MASSIVIZING COMPUTER SYSTEMS

MASSIVIZING GRAPH PROCESSING:

THE SCIENCE, DESIGN, AND ENGINEERING OF A COMPLEX ECOSYSTEM

Sponsored by:

NWO

@Large Research
Massivizing Computer Systems

http://atlarge.science

bit.ly/AtLargeGraphSys22

Massivizing =
Rich challenge of computer science □ high societal impact!

in @AIosup

VU VRIJE UNIVERSITEIT AMSTERDAM

Contributions from the MCS/AtLarge teams. Many thanks!
Many thanks to our collaborators, international working groups, authors of all images included here. Also thanks to Gabriela Ciocârlie, UTSA AI Consortium for Human Well-Being.

Prof.dr.ir. Alexandru IOSUP

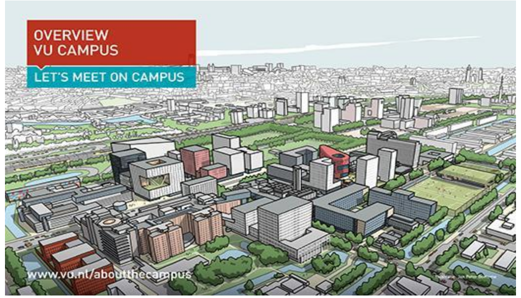
US IN 1 MINUTE

WE’RE MASSIVIZING COMPUTER SYSTEMS!

VU AMSTERDAM < SCHIPHOL < THE NETHERLANDS < EUROPE

Amsterdam

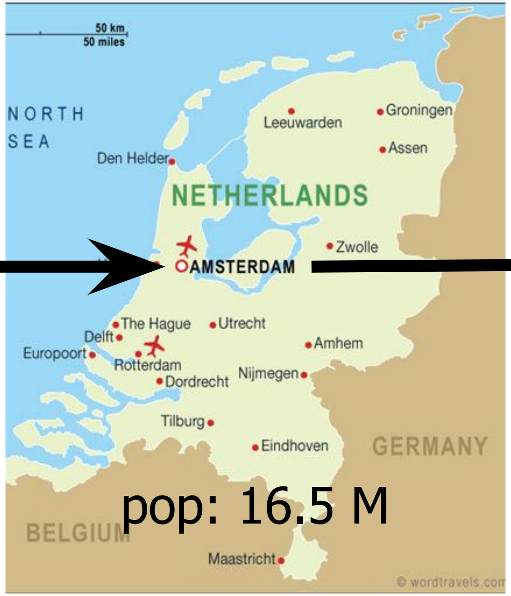
founded 10th century
pop: 850,000



VU

founded 1880
pop: 23,500

The Netherlands



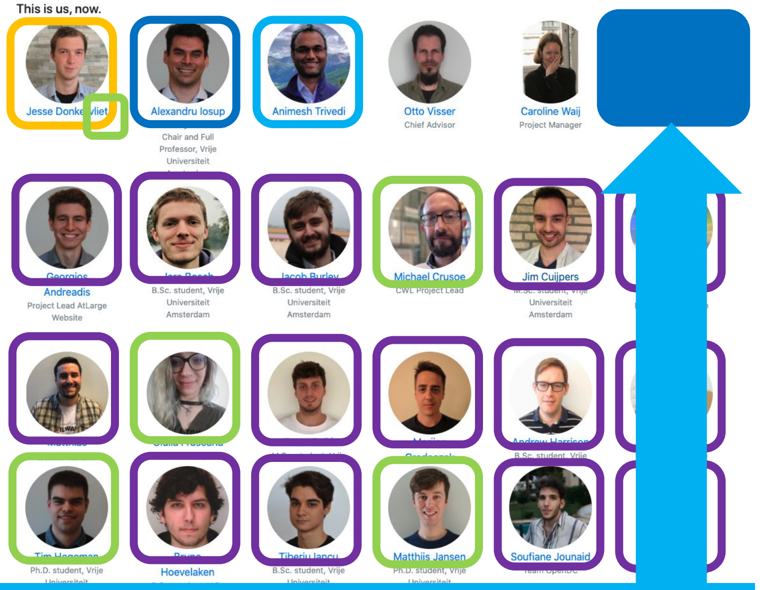
pop: 16.5 M

Europe

North Atlantic Ocean
North Sea
FINLAND
SWEDEN
RUSSIA
IRELAND
UNITED KINGDOM
DENMARK
POLAND
BELARUS
UKRAINE
GERMANY
FRANCE
SPAIN
PORTUGAL
ITALY
ROMANIA
BULGARIA
GREECE
TURKEY
Mediterranean Sea
Black Sea
London
Paris
Berlin
Amsterdam

http://atlarge.science

CURRENT TEAM



This is us, now.

Jesse Donkervliet

Alexandru Iosup
Chair and Full Professor, Vrije Universiteit

Animesh Trivedi

Otto Visser
Chief Advisor

Caroline Waij
Project Manager

Georgios Andreadis
Project Lead AtLarge Website

B.Sc. student, Vrije Universiteit Amsterdam

Jacob Burley
B.Sc. student, Vrije Universiteit Amsterdam

Michael Crusoe
CWL Project Lead

Jim Cuijpers
M.Sc. student, Vrije Universiteit Amsterdam

Andrew Harrison
B.Sc. student, Vrije

Tim Hegeman
Ph.D. student, Vrije Universiteit

Bruno Hoevelaken

Tiberiu Iancu
B.Sc. student, Vrije Universiteit

Matthijs Jansen
Ph.D. student, Vrije Universiteit

Soufiane Jounaid
Team OpenDC

🟦 Professor

🟦 Assistant Prof.

🟨 Teacher

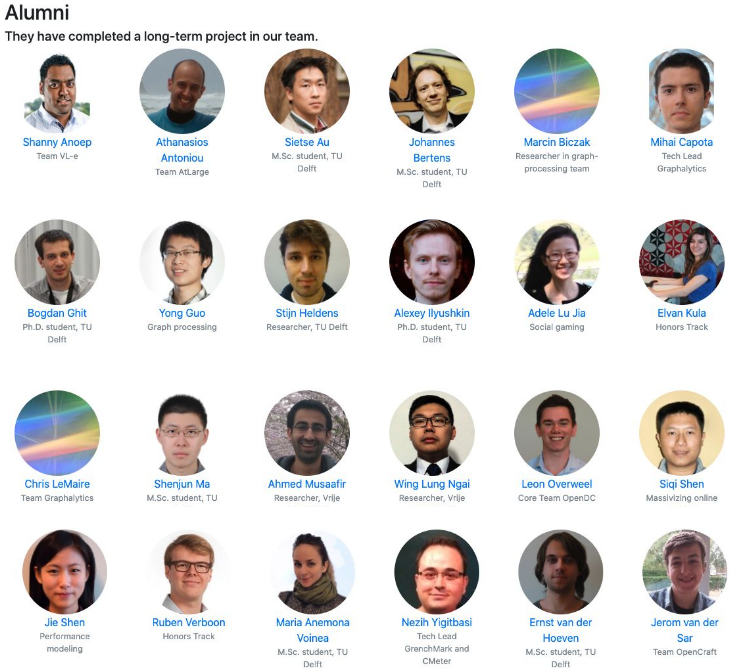
🟩 Visitor/P.-doc

🟩 Ph.D. student

🟪 Early Scientist

WE ARE HIRING
A NEW ASST. PROF.!

ALUMNI



Alumni

They have completed a long-term project in our team.

Shanny Anoep
Team VL-e

Athanasios Antoniou
Team AtLarge

Sietse Au
M.Sc. student, TU Delft

Johannes Bertens
M.Sc. student, TU Delft

Marcin Biczak
Researcher in graph-processing team

Mihai Capota
Tech Lead Graphalytics

Bogdan Ghit
Ph.D. student, TU Delft

Yong Guo
Graph processing

Stijn Heldens
Researcher, TU Delft

Alexey Ilyushkin
Ph.D. student, TU Delft

Adele Lu Jia
Social gaming

Elvan Kula
Honors Track

Chris LeMaire
Team Graphalytics

Shenjun Ma
M.Sc. student, TU

Ahmed Musaafir
Researcher, Vrije

Wing Lung Ngai
Researcher, Vrije

Leon Overweel
Core Team OpenDC

Siqi Shen
Massivizing online

Jie Shen
Performance modeling

Ruben Verboon
Honors Track

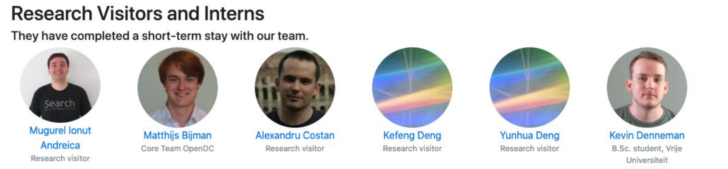
Maria Anemona Voinea
M.Sc. student, TU Delft

Nezih Yigitbasi
Tech Lead GrenchMark and CMeter

Ernst van der Hoeven
M.Sc. student, TU Delft

Jerom van der Sar
Team OpenCraft

VISITORS



Research Visitors and Interns

They have completed a short-term stay with our team.

Mugurel Ionut Andreica
Research visitor

Matthijs Bijman
Core Team OpenDC

Alexandru Costan
Research visitor

Kefeng Deng
Research visitor

Yunhua Deng
Research visitor

Kevin Denneman
B.Sc. student, Vrije Universiteit

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

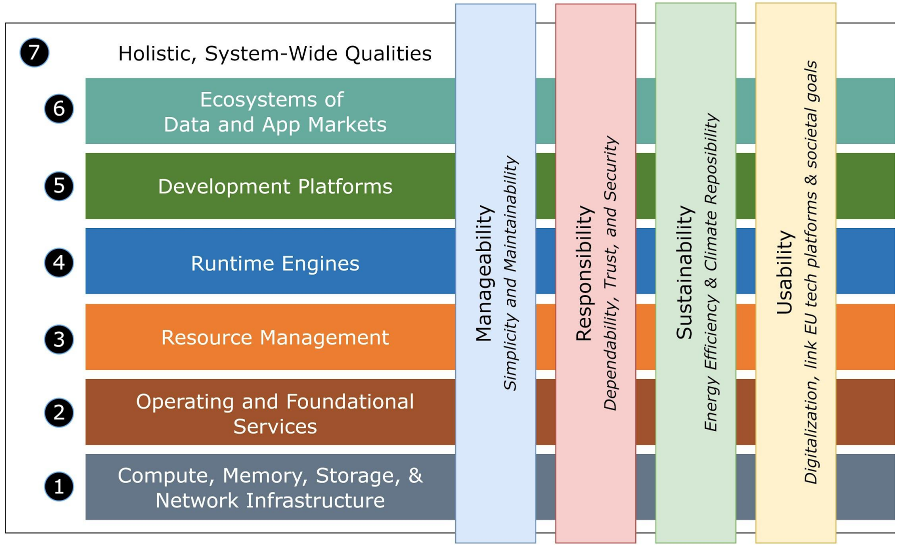
WHO AM I?
PROF. DR. IR. ALEXANDRU IOSUP

- Education, my courses:
  - Honours Programme, Computer Org. (BSc)
  - Distributed Systems, Cloud Computing (MSc)
- Research, 15 years in DistribSys:
  - Massivizing Computer Systems
  - About 30 young researchers in the team
- About me:
  - Worked in 7 countries, NL since 2004
  - 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

