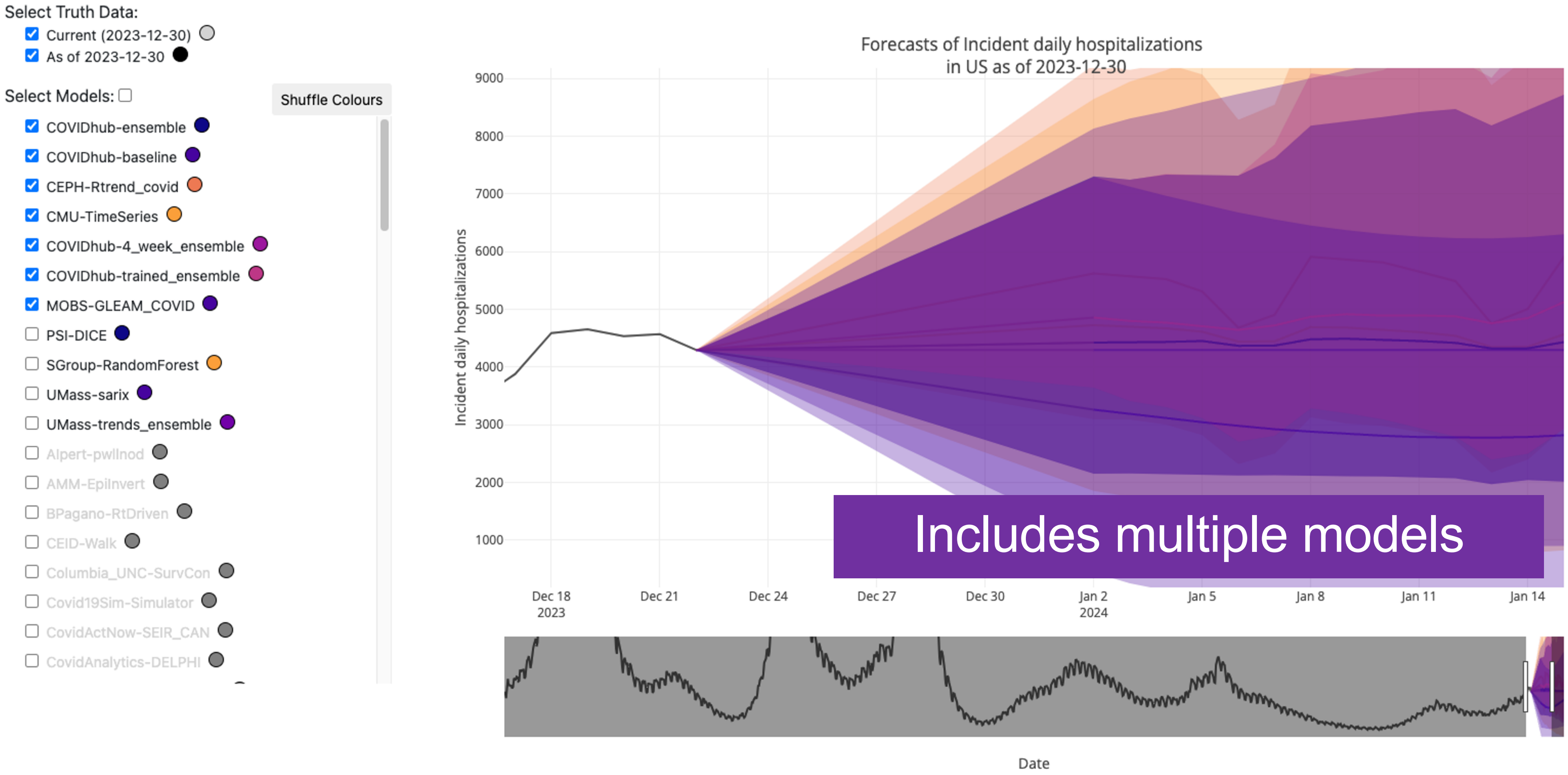
Source: COVID-19 Forecast Hub

Multi-Model in Virology - Covid19ForecastHub



Source: COVID-19 Forecast Hub

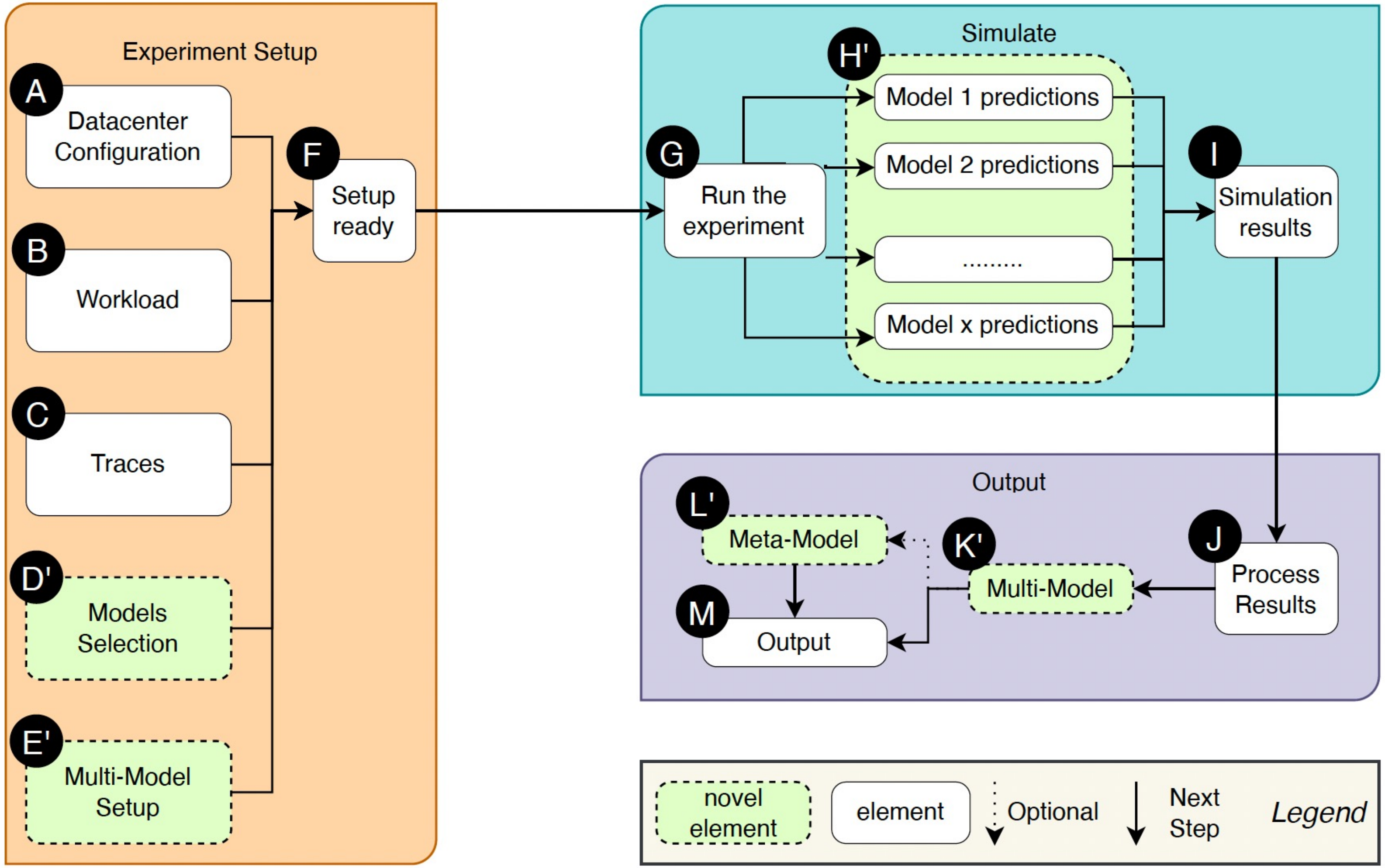
Multi-Model in Virology - Covid19ForecastHub