WE ARE ALIGNED WITH COMMUNITY CONCERNS…

The Manifesto on Computer Systems and Networking Research

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



7 Holistic, System-Wide Qualities

6 Ecosystems of Data and App Markets

5 Development Platforms

4 Runtime Engines

3 Resource Management

2 Operating and Foundational Services

1 Compute, Memory, Storage, & Network Infrastructure

Manageability *Simplicity and Maintainability*

Responsibility *Dependability, Trust, and Security*

Sustainability *Energy Efficiency & Climate Reposibility*

Usability *Digitalization, link EU tech platforms & societal goals*

Signed

50+ PIs / Leads
7 universities
5 relevant societal stakeholders

Available

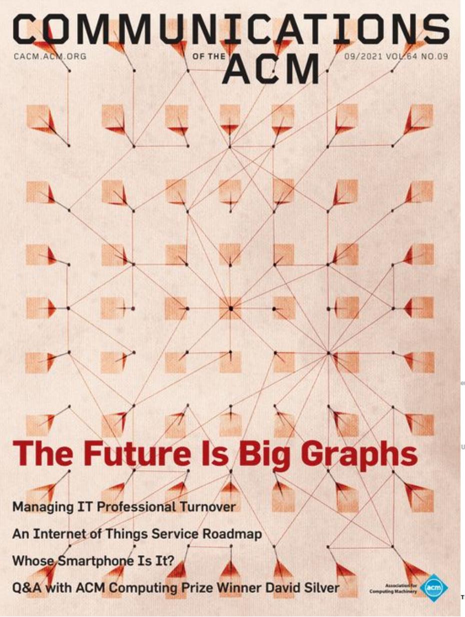
Full version (40+ pages) https://arxiv.org/pdf/2206.03259
Who’s Who in CompSysNL? https://bit.ly/CompSysNLWhosWho

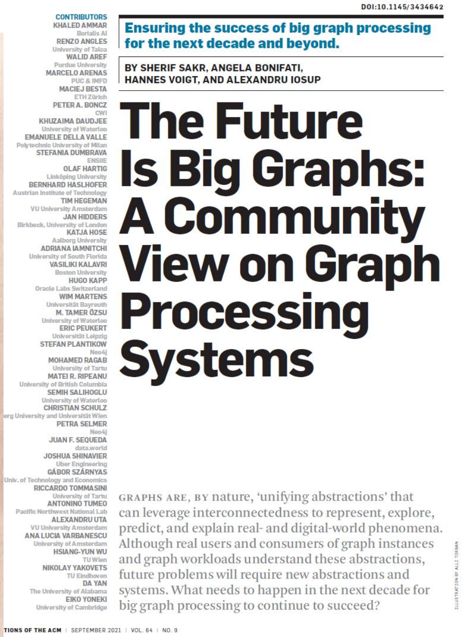
ONE PROJECT TO MENTION…

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

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

contributed articles


COMMUNICATIONS OF THE ACM
CACM.ACM.ORG
09/2021 VOL.64 NO.09
The Future Is Big Graphs
Managing IT Professional Turnover
An Internet of Things Service Roadmap
Whose Smartphone Is It?
Q&A with ACM Computing Prize Winner David Silver



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

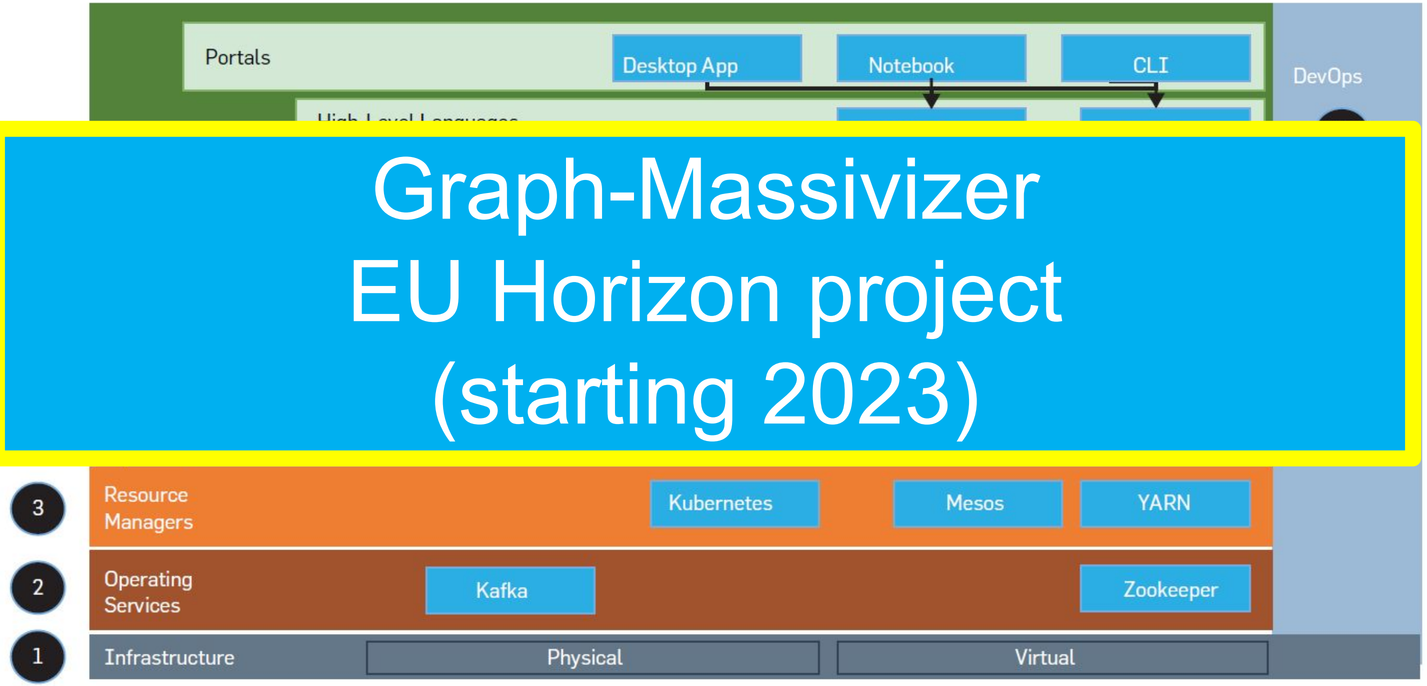
CONTRIBUTORS
KHALED AMMAR
Borialis AI
RENZO ANGLES
University of Talca
WALID AREF
Purdue University
MARCELO ARENAS
PUC & IMFD
MACIEJ BESTA
ETH Zürich
PETER A. BONCZ
CWI
KHUZAIMA DAUDJEE
University of Waterloo
EMANUELE DELLA VALLE
Polytechnic University of Milan
STEFANIA DUMBRAVA
ENSIIE
OLAF HARTIG
Linköping University
BERNHARD HASLHOFER
Austrian Institute of Technology
TIM HEGEMAN
VU University Amsterdam
JAN HIDDERS
Birkbeck, University of London
KATJA HOSE
Aalborg University
ADRIANA IAMNITCHI
University of South Florida
VASILIKI KALAVRI
Boston University
HUGO KAPP
Oracle Labs Switzerland
WIM MARTENS
Universität Bayreuth
M. TAMER ÖZSU
University of Waterloo
ERIC PEUKERT
Universität Leipzig
STEFAN PLANTIKOW
Neo4j
MOHAMED RAGAB
University of Tartu
MATEI R. RIPEANU
University of British Columbia
SEMIH SALIHOGLU
University of Waterloo
CHRISTIAN SCHULZ
erg University and Universität Wien
PETRA SELMER
Neo4j
JUAN F. SEQUEDA
data.world
JOSHUA SHINAVIER
Uber Engineering
GÁBOR SZÁRNYAS
Univ. of Technology and Economics
RICCARDO TOMMASINI
University of Tartu
ANTONINO TUMEO
Pacific Northwest National Lab
ALEXANDRU UTA
VU University Amsterdam
ANA LUCIA VARBANESCU
University of Amsterdam
HSIANG-YUN WU
TU Wien
NIKOLAY YAKOVETS
TU Eindhoven
DA YAN
The University of Alabama
EIKO YONEKI
University of Cambridge

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?

ILLUSTRATION BY ALLI TORBAN

TIONS OF THE ACM | SEPTEMBER 2021 | VOL. 64 | NO. 9

CACM Cover/Featured article, Sep 2021



Graph-Massivizer
EU Horizon project
(starting 2023)

(*) Digital twin for datacenters, with partners CINECA, UniBo, etc.

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

@L

THIS IS THE GOLDEN AGE OF COMPUTER ECOSYSTEMS

1

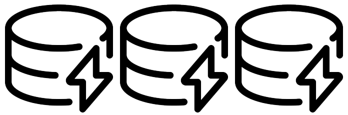
…but >>

THIS IS THE GOLDEN AGE OF MASSIVE COMPUTER ECOSYSTEMS

+ AI, esp. ML

Education for Everyone (Online)

Business Services



Big Data

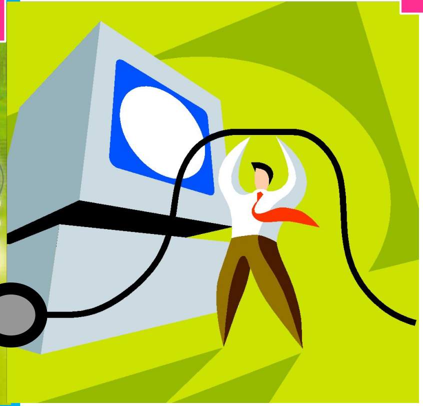
Edge / IoT / Fog / … Computing

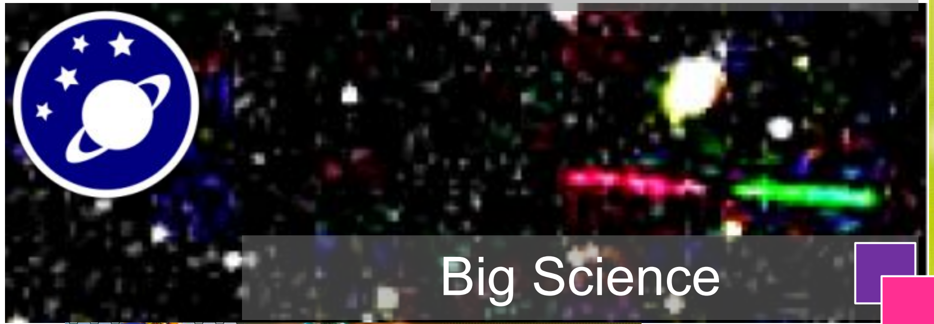
Internet

5G/6G/*

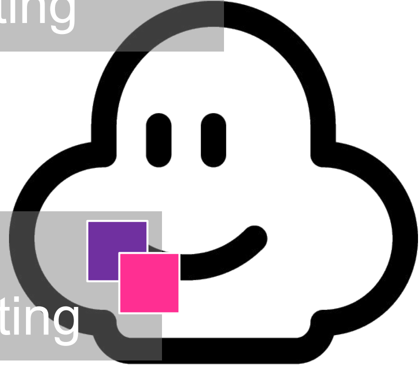
THz

VLC





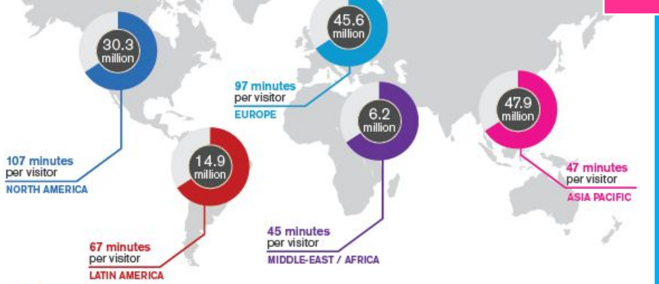
Big Science



Cloud Computing

CGACATAT
CTAAAGATGATCTTTAGTCCCGGTTCGAA
TCTTTAGTCCCGGTTGATAACACCAACC
GTAATACCAACCGGGACTAAAGATCCCG
GGGACTAAAGTCCCACCCCTATATATATG
TTCAAAATTTCTTCAAAAAAGAGGGGAG