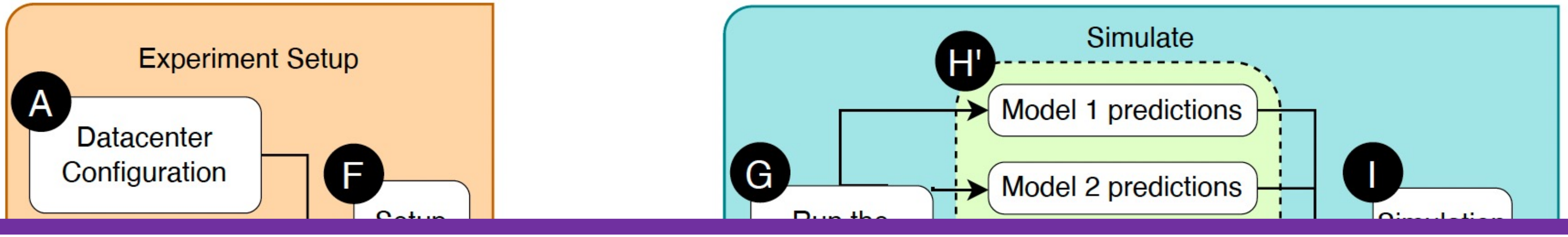
Includes multiple models

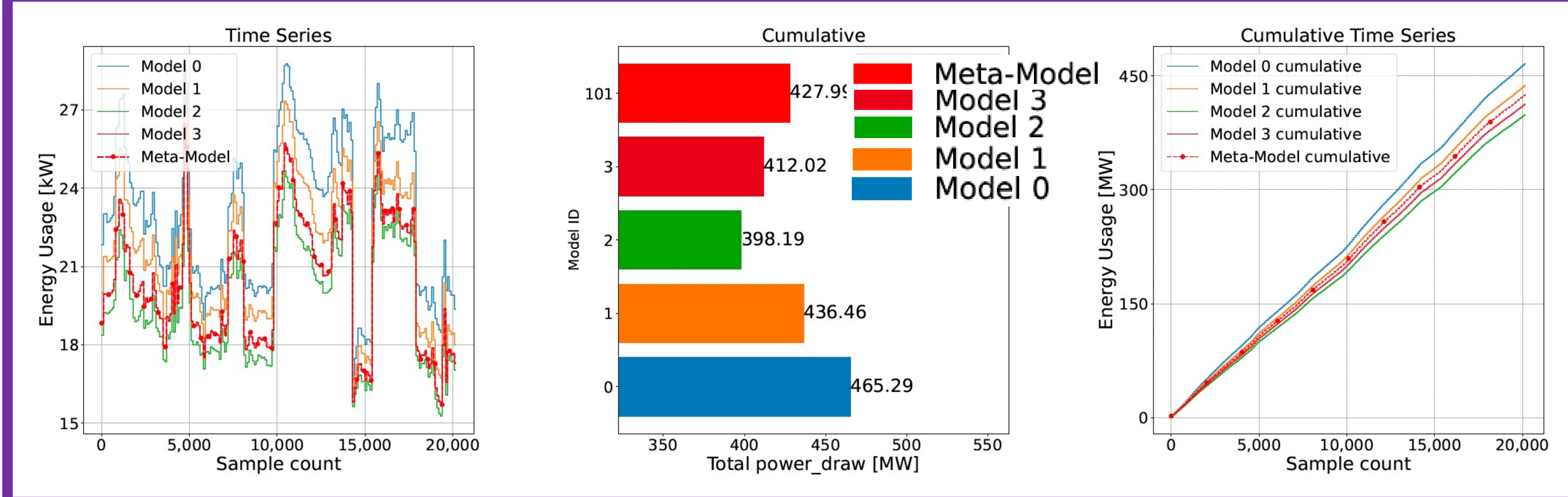
Source: COVID-19 Forecast Hub

The OpenDC Metamodel Vision



The OpenDC Metamodel – Example





B

02

ExDe: Design Space Exploration of Scheduler Architectures and Mechanisms for Serverless Data-processing

Sacheendra Talluri[1], Nikolas Herbst[2], Cristina Abad[3], Tiziano De Matteis[1], Alexandru Iosup[1]