Online Gaming

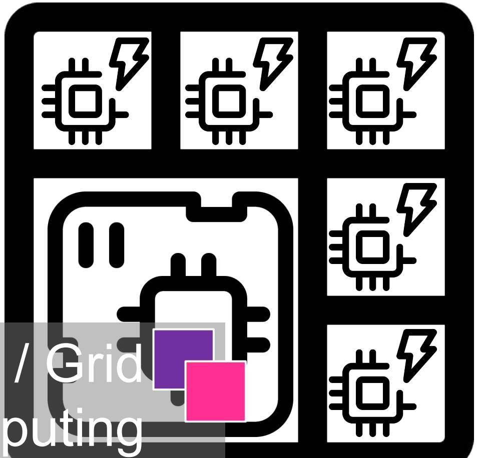


+ data as graphs

ABN·AMRO

Daily Life

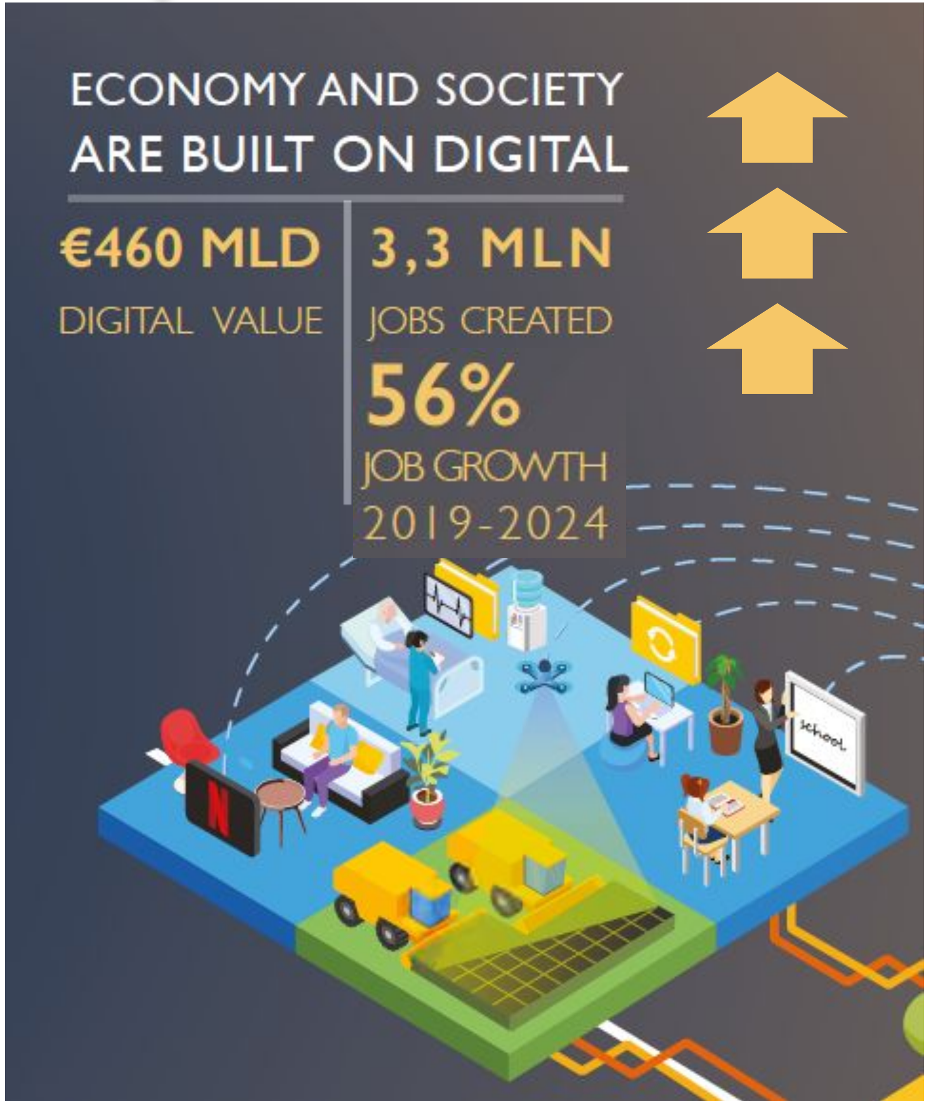
Datacenter



HPC / Grid Computing

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

THE ECONOMIC IMPACT OF MASSIVE COMPUTER ECOSYSTEMS



DIVERSE SERVICES FOR ALL

EVERY €1 → €15 ADDED VALUE

Impacting >60% of the NL GDP (1 trillion EUR/y)

Attracting >20% of all foreign direct investments in NL

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.

BUT WE CANNOT TAKE THIS TECHNOLOGY FOR GRANTED

2

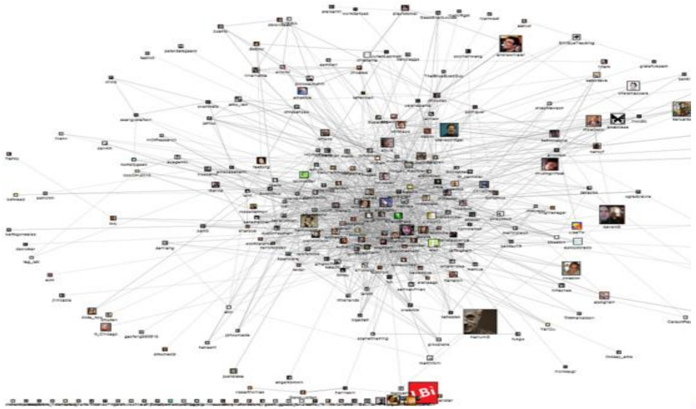
(So, this is why I am giving this talk)

Golden Age <<

This talk >>

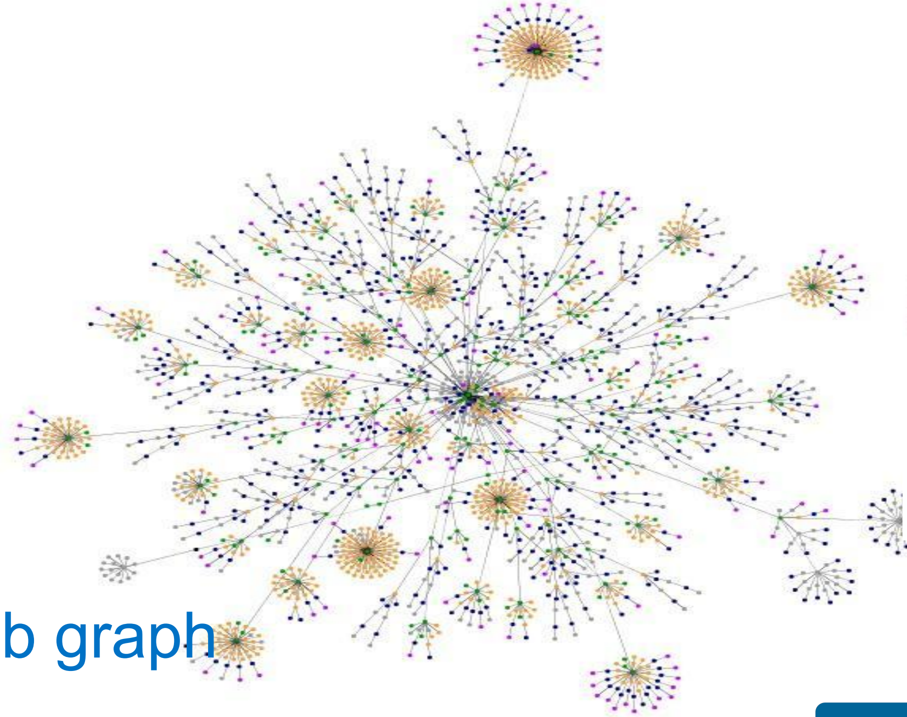
GRAPH DATA EVERYWHERE, AT UNPRECEDENTED SCALE

NEED TO MASSIVIZE GRAPH PROCESSING



Social / gaming network

~1 billion vertices
~100 billion connections



Web graph

~50 billion pages
~1 trillion hyperlinks

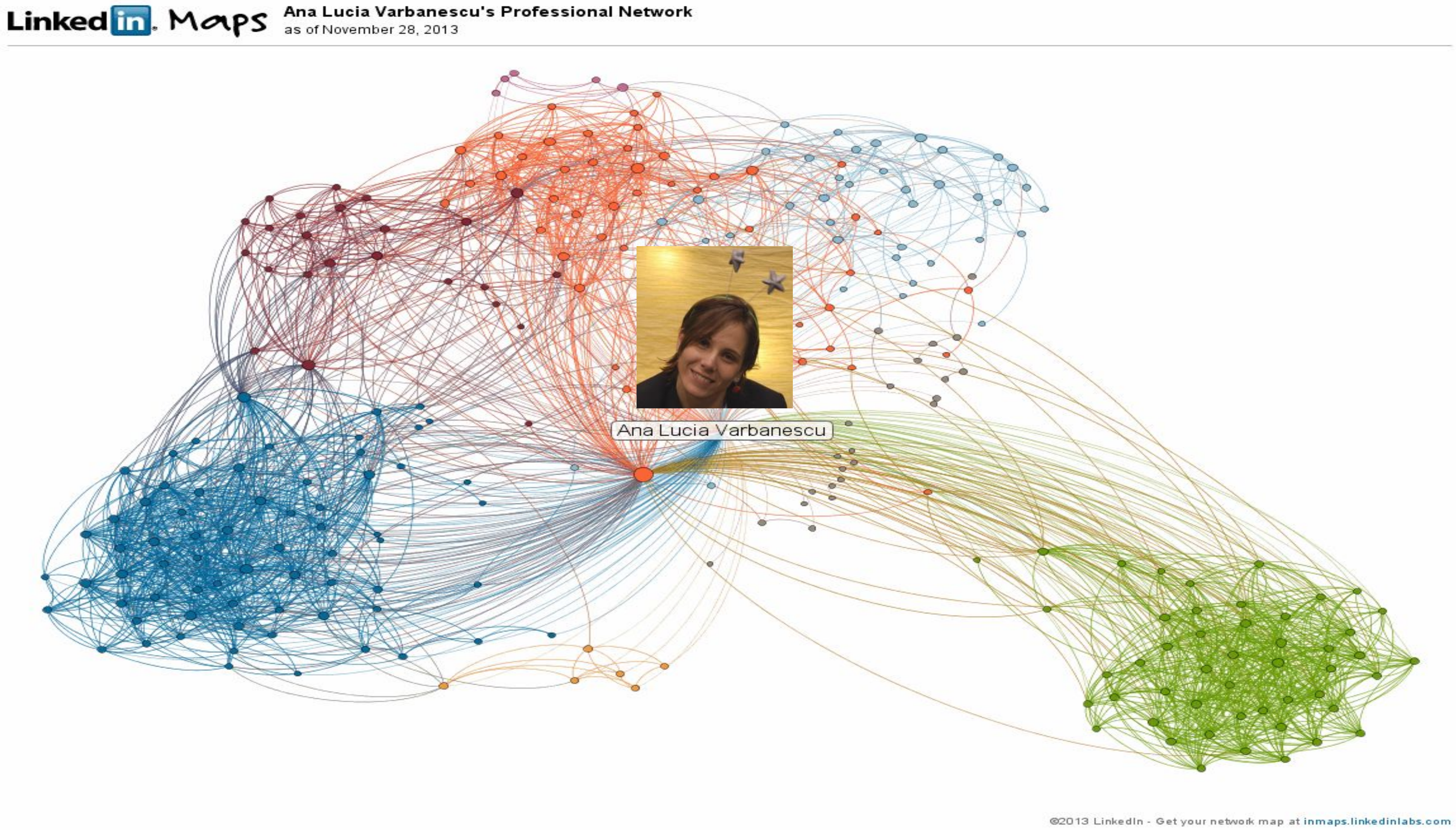
LinkedIn

Brain network

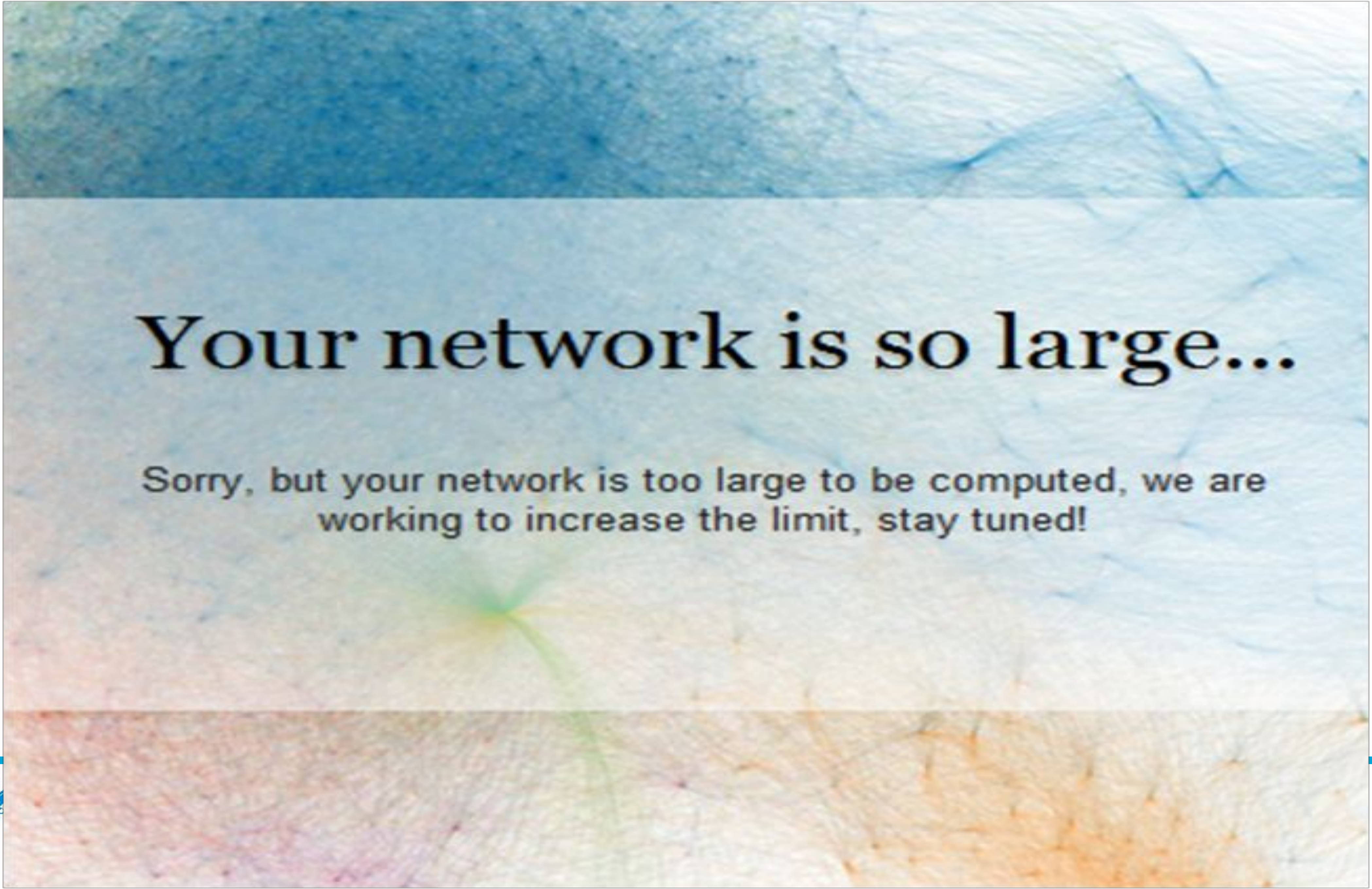
~100 billion neurons
~100 trillion connections

VU

Sources: Smith, CHI’10; Blog webpage; Gigandet et al., PLoS ONE 3(12)]



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

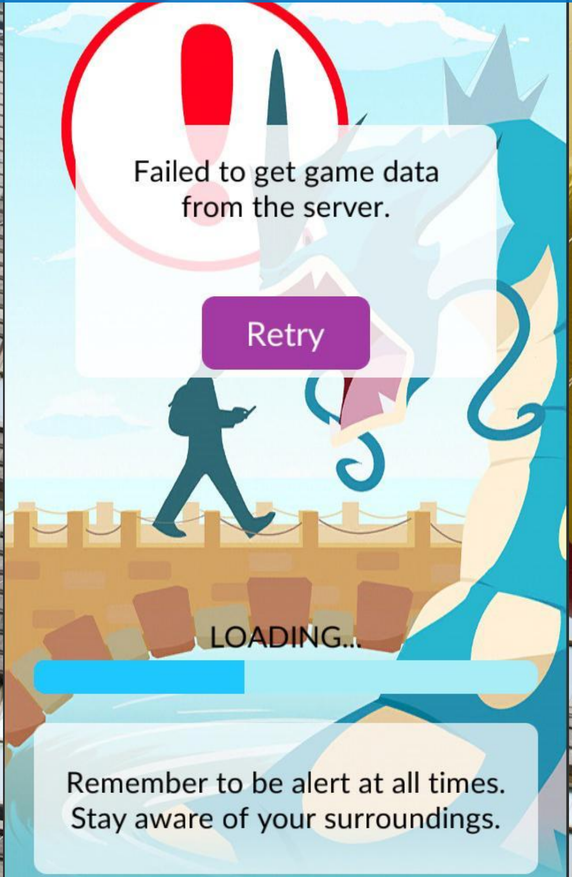


Your network is so large...

Sorry, but your network is too large to be computed, we are working to increase the limit, stay tuned!

PHENOMENON: FAILURES IN COMPUTER ECOSYSTEMS

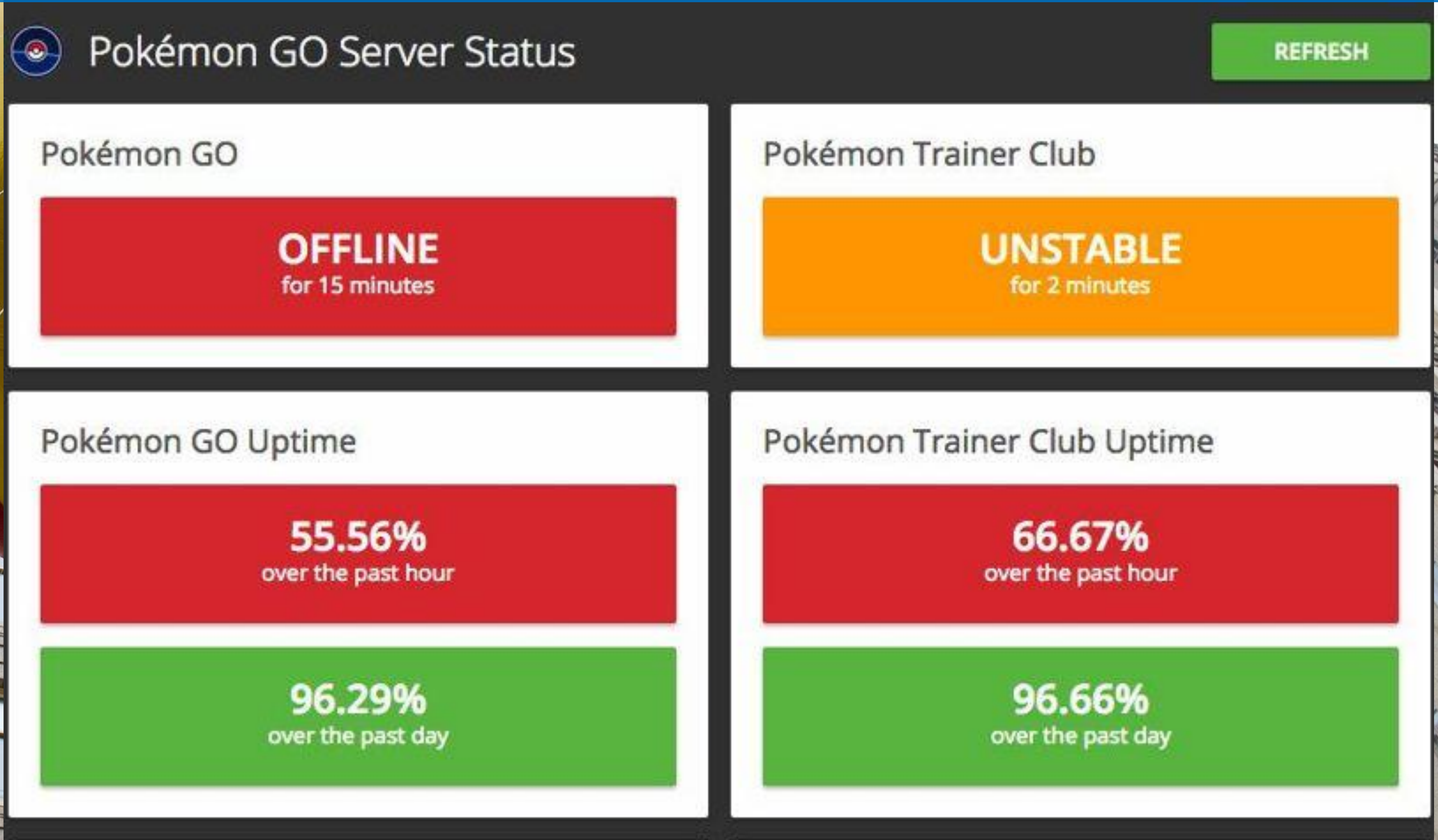
UNCOVERING THE PRESENCE OF FAILURES



Failed to get game data from the server.

Remember to be alert at all times. Stay aware of your surroundings.

Source: gamenewstoday.com



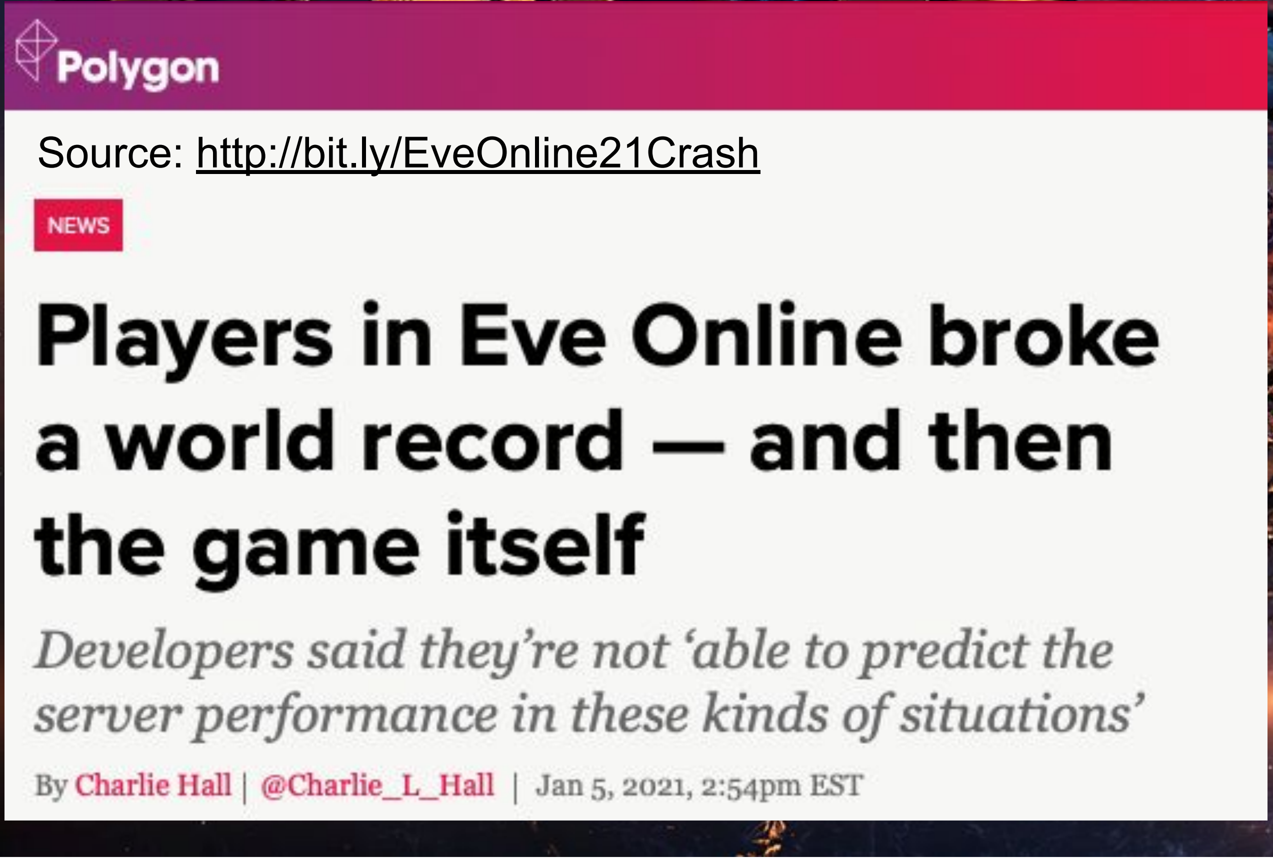
Pokémon GO Server Status

Pokémon GO	Pokémon Trainer Club
OFFLINE for 15 minutes	UNSTABLE for 2 minutes

Pokémon GO Uptime	Pokémon Trainer Club Uptime
55.56% over the past hour	66.67% over the past hour
96.29% over the past day	96.66% over the past day