[1]VU Amsterdam, [2]Würzburg University, [3]ESPOL

https://atlarge-research.com/pdfs/exde_fgcs23.pdf

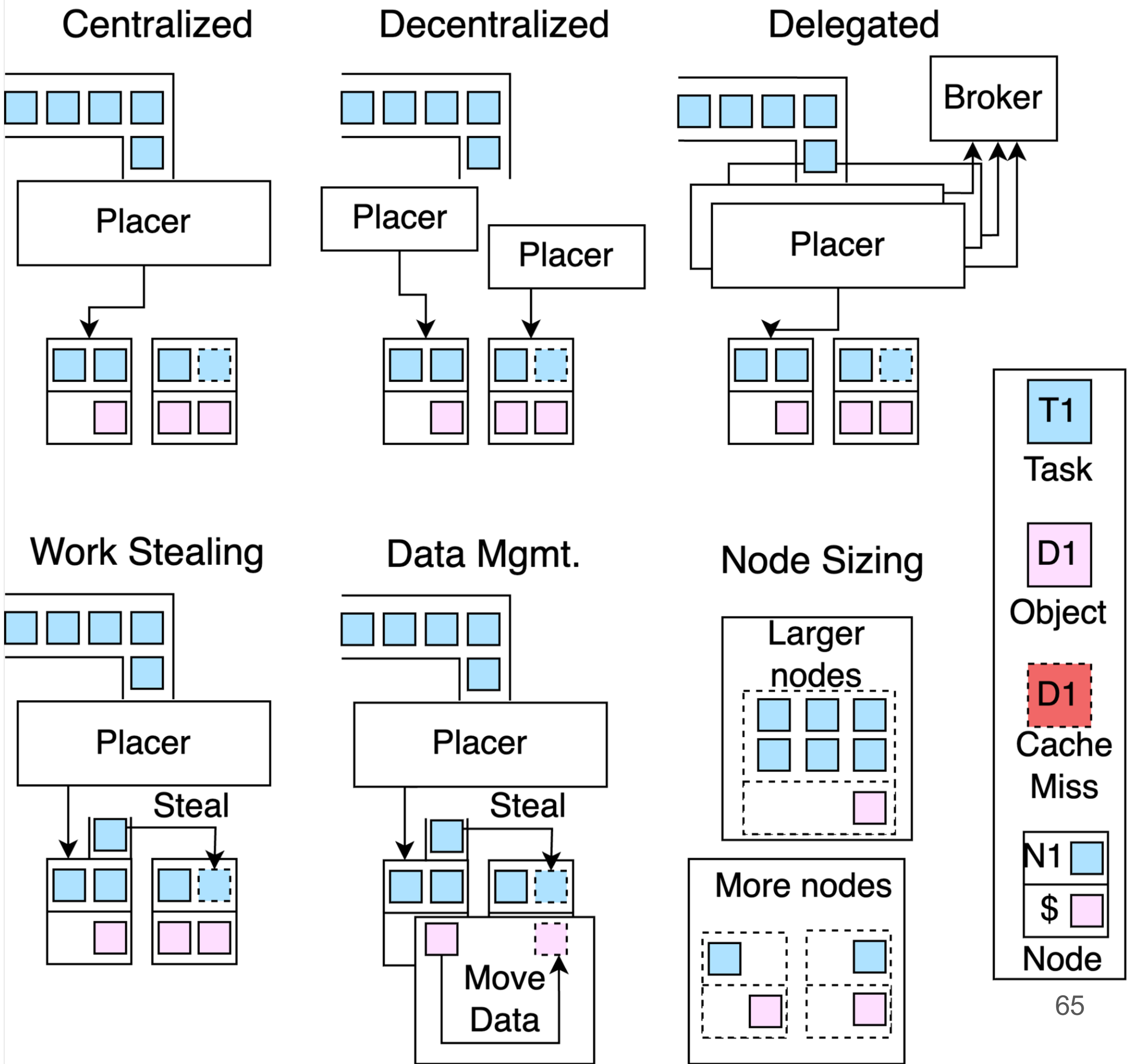
@Large Research

B

02

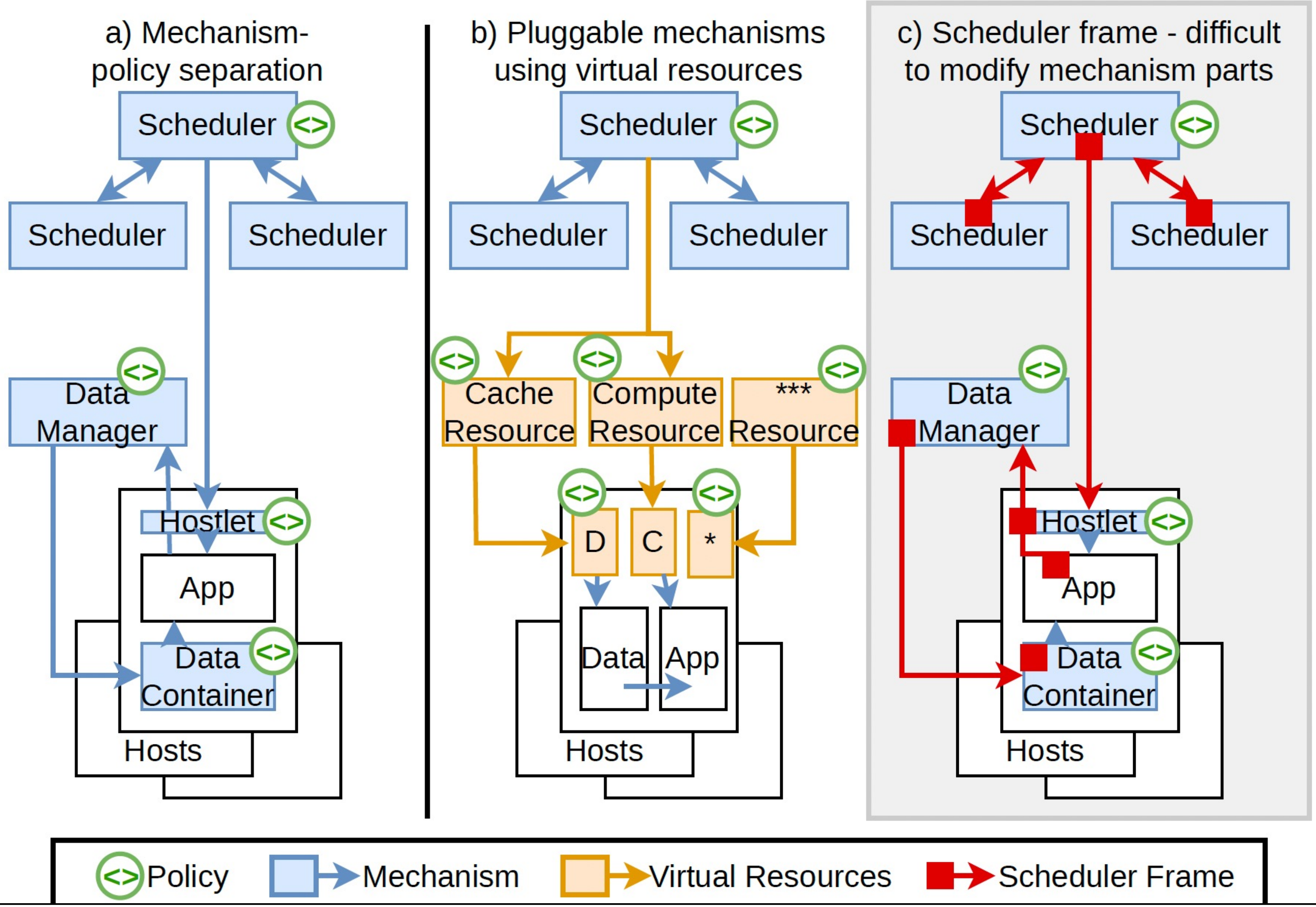
Serverless Requires A Variety of Schedulers & Mechanisms

Large Design Space



https://atlarge-research.com/pdfs/exde_fgcs23.pdf

SCHEDULING CONCEPTS [3/3]: FRAMES (Talluri et al., 2023)



Scheduler Frame

the set of all mechanisms that enable actions not possible by any local modification of the scheduler algorithm and policy.

Instead, a frame requires coordination between multiple scheduler components.

Sacheendra Talluri, Nikolas Herbst, Cristina Abad, Tiziano De Matteis, Alexandru Iosup, (2023) ExDe: Design Space Exploration of Scheduler Architectures and Mechanisms for Serverless Data-processing. FGCS. (in print)

Existing Mechanisms Characterized Using Frames

Mechanism	Components used						Implementations
	Placer(s)	Broker	Host	Client	Metadata	DM	
Architecture	✓	✓		✓			Centralized [28, 29], decentralized [30, 4], delegated [31, 13], hybrid [32, 33]
Preemption	✓		✓				Threshold-based [34], fair sharing [24]
Control-flow	✓	✓	✓				Push/pull [35], speculative exec. [36]
Data placement	✓				✓	✓	Shuffle [27], intermediate data [21]
Fault tolerance	✓			✓	✓	✓	Checkpoint [37], retry [38]
Networking	✓				✓		NetHint [39]
Barriers	✓	✓					Gang scheduling [18, 40]

https://atlarge-research.com/pdfs/exde_fgcs23.pdf

B

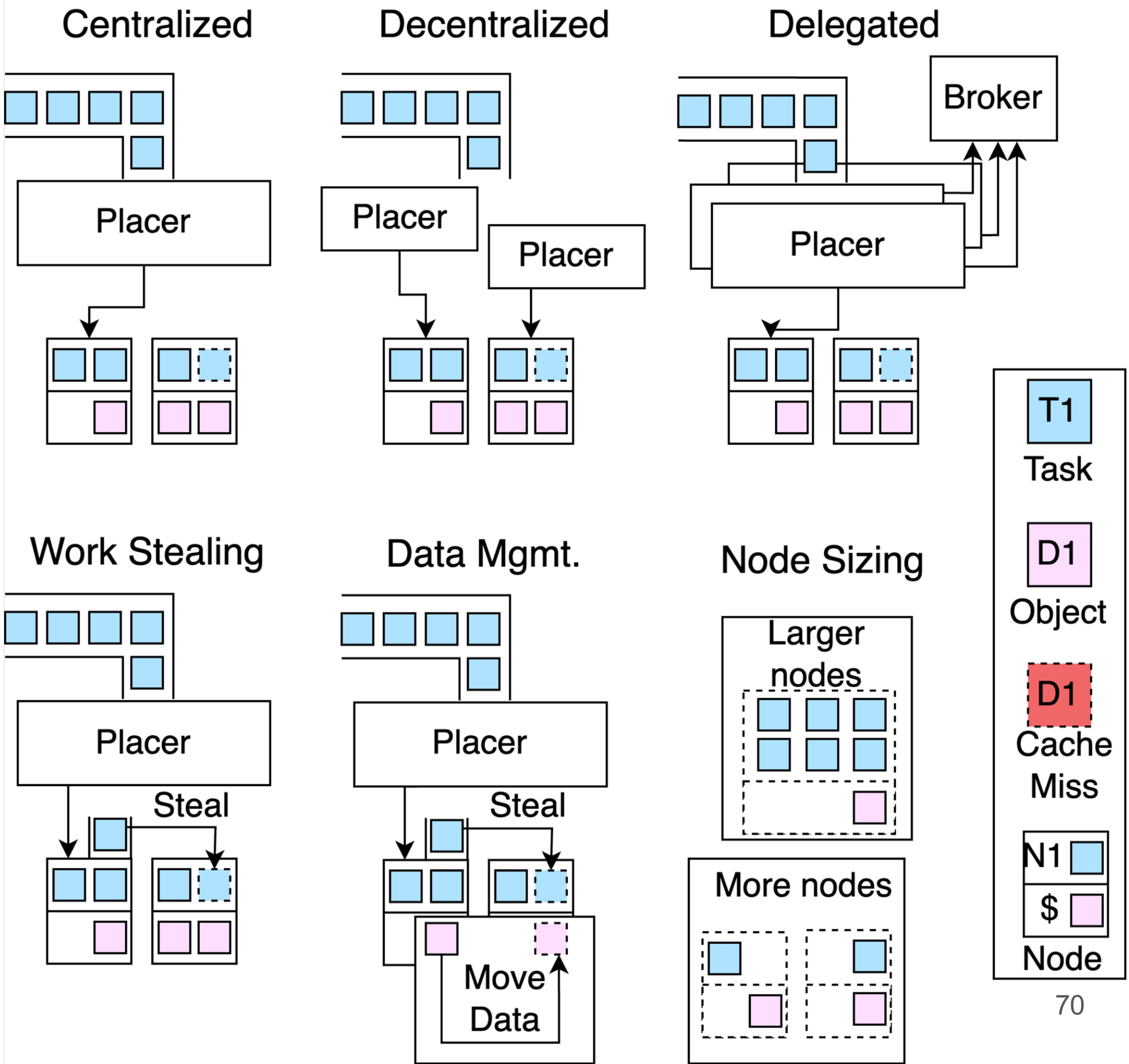
02 Simulator Setup

OpenDC Simulator

- 54 traces from IBM
- 3 million tasks per trace
- 15-31 node clusters based on the trace
- 4 CPUs per node
- FIFO + Least loaded node placement policy

Metric:

Slowdown = executime time / ideal execution time

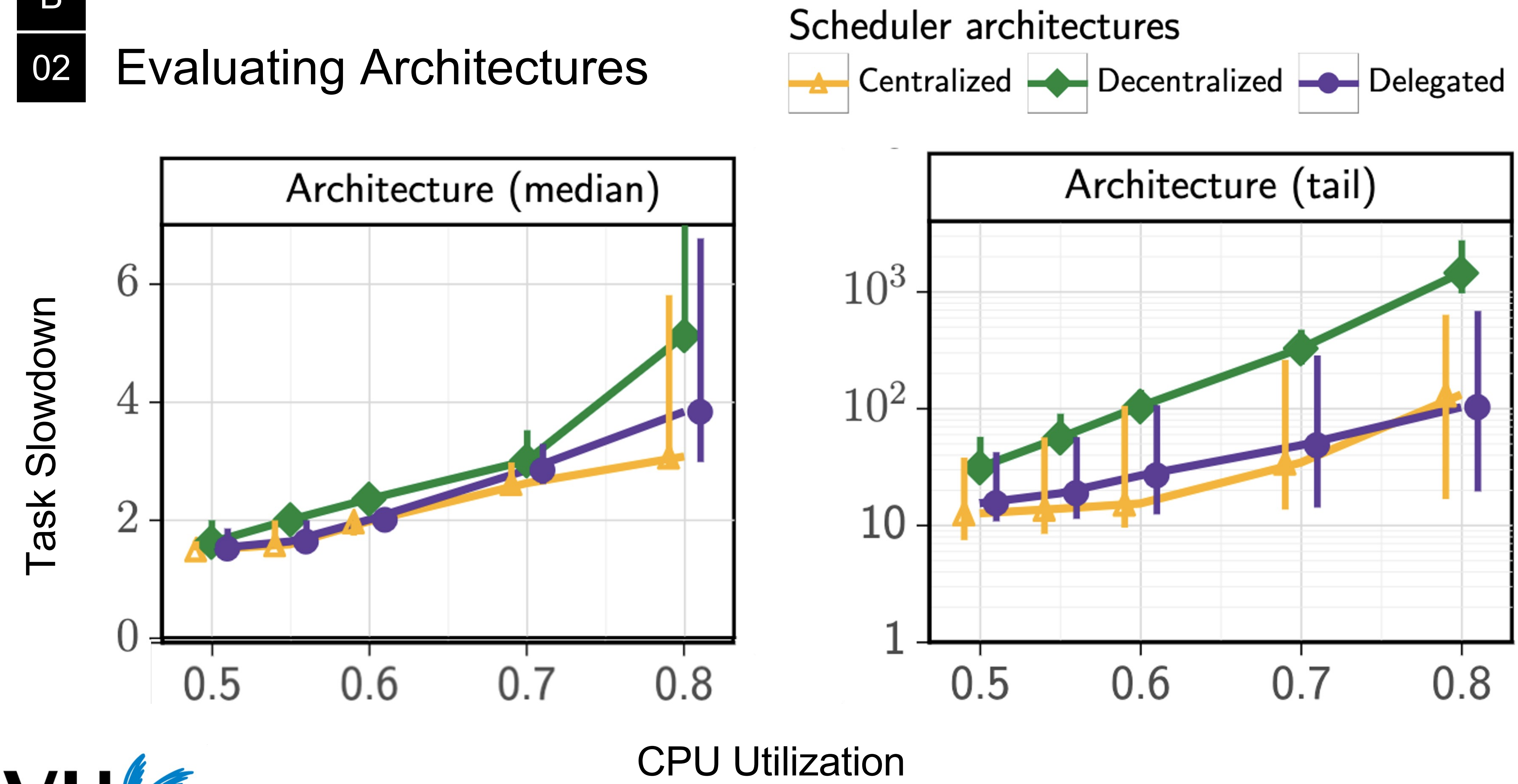


https://atlarge-research.com/pdfs/exde_fgcs23.pdf

B

02

Evaluating Architectures



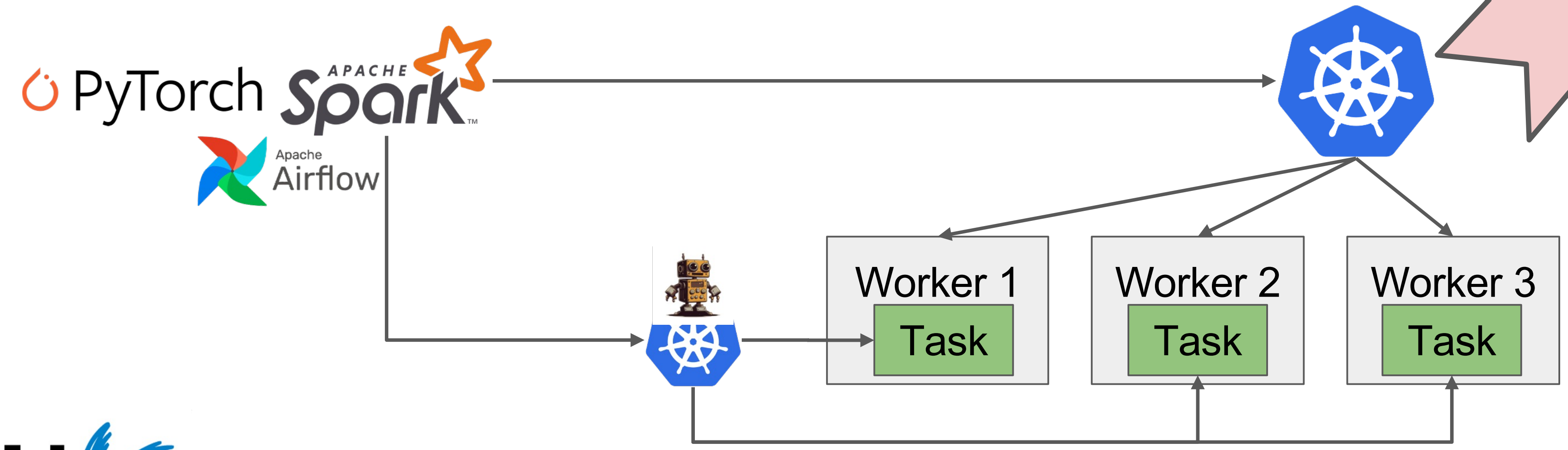
https://atlarge-research.com/pdfs/exde_fgcs23.pdf

To Real World: Kubernetes-based Scheduler as a Service [WIP]

X2

- Distributed component support based on Scheduler Frames
- Plug-n-play for multiple systems
- Advanced scheduling policy and mechanism library

Join us!



https://atlarge-research.com/pdfs/exde_fgcs23.pdf

WHAT IS A DIGITAL TWIN FOR ICT INFRASTRUCTURE?

X1 monitoring and datagen

A Real-World ICT Infrastructure

Real-World Applications

Real-World SW + Service

Real-World HW

B Virtu

02

03

“Holistic, end-to-end perspective is important”
-- Dejan Milojcic at ModSim’24

goal-oriented steering, RM&S decisions X2

VU

X1+ ODA 02

OUR VISION: THE ICT INFRASTRUCTURE MEMEX

Inspired by Bush (1945) As we may think. The Atlantic, Jul 1945.

UNCOVERING THE MYSTERIES OF OUR UNIVERSE, PHYSICAL AND DIGITAL

Find and eradicate performance issues

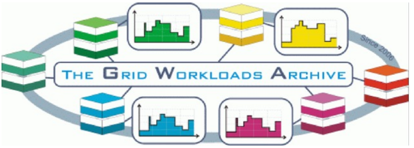
Get quantitative evidence

Enable new designs and automation

Consider culture and ethics

Understand how entire ecosystems behave and evolve

Cloud, Grid, Edge, Fog, etc.



[Iosup et al. FGCS'08]

One aspect: BigData, P2P

P2P T.A.

[Zhang et al. CoNext'10]

Sci.&Eng. Apps+Sys.

[Versluis et al. TPDS'21]

Groups, Workflows

[Iosup et al. IEEE IC'11]

Consumer Apps+Sys.

Game Trace Archive

[Guo et al. NETGAMES'12]

Enterprise Sys.