PHENOMENON: PERFORMANCE IN CLOUD SERVICES

UNCOVERING THE PRESENCE OF PERFORMANCE ISSUES, EVEN LEADING TO CRASHES



Polygon

Source: http://bit.ly/EveOnline21Crash

NEWS

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

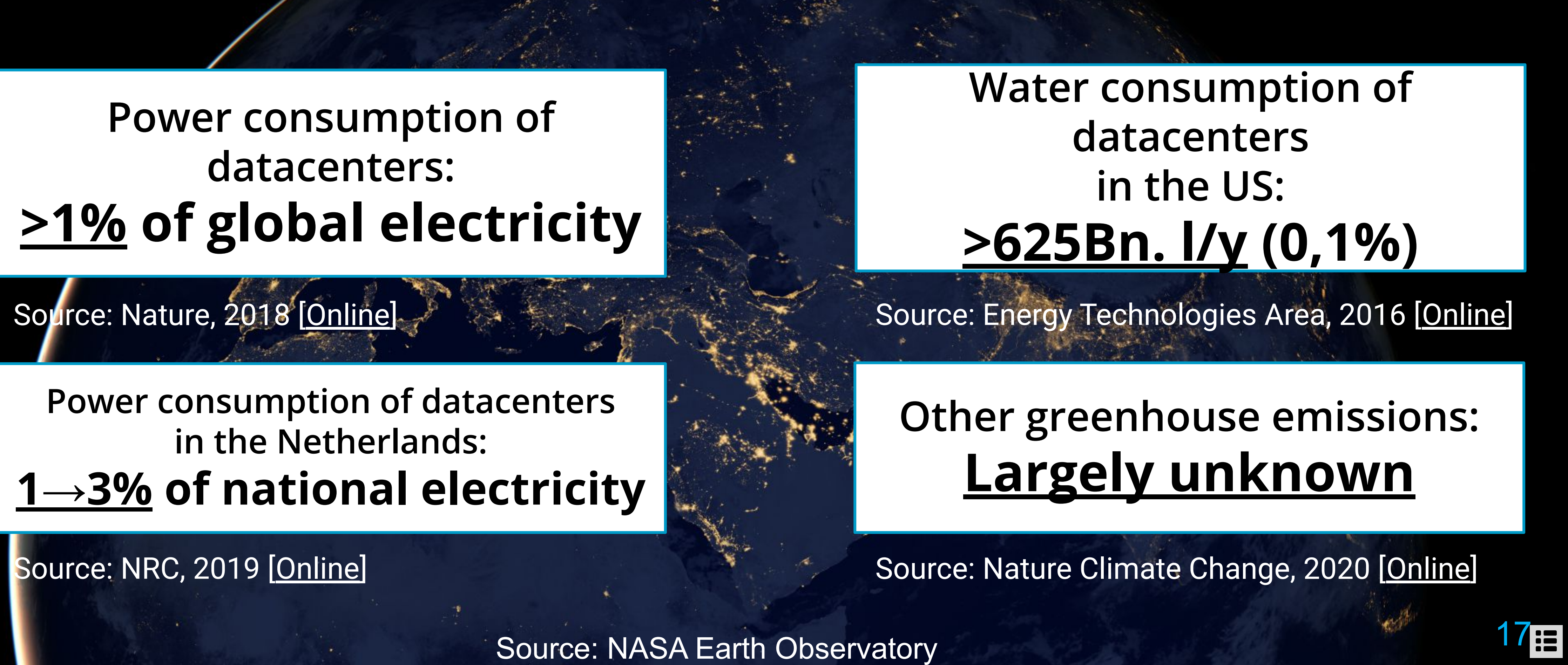
Developers said they're not 'able to predict the server performance in these kinds of situations'

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

Source: Razorien/CCP Games

PHENOMENON: IT SUSTAINABILITY

UNCOVERING THE USE OF ENERGY AND WATER, THE IMPACT ON CLIMATE



Power consumption of datacenters:
>1% of global electricity

Source: Nature, 2018 [Online]

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

Source: Energy Technologies Area, 2016 [Online]

Power consumption of datacenters in the Netherlands:
1→3% of national electricity

Source: NRC, 2019 [Online]

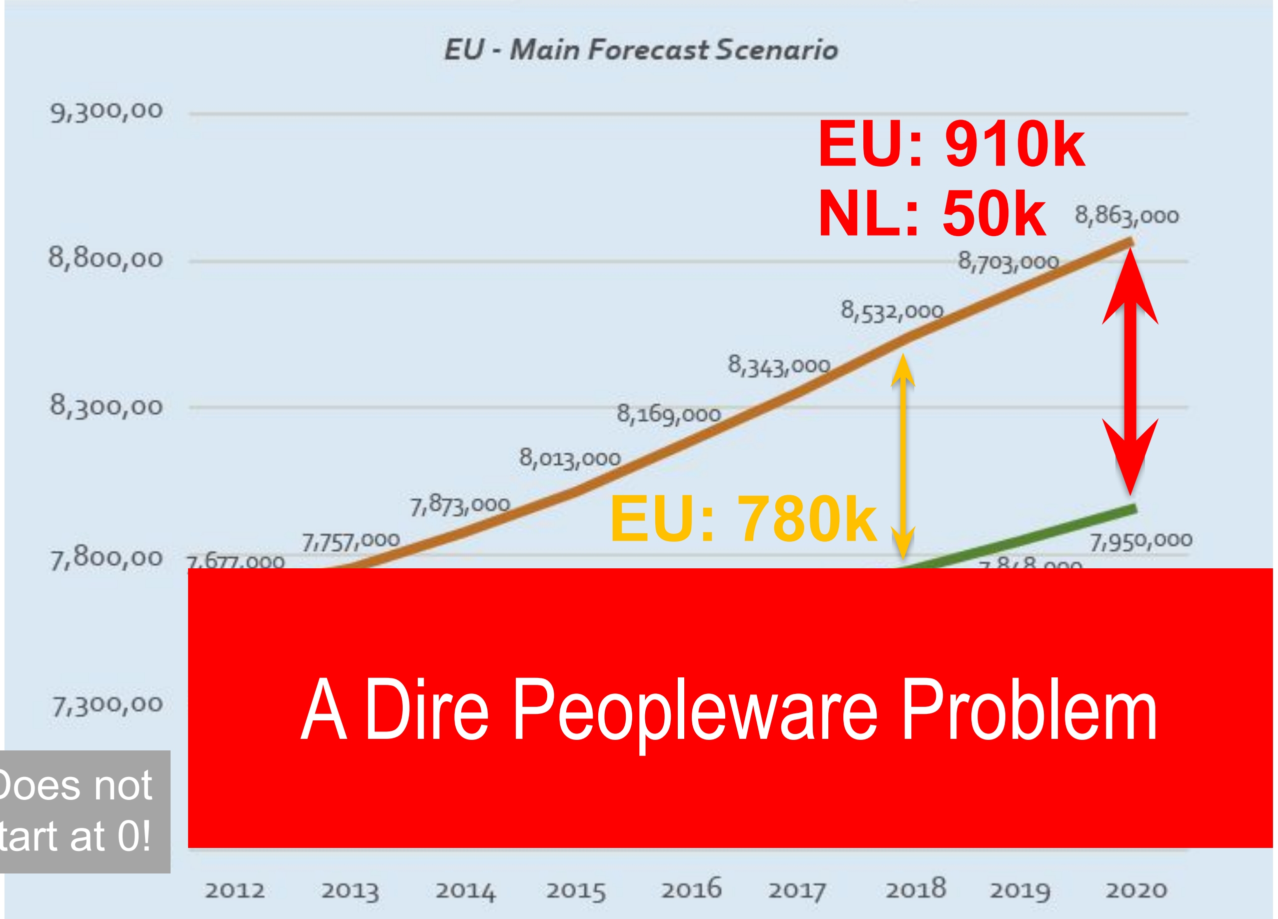
Other greenhouse emissions:
Largely unknown

Source: Nature Climate Change, 2020 [Online]

Source: NASA Earth Observatory

PHENOMENON: FEW CAN OPERATE COMPLEX IT ECOSYSTEMS

THE WORKFORCE GAP, IN THE NETHERLANDS & IN EUROPE



EU: 910k
NL: 50k

EU: 780k

A Dire Peopleware Problem

Does not start at 0!

Source: e-Skills for Jobs in Europe, 2014

THIS TALK: MASSIVIZING = LET’S THINK ECOSYSTEMS!

WE TAKE A HOLISTIC VIEW, BASED ON COMPUTER ECOSYSTEMS

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.

3 COMBINE SCIENCE, DESIGN, & ENGINEERING

Idea:
Meaningful discovery requires a mix of experimental science, design, and engineering.

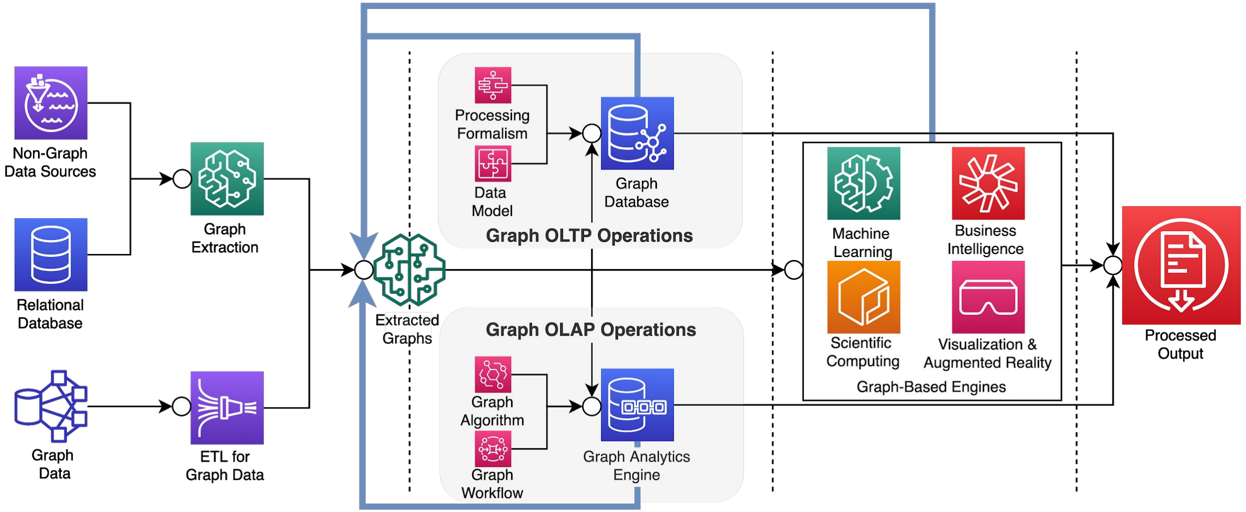
These are very different activities!

LEONARDO’S APPROACH: ADDRESS COMPLEXITY WITH MIX OF SCIENCE, DESIGN, AND ENGINEERING (AND ART, BUT HE WAS PEERLESS!)

Source: Leonardo da Vinci (15th-16th century), Jean Paul Richter (1883) The Literary Works of Leonardo da Vinci.

ENGINEERING COMPLEX WORKFLOWS

THE NEED FOR SPEED IN GRAPH PROCESSING FOR COVID-19 PREVENTION



Social network □ Epidemiology

Who is a super-spreader? Where? Why?
How would the spreading evolve?
Which measures/policies can curb spreading?

Biological network □ Drug synthesis

Which diseases are similar? Which genes are related to this disease?
Which inhibitors could work? Which drugs could be targeted/synthesized to cure COVID-19?

Source: Sakr, Bonifati, Voigt, Iosup. The Future is Big Graphs! A Community View on Graph Processing Systems. Dagstuhl Seminar 19491 Big Graph Processing Systems and CACM article.