Business -Critical

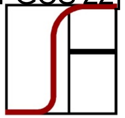
[Shen et al. CCGRID'15]

Systems, Ecosystems

[Ghiț et al. CCGRID'14]

Performance, Availability,

[Versluis et al. FGCS'22]



[Iosup et al. CCGRID'10]

DISCOVERY = LARGE-SCALE, LONG-TERM STUDY

X1

UNCOVERING THE MYSTERIES OF OUR PHYSICAL UNIVERSE

GEORGE SMOOT
NOBEL PRIZE 2006

SKA FUNDING: 500+ FTE, EUR 1.5B

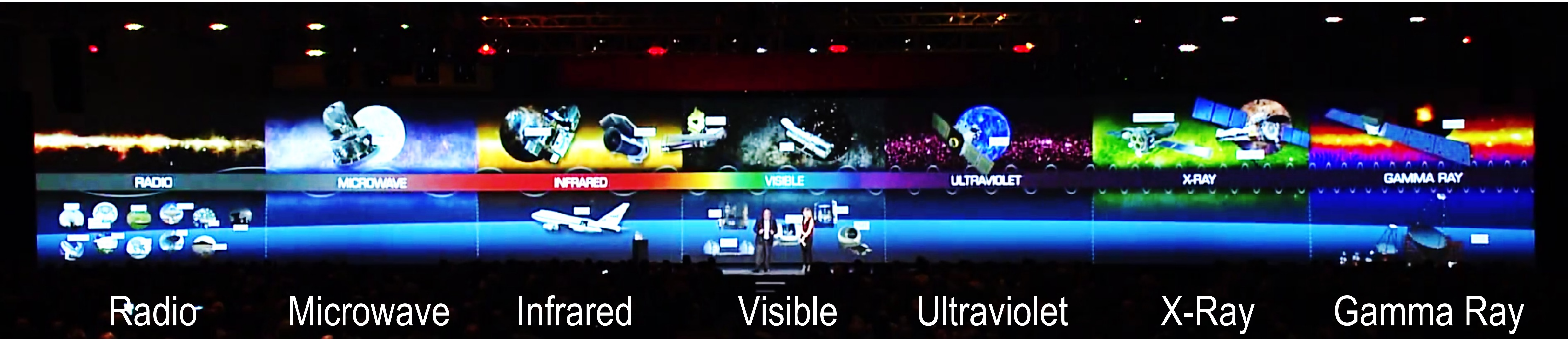
Radio Microwave Infrared Visible Ultraviolet X-Ray Gamma Ray

James Cordes. The Square Kilometer Array. Project Description. 2009 [Online]
The Square Kilometer Array Factsheet, How much will it cost?, 2012 [Online]
Phil Diamond and Rosie Bolton, Life, the Universe & Computing: The story of the SKA Telescope, SC17 Keynote. [Online]

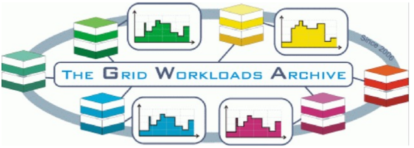
DISCOVERY = LARGE-SCALE, LONG-TERM STUDY

X1

UNCOVERING THE MYSTERIES OF OUR UNIVERSE, PHYSICAL AND DIGITAL



Cloud, Grid, Edge, Fog, etc.



[Iosup et al. FGCS'08]

One aspect: BigData, P2P

P2P T.A.

[Zhang et al. CoNext'10]

Sci.&Eng. Apps+Sys.

[Versluis et al. TPDS'21]

Groups, Workflows

[Iosup et al. IEEE IC'11]

Consumer Apps+Sys.

Game Trace Archive

[Guo et al. NETGAMES'12]

Enterprise Sys.

Business -Critical

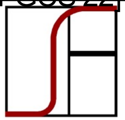
[Shen et al. CCGRID'15]

Systems, Ecosystems

[Ghiț et al. CCGRID'14]

Performance, Availability,

[Versluis et al. FGCS'22]



[Iosup et al. CCGRID'10]

THE WORKFLOW TRACE ARCHIVE

96 traces

X1

Laurens Versluis

METADATA AND TRACES FOR YOUR WORKFLOW SYSTEMS

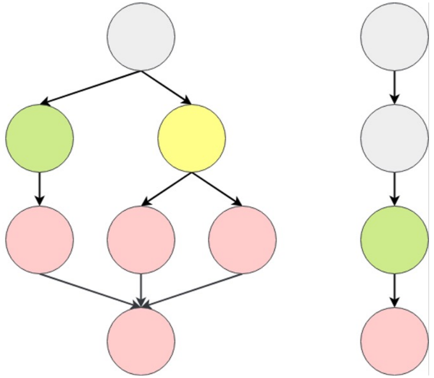
WORKFLOWS ARE COMMON IN MANY DOMAINS

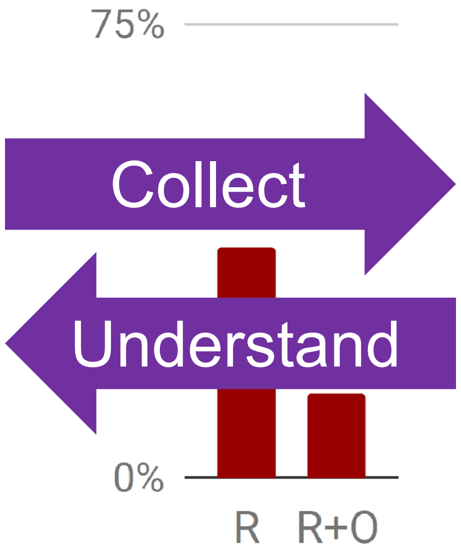
EXCEPT IN SCI., DESIGN, & ENG.

THE WORKFLOW TRACE ARCHIVE
→ WORKFLOWS ARE DIVERSE!

Protein BLAST
protein ▸ protein

Shell





universität innsbruck

Alibaba.com Global trade starts here.™

WorkflowHub

Shell

TWO SIGMA

Los Alamos NATIONAL LABORATORY

Google

TU Delft Delft University of Technology

Pegasus

VU UNIVERSITY AMSTERDAM

spec Research

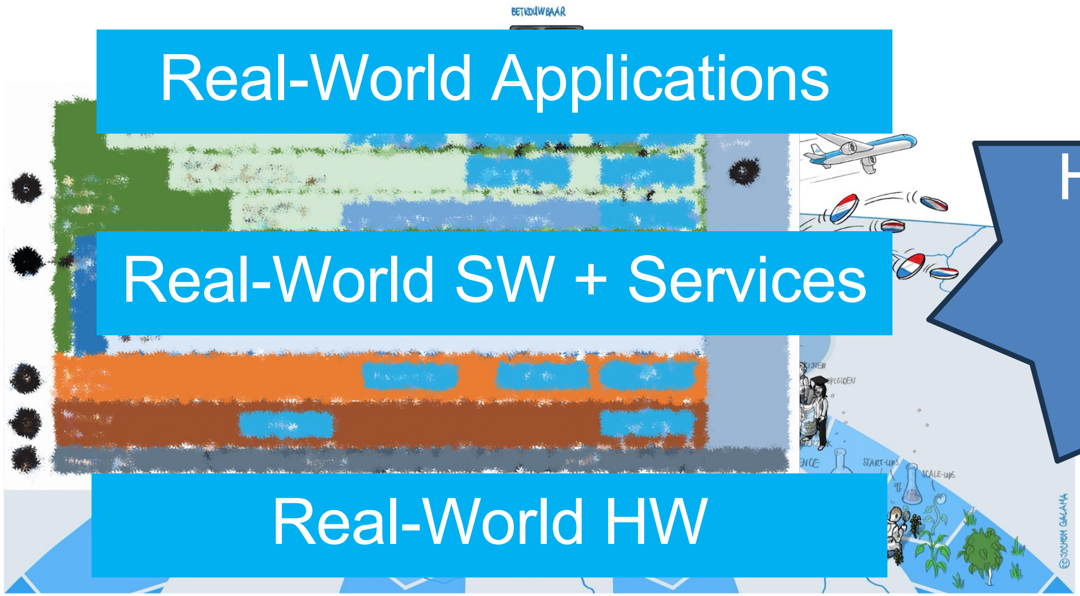
[Versluis et al. The Workflow Trace Archive]
IEEE TPDS + http://arxiv.org/pdf/1906.07471

http:// wta.atlarge.science

WHAT IS A DIGITAL TWIN FOR ICT INFRASTRUCTURE?