COMPUTER SCIENCE IMPACTING SCIENCE

MANY GRAPH DATASETS, MANY ALGORITHMS □ 10S OF SPECIALIZED PLATFORMS

ORACLE PGX

Intel Graphmat

hadoop 2

NetSysLab

TOTEM

APACHE GIRAPH

Neo4j the graph database

IBM System G

medusa-gpu

Medusa: Simplified Graph Processing on GPUs

GraphLab

PROJECT PEGASUS

HyperGraphDB

Trinity

StratoSphere Above the Clouds

GraphX

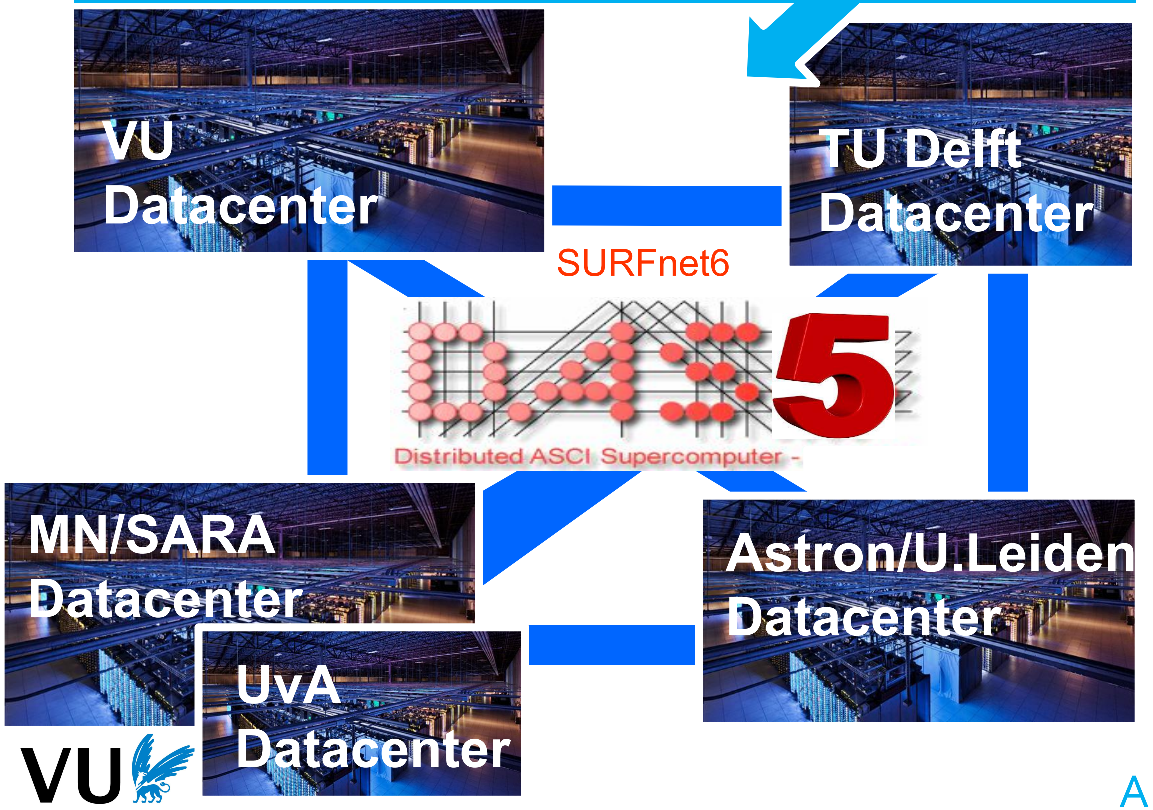
How well do graph processing platforms perform?

VU

EXPERIMENTAL METHODS OF DISCOVERY

UNIQUE OPPORTUNITY: WE DRINK OUR OWN CHAMPAGNE (*IN VIVO*)!

Our Prototypes (*in phyisico/in vitro*)



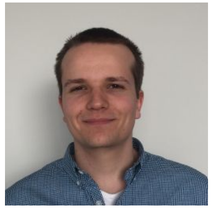
Alex Uta

Georgios Andreadis

Fabian Mastebroek

acheendra

Maria Voinea


Laurens Versluis

Alexey Ilyushkin

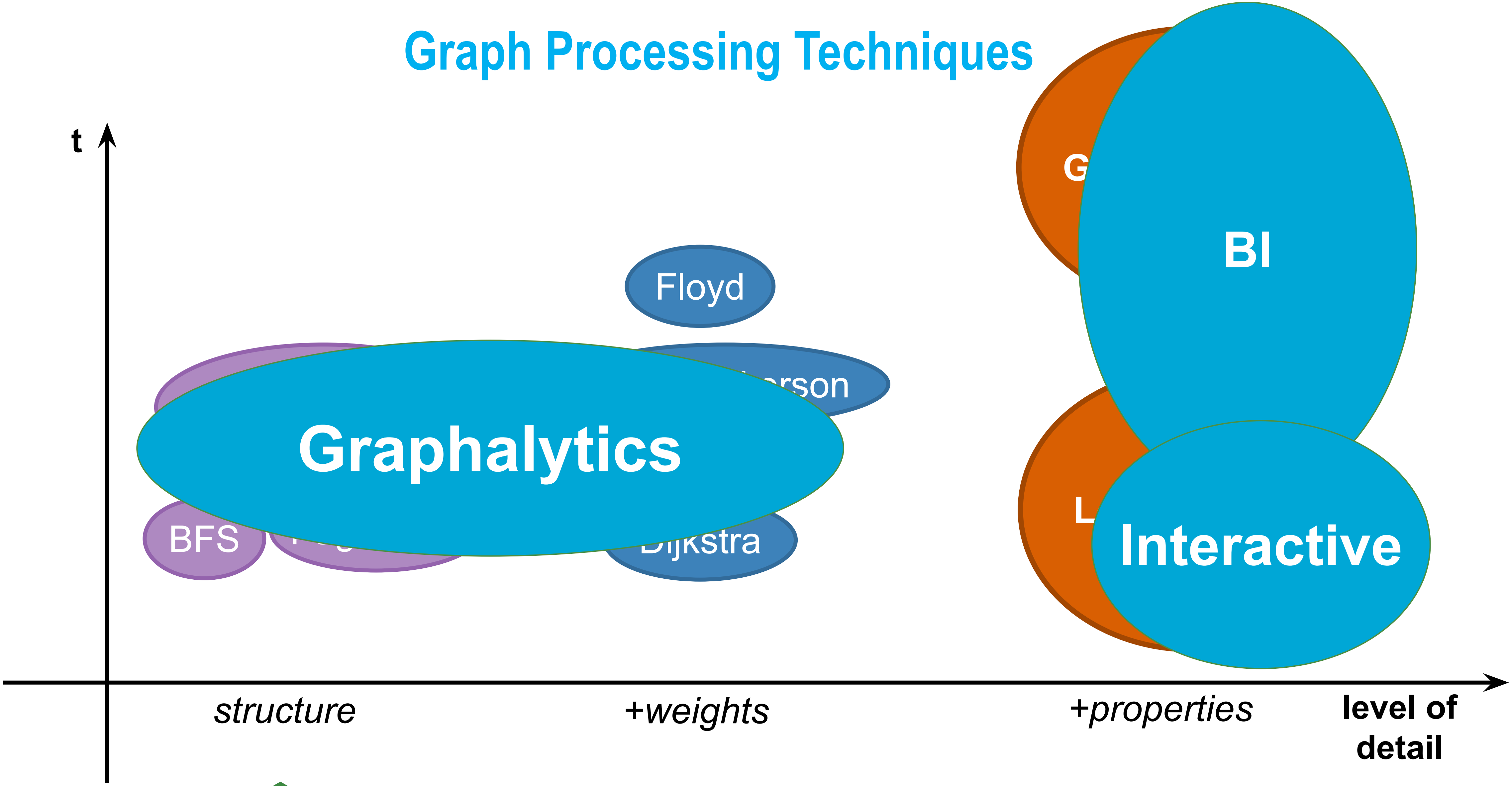
+ amazon web services™

We also use clouds

+ OpenDC

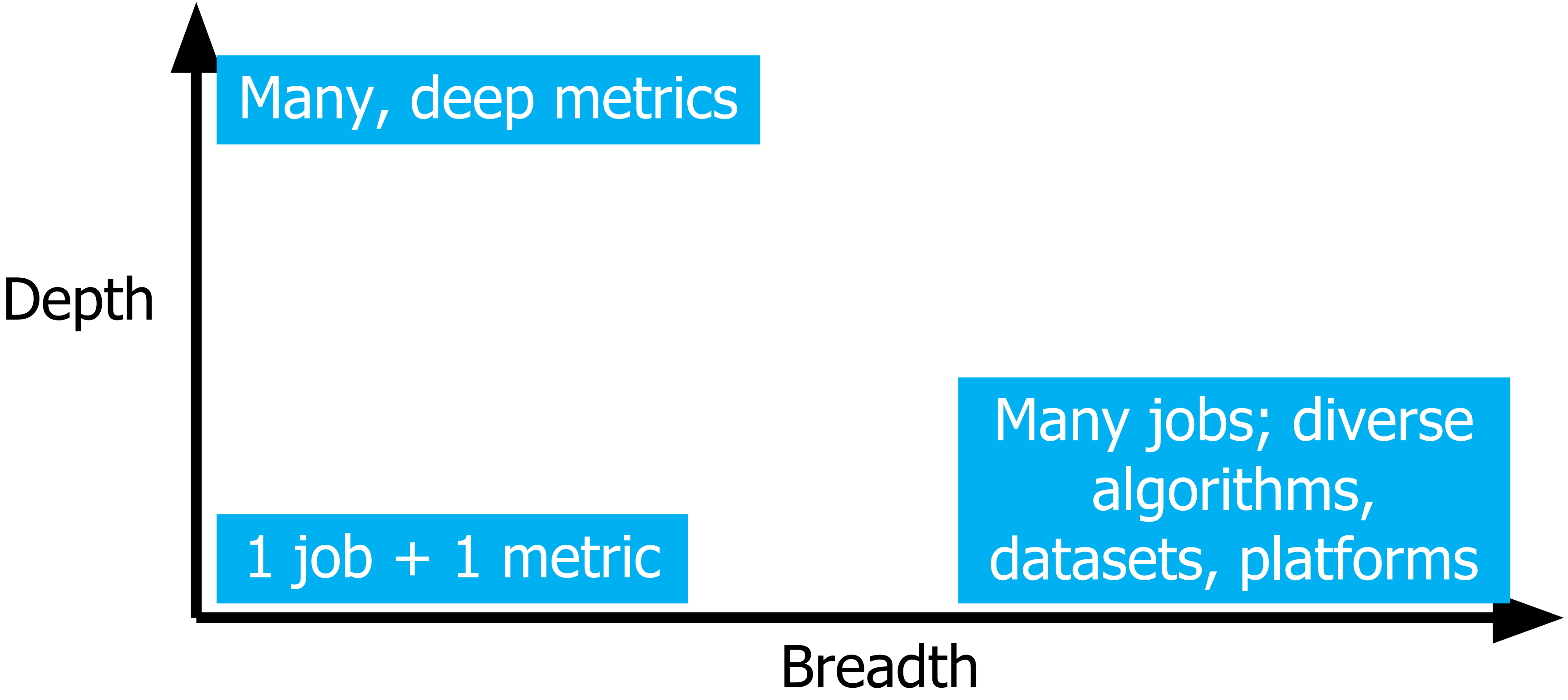
And simulators (*in silico*)

Graph Processing Techniques



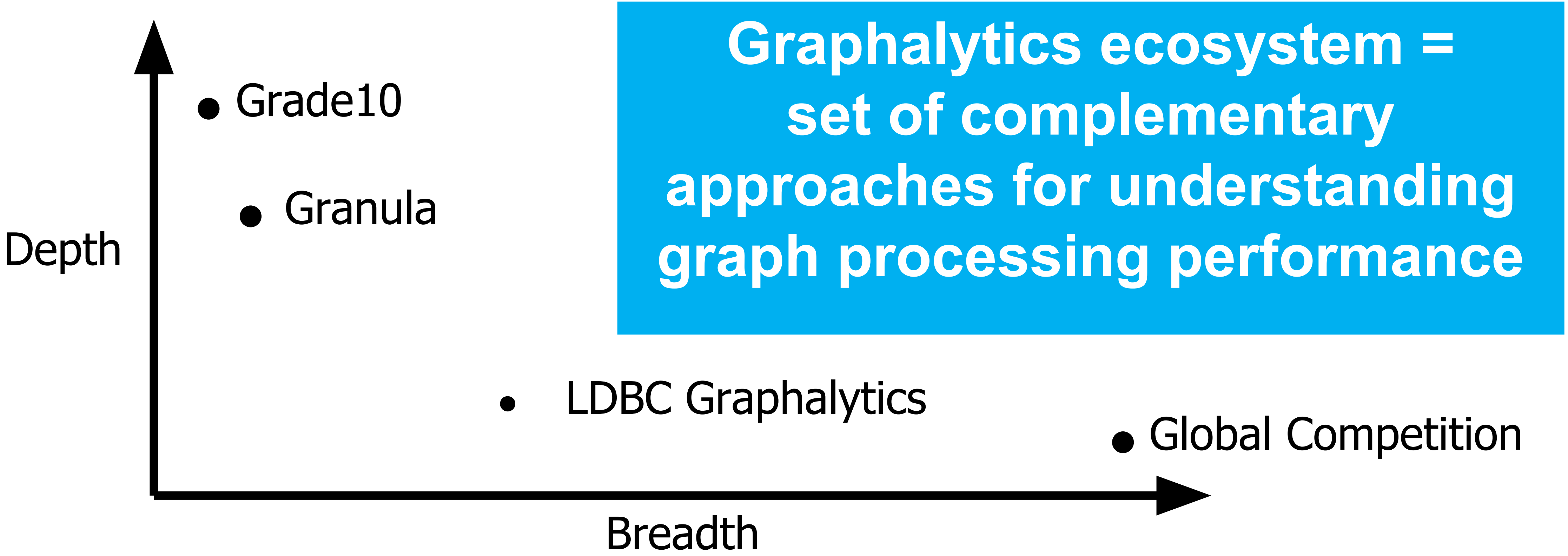
Slide by Gábor Szárnyas

There is More to Understanding Performance



The Graphalytics Ecosystem

Graphalytics ecosystem = set of complementary approaches for understanding graph processing performance



- Grade10
- Granula
- LDBC Graphalytics
- Global Competition

Depth

Breadth

AUTOMATED TESTING FOR DISTRIBUTED ECOSYSTEMS?

LDBC GRAPHALYTICS: BENCHMARKING LEADING TO DISCOVERY

Ahmed Musaafir

LDBC *The graph & RDF benchmark reference*

Graphalytics

> Benchmark, 7+ years engineering
> Many classes of algorithms used in practice
> Diverse real and auto-gen datasets
> Diverse experiments, representative for practice
> Renewal process to keep the workload relevant
> Automated comparison of many platforms, community-driven and industrial
> Automated global competition, manual auditing

ORACLE intel Georgia Tech IBM HUAWEI CWI UPC

- Community endorsed:

graphalytics.org

- Surprising finding, selected:

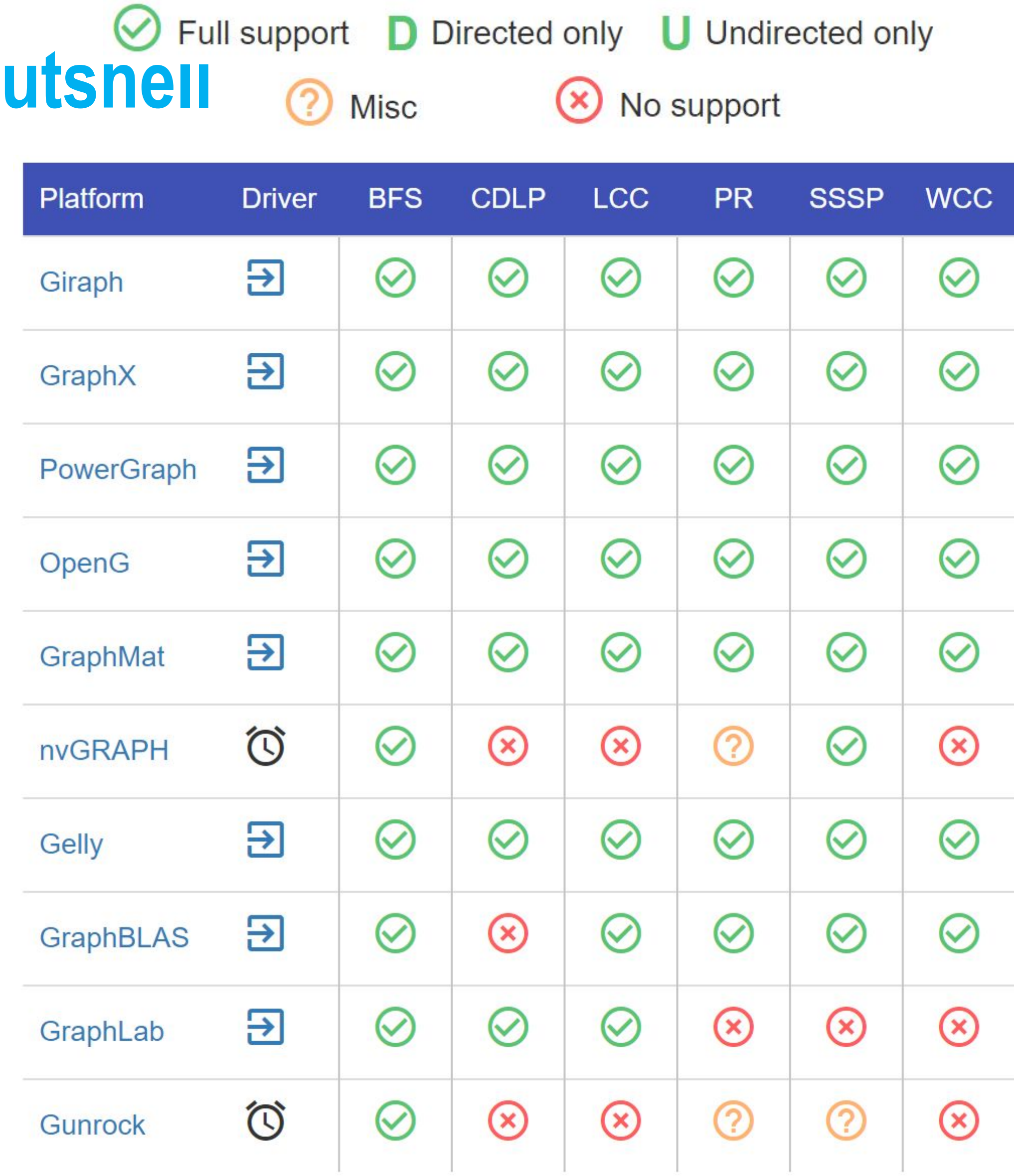
Performance: orders of magnitude difference due to each of platform, algorithm, dataset, and hardware

[Iosup et al. arXiv'21] [Iosup et al. PVLDB'16] [Guo et al. CCGRID'15] [Guo et al. IPDPS'14]

https://graphalytics.org

Graphalytics in a Nutshell

- An LDBC benchmark
- Advanced benchmarking harness
- Many classes of algorithms
- Diverse real and synthetic datasets
- Diverse set of experiments
- Renewal process to keep workload relevant
- Enables comparison of many platforms, community-driven and industrial

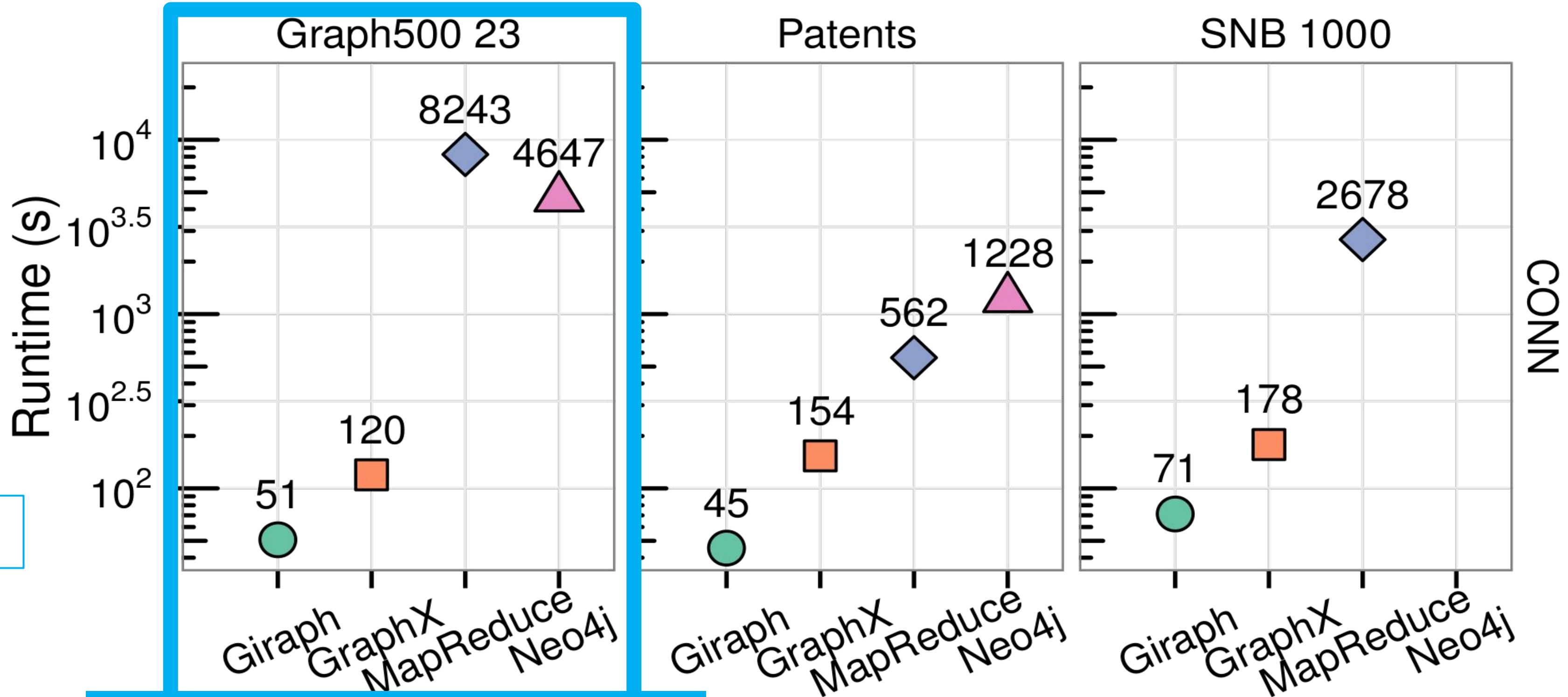


✓ Full support D Directed only U Undirected only

? Misc ✗ No support

Platform	Driver	BFS	CDLP	LCC	PR	SSSP	WCC
Giraph	⇥	✓	✓	✓	✓	✓	✓
GraphX	⇥	✓	✓	✓	✓	✓	✓
PowerGraph	⇥	✓	✓	✓	✓	✓	✓
OpenG	⇥	✓	✓	✓	✓	✓	✓
GraphMat	⇥	✓	✓	✓	✓	✓	✓
nvGRAPH	⏰	✓	✗	✗	?	✓	✗
Gelly	⇥	✓	✓	✓	✓	✓	✓
GraphBLAS	⇥	✓	✗	✓	✓	✓	✓
GraphLab	⇥	✓	✓	✓	✗	✗	✗
Gunrock	⏰	✓	✗	✗	?	?	✗