X1 monitoring and datagen

A Real-World ICT Infrastructure



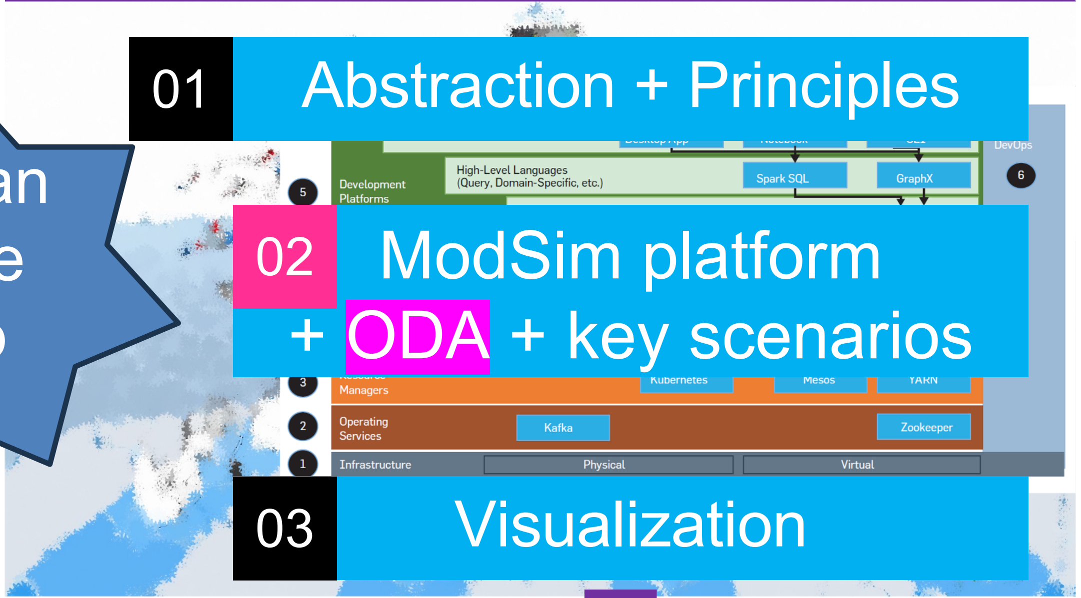
Real-World Applications

Real-World SW + Services

Real-World HW

Human in the loop

B Virtual ICT Infrastructure



01 Abstraction + Principles

02 ModSim platform + ODA + key scenarios

03 Visualization

goal-oriented steering, RM&S decisions X2

VU

ODA: DISCOVER PHENOMENA, EXPLAIN WHAT'S HAPPENING 02

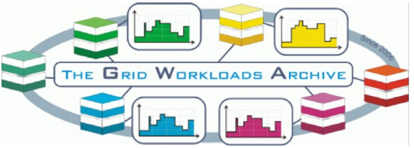
UNCOVERING THE MYSTERIES OF OUR UNIVERSE, PHYSICAL AND DIGITAL

SOME OF OUR DISCOVERIES

…

BOTS, NOT PARALLEL JOBS

Cloud, Grid, Edge, Fog, etc.



[Iosup et al. FGCS'08]

…

GROUPS NOT RARE, DOMINANT

One aspect: BigData, P2P

P2P T.A.

[Zhang et al. CoNext'10]

Sci.&Eng. Apps+Sys.

[Versluis et al. TPDS'21]

Groups, Workflows

[Iosup et al. IEEE IC'11]

…

COMMUNITY DYNAMICS

Consumer Apps+Sys.

Game Trace Archive

[Guo et al. NETGAMES'12]

…

SYSTEMIC VARIABILITY

Enterprise Sys.

Business -Critical

[Shen et al. CCGRID'15]

Systems, Ecosystems

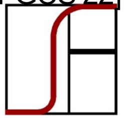
[Ghiț et al. CCGRID'14]

…

CORRELATED, NOT IID FAILURES

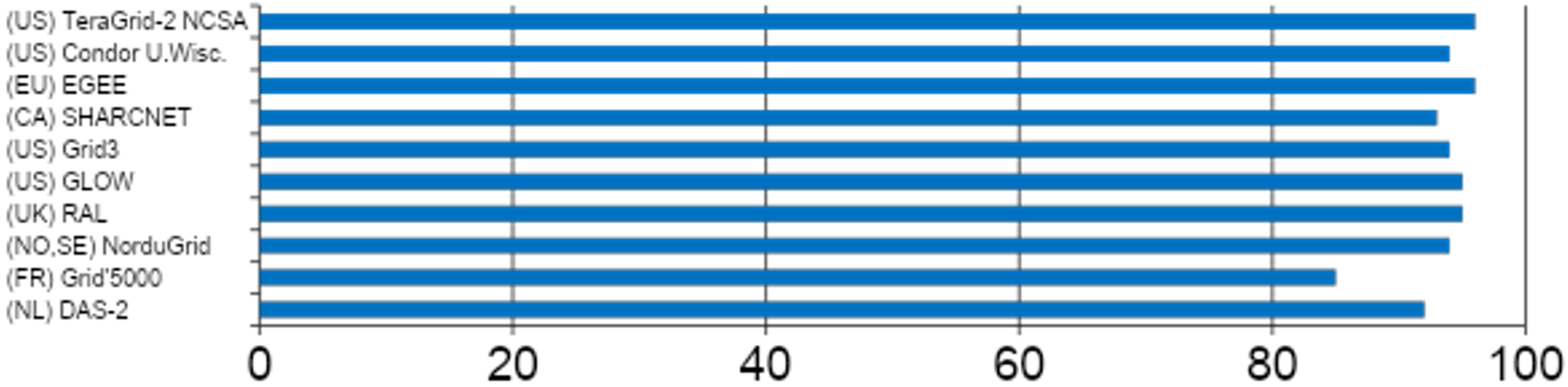
Performance, Availability,

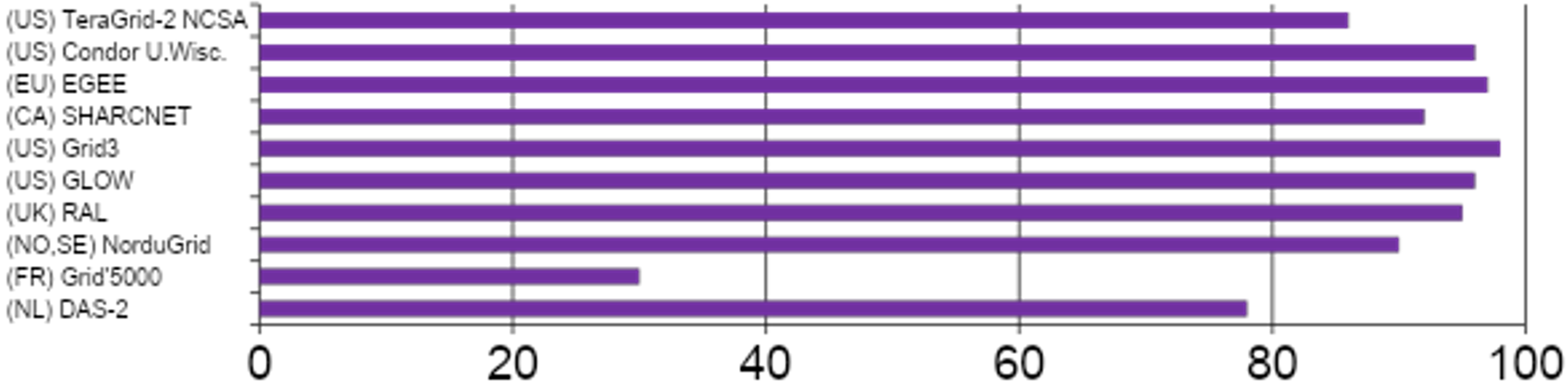
[Versluis et al. FGCS'22]



[Iosup et al. CCGRID'10]

One Phenomenon: BoTs = Dominant Programming Model for Grid Computing





Each bar is for one grid

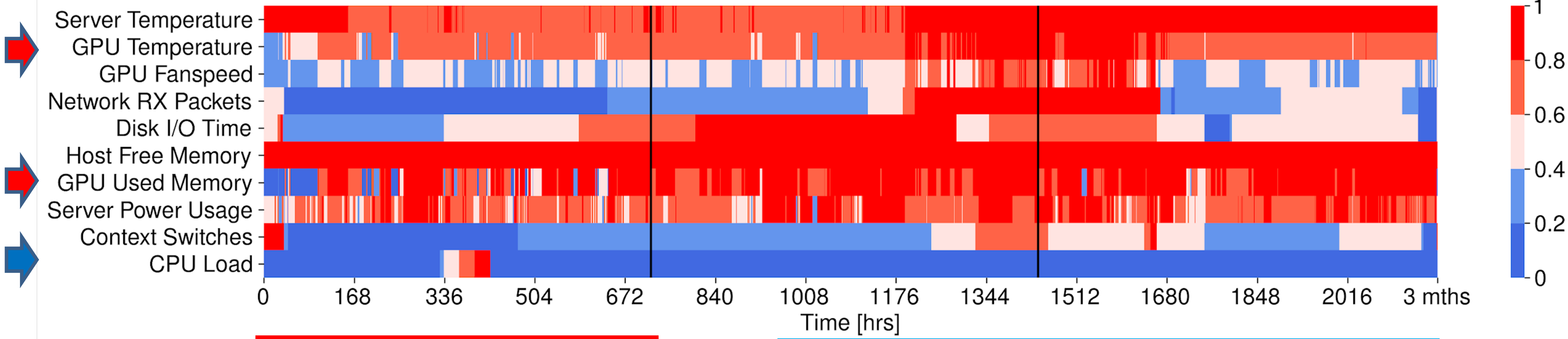
Each bar is a long-term workload trace

Iosup and Epema: Grid Computing Workloads. IEEE Internet Computing 15(2): 19-26 (2011)

VU VRIJE UNIVERSITEIT AMSTERDAM

One Basic Result

Uta et al., Beneath the SURFace: An MRI-like View into the Life of a 21st-Century Datacenter. USENIX; login 2020.



Hot GPUs

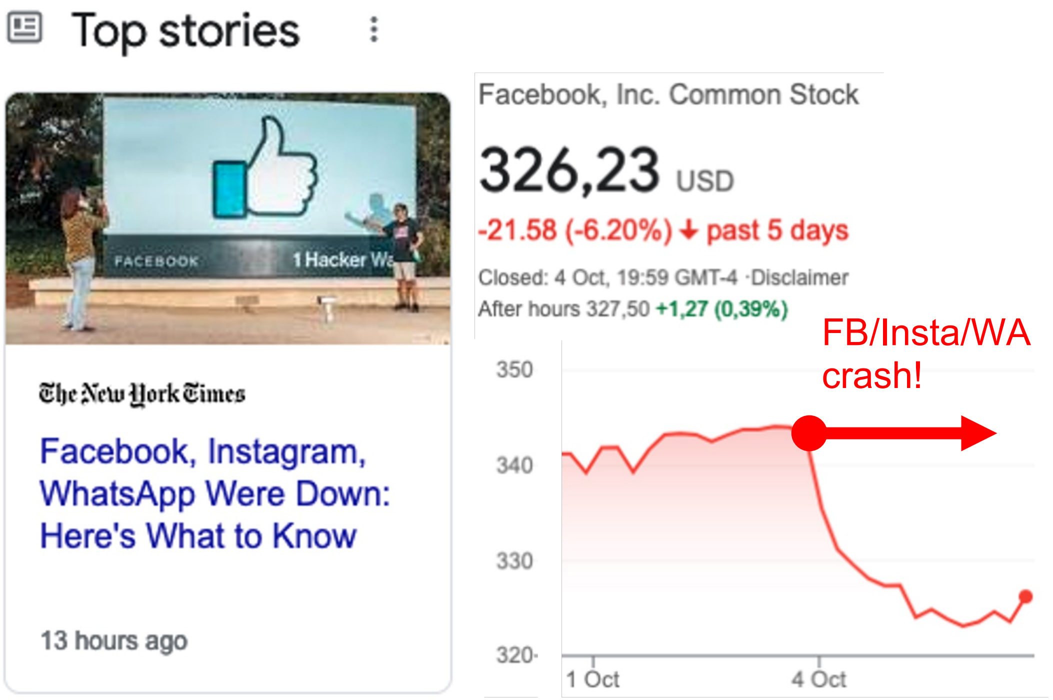
Cold CPUs*

* when GPUs

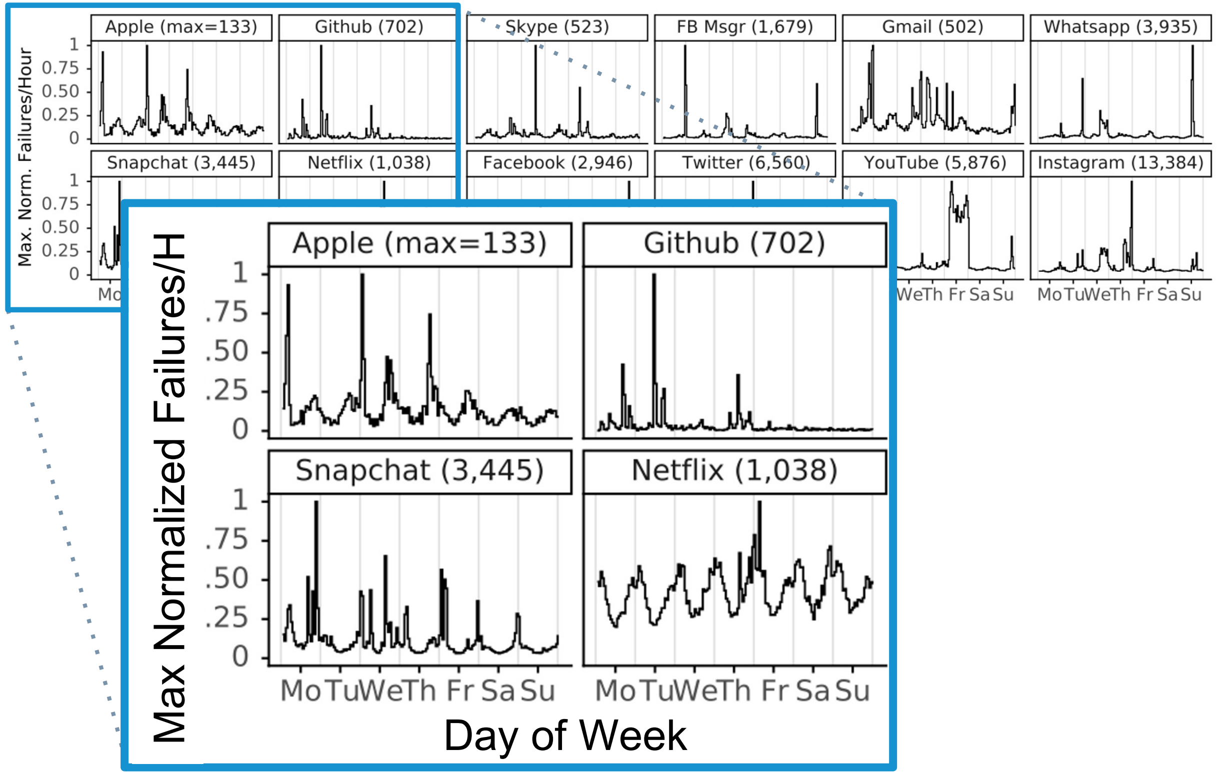
Could use less powerful, less expensive CPUs

Do Cloud Operators Tell the Truth and Nothing But the Truth? User Reports Help!

Vision: Data-driven, computer systems approaches are key to enable understanding and improving cloud dependability



Source: NYTimes, Google market tracker, Oct 5, 2021



Talluri, Overweel, Versluis, Trivedi, Iosup (2021) Empirical Characterization of User Reports about Cloud Failures. ACSOS

AGENDA

IN THIS TALK:

1. Golden Age of Computer Ecosystems
2. But we cannot take this technology for granted
3. We can try to understand and improve things for targeted applications – with Digital Twins (DT)
4. We defined a high-level reference architecture for DT
5. Let’s detail the main components (not comprehensively, not necessarily in order, and time-allowing)
6. Let’s conclude

TAKE-HOME

https://github.com/atlarge-research/opendc

We're building a Digital Twin for 21st century ICT Infrastructure (DT21). We have many (developing) theories and practical results. We know others are working on this topic and seek discussion and an active collaboration. To discuss:

- Theory: What are the core components of DT21? How do they relate - what is a good reference architecture for DT21?
- Theory: Are performance and availability just parts of a conceptual continuum? How to express sustainability? What else is in there and how to include it in DT21?
- Theory: We say: Just touch it with your lower lip, briefly, and move away if too hot. How to reason about energy use, both short- and long-term? What scenarios?
- Theory/Practice: Finding the right abstractions, principles, architectures, systems for scaling data.
- Theory/Practice: What are the interfaces? How to enable interoperable ecosystems/DTs?
- Practice: Sharing traces collected in ICT infrastructure. (Do we need ontologies? What is a good middle ground between implicit ontologies and exhaustive ontologies?)
- Practice: What kinds of DT21 benefit from more complex types of analysis, e.g., simulation combined with graph analytics and (ML-based) learning?
- Hiring: Education with DT21 is much cheaper to organize, can reveal deep operational aspects and engage a diverse new generation of experts.
- Ethics: If a tree falls in front of you, do you have to see it? From plausible deniability to responsibility in ICT infrastructure management.

Talk available: bit.ly/VUDigitalTwin24

WANT TO READ MORE ON THE TOPIC?

FURTHER READING

https://atlarge-research.com/publications.html

1. Alexandru Iosup, Fernando Kuipers, Ana Lucia Varbanescu, Paola Grosso, Animesh Trivedi, Jan S. Rellermeyer, Lin Wang, Alexandru Uta, Francesco Regazzoni (2022) Future Computer Systems and Networking Research in the Netherlands: A Manifesto. CoRR abs/2206.03259.
2. Talluri, S., Herbst, N., Abad, C., De Matteis, T., & Iosup, A. (2024). ExDe: Design space exploration of scheduler architectures and mechanisms for serverless data-processing. Future Generation Computer Systems, 153, 84-96.
3. Dante Niewenhuis et al. (2024) FootPrinter: Quantifying Data Center Carbon Footprint. In ACM/SPEC ICPE. 189–195.
4. Kounev, Herbst, Abad, Iosup, et al. (2023) Serverless Computing: What It Is, and What It Is Not? CACM 66(9).
5. Jansen et al. (2023) The SPEC-RG Reference Architecture for The Compute Continuum. CCGRID.
6. Versluis et al. (2023) Less is not more: We need rich datasets to explore. FGCS 142.
7. Crusoe et al. (2022) Methods included: standardizing computational reuse and portability with the Common Workflow Language. CACM 65(6).
8. Andreadis et al. (2022) Capelin: Data-Driven Capacity Procurement for Cloud Datacenters using Portfolios of Scenarios. IEEE TPDS 33(1).
9. Eismann et al. (2022) The State of Serverless Applications: Collection, Characterization, and Community Consensus. IEEE Trans. Software Eng. 48(10).
10. Sakr, Bonifati, Voigt, Iosup, et al. (2021) The future is big graphs: a community view on graph processing systems. Commun. ACM 64(9).
11. Mastenbroek et al. (2021) OpenDC 2.0: Convenient Modeling and Simulation of Emerging Technologies in Cloud Datacenters. CCGRID.
12. Versluis and Iosup (2021) A survey of domains in workflow scheduling in computing infrastructures: Community and keyword analysis, emerging trends, and taxonomies. FGCS 123.
13. Papadopoulos et al. (2021) Methodological Principles for Reproducible Performance Evaluation in Cloud Computing. IEEE Trans. Software Eng. 47(8).
14. Versluis et al. (2020) The Workflow Trace Archive: Open-Access Data From Public and Private Computing Infrastructures. IEEE TPDS 31(9).
15. Alexandru Uta, Alexandru Custura, Dmitry Duplyakin, Ivo Jimenez, Jan S. Rellermeyer, Carlos Maltzahn, Robert Ricci, Alexandru Iosup (2020) Is Big Data Performance Reproducible in Modern Cloud Networks? NSDI.
16. Erwin Van Eyk, Lucian Toader, Sacheendra Talluri, Laurens Versluis, Alexandru Uta, Alexandru Iosup (2018) Serverless is More: From PaaS to Present Cloud Computing. IEEE Internet Comput. 22(5).

EXTRA! EXTRA! WANT EVEN MORE ON THIS TOPIC? LET’S TALK.

@L EDUCATION

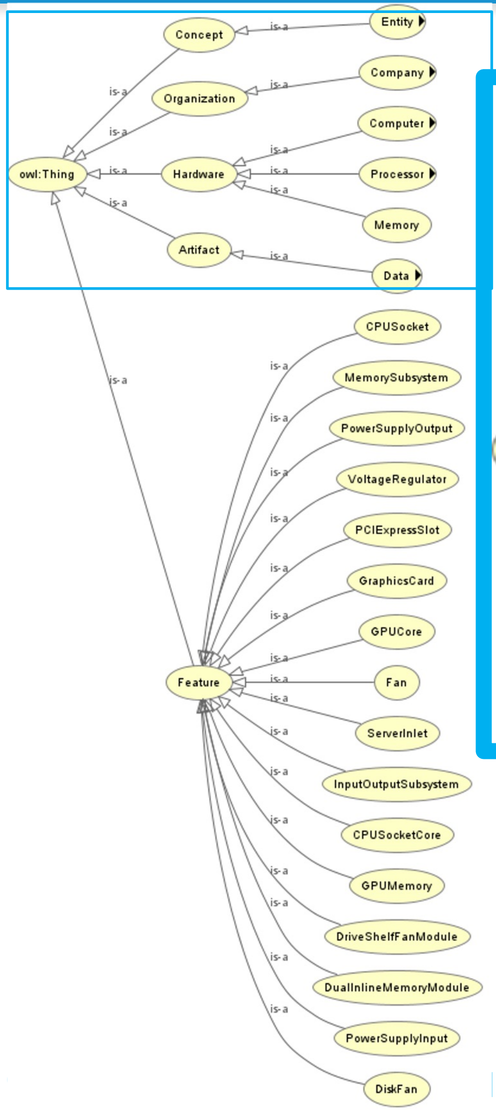
Integrate the ICT digital twin into our compsys coursework and education processes.
It's fun, stimulates curiosity, leads to learning!

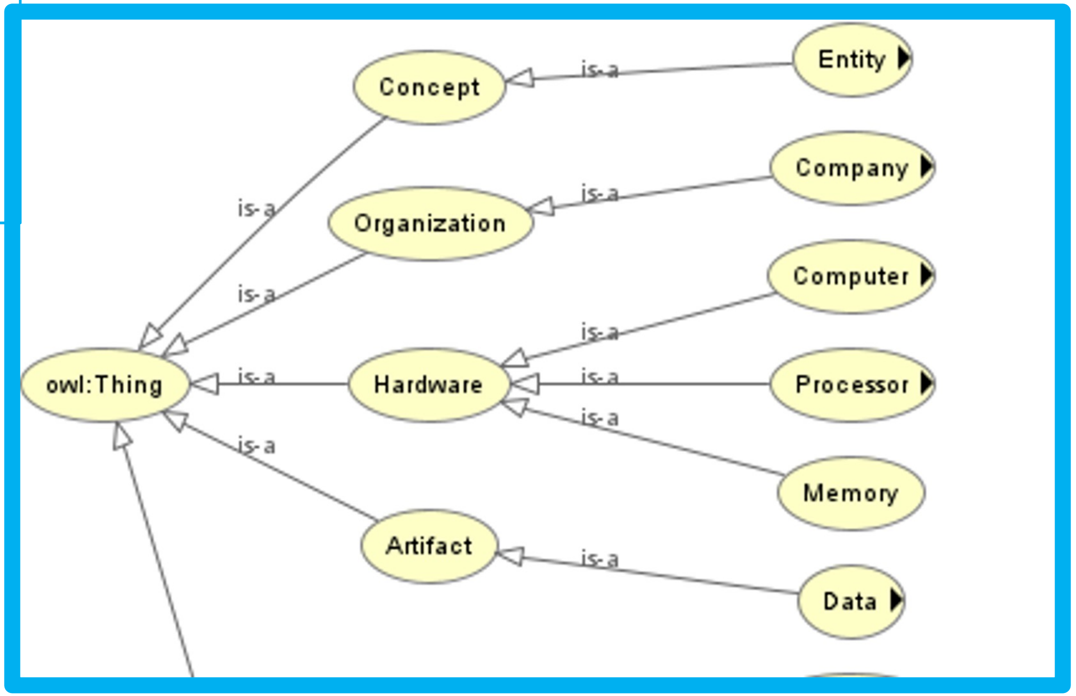
INTEROPERABILITY THROUGH A NARROW ONTOLOGY FOR ODA (STANDARDIZE 'D')

Problem:
Different data formats, collection processes.

Solution:

1. Define a narrow ontology
2. Integrate into data science process
3. Implement for usability, e.g., graph, time-series, relational database

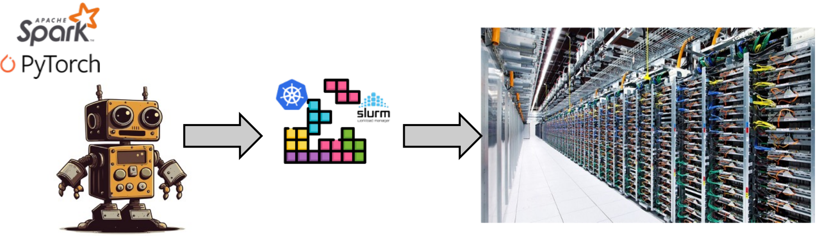


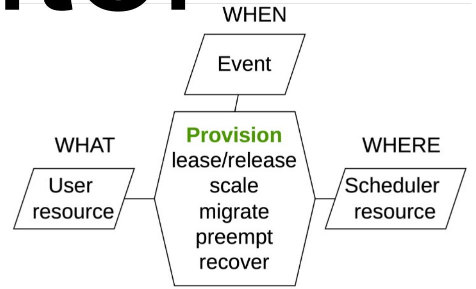


Ongoing work

Xiaoyu Chu, Shekhar Suman, Martin Molan, Andrea Bartolini (UniBo), Iosup, et al. (2023) Ontology for HPC Infrastructure Data

The Cost of Simplicity: Understanding Datacenter Scheduler Programming Abstractions





Aratz Manterola Lasa[1], **Sacheendra Talluri**[2], Tiziano De Matteis[2], Alexandru Iosup[2]

[1]WarpStream Labs
[2]Vrije Universiteit Amsterdam

Manterola Lasa, A., Talluri, S., De Matteis, T., & Iosup, A. (2024, May). The Cost of Simplicity: Understanding Datacenter Scheduler Programming Abstractions. In Proceedings of the 15th ACM/SPEC International Conference on Performance Engineering (pp. 166-177).

https://atlarge-research.com/pdfs/2024-icpe-datacenter-scheduler.pdf

@Large Research