Runtime: The Platform Has Large Impact

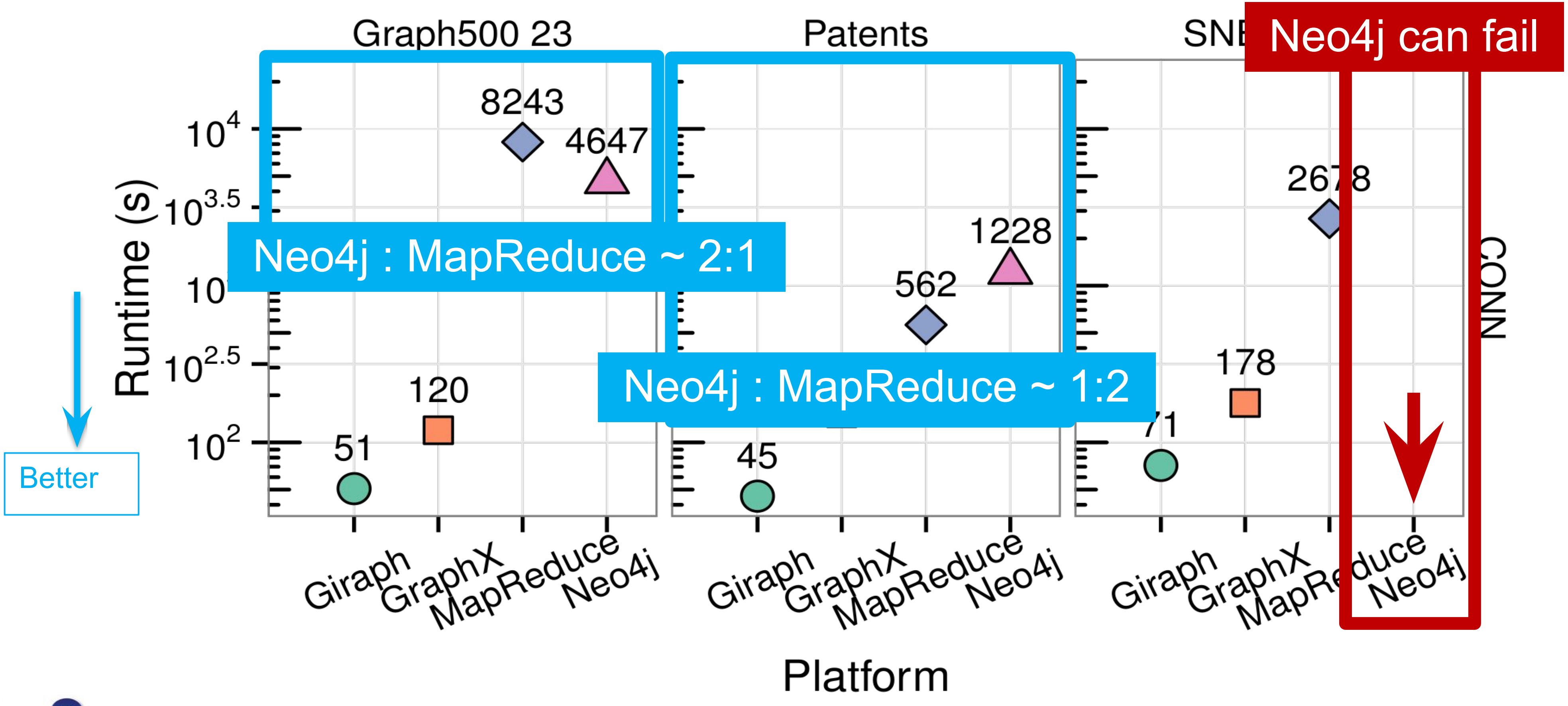


Better

2 orders of magnitude difference due to

M. Capota et al., Graphalytics: A Big Data Benchmark for Graph-Processing Platforms. SIGMOD GRADES 2015

Runtime: The Dataset Has Large Impact



Neo4j can fail

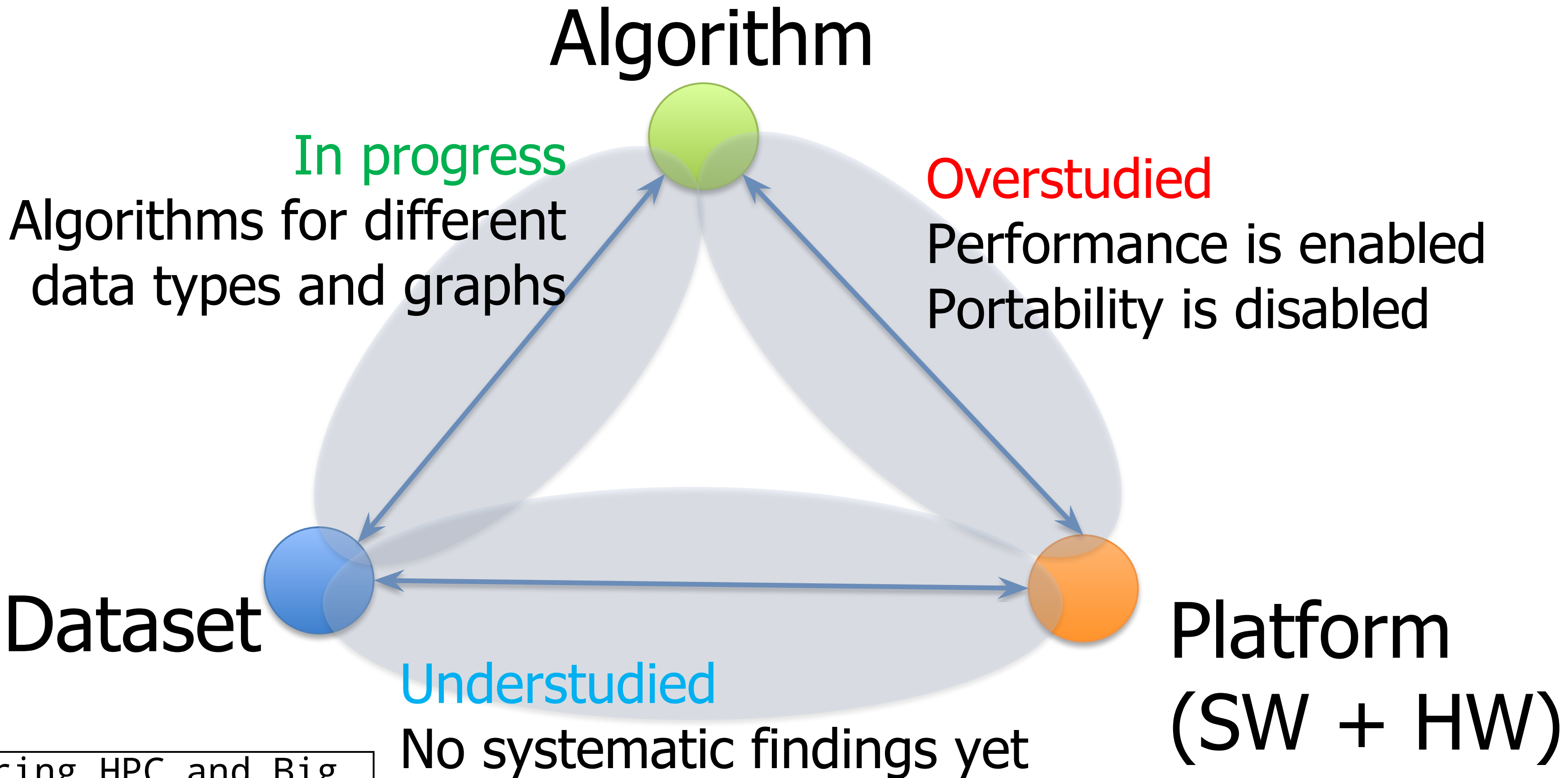
Neo4j : MapReduce ~ 2:1

Neo4j : MapReduce ~ 1:2

M. Capota et al., Graphalytics: A Big Data Benchmark for Graph-Processing Platforms. SIGMOD GRADES 2015

Introduced by
Ana Lucia
Varbanescu.

The Platform-Algorithm-Dataset (PAD) Triangle for Performance Engineering of Graph-Processing Systems



In progress
Algorithms for different data types and graphs

Overstudied
Performance is enabled
Portability is disabled

Understudied
No systematic findings yet
Must be correlated with the algorithm

Capota et al.’15

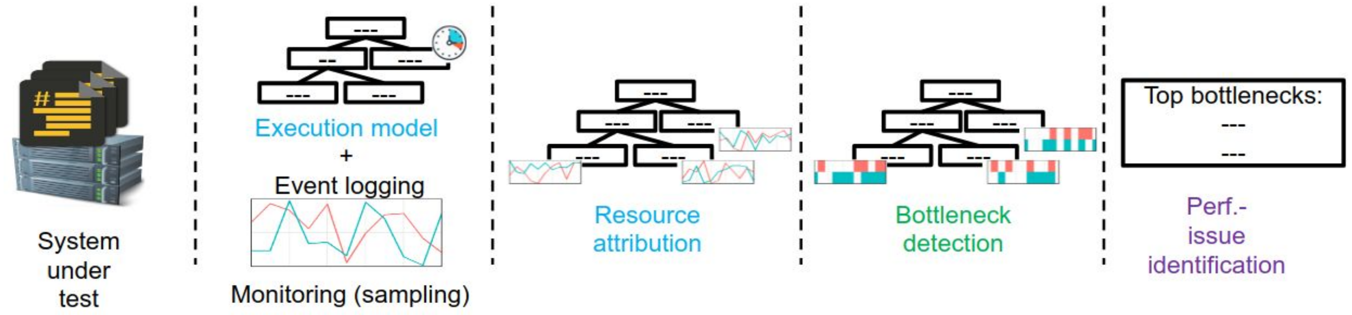
A. Uta et al., Exploring HPC and Big Data Convergence. CLUSTER 2018

LOCALIZATION OF BOTTLENECKS □ PERF. ISSUES

GRADE10: MODELING LEADS TO DISCOVERY

LDBC The graph & RDF benchmark reference

- Graphalytics Grade10:
  > Automated bottleneck detection
  > Automated identification of performance issues



Multi-stage process, works in ecosystem

APACHE GIRAPH

Tim Hegeman

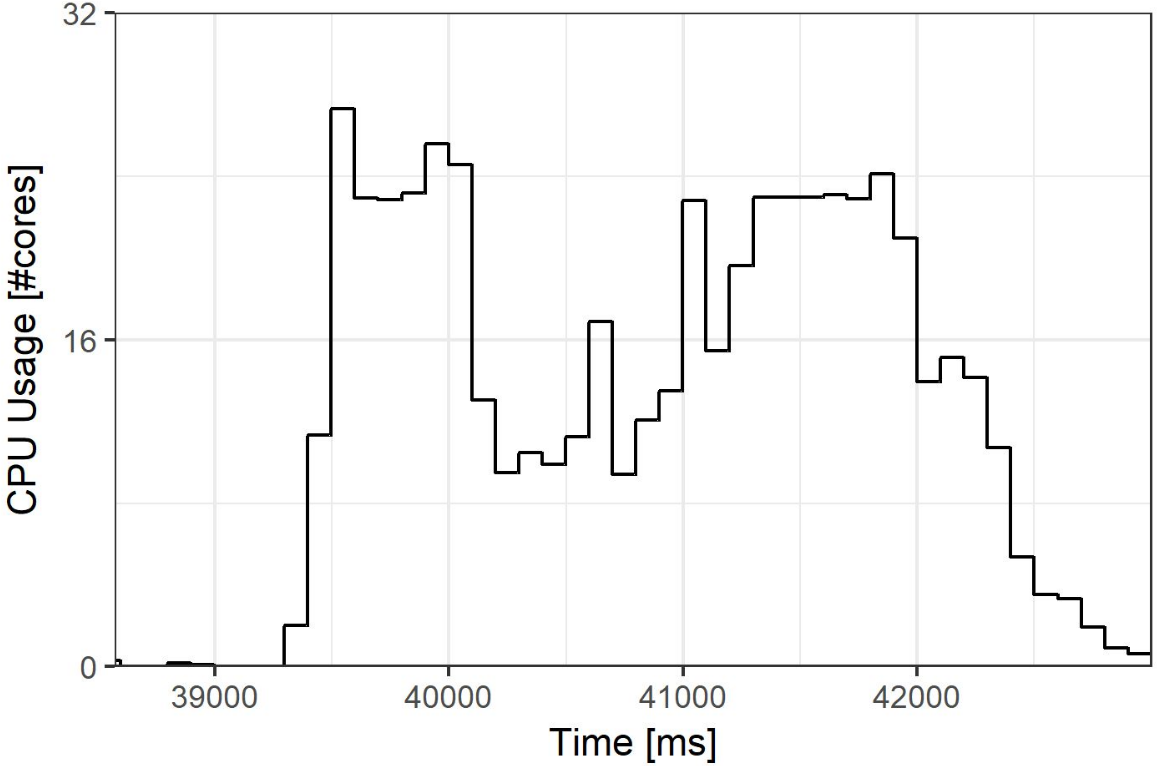
- Without Grade10:

No bottleneck at all

- With Grade10:

Always bottleneck
Can explain causes:
+ Message queue full
+ Garbage collector
+ CPU
+ Others

Grade10 Result: Analysing a Giraph Job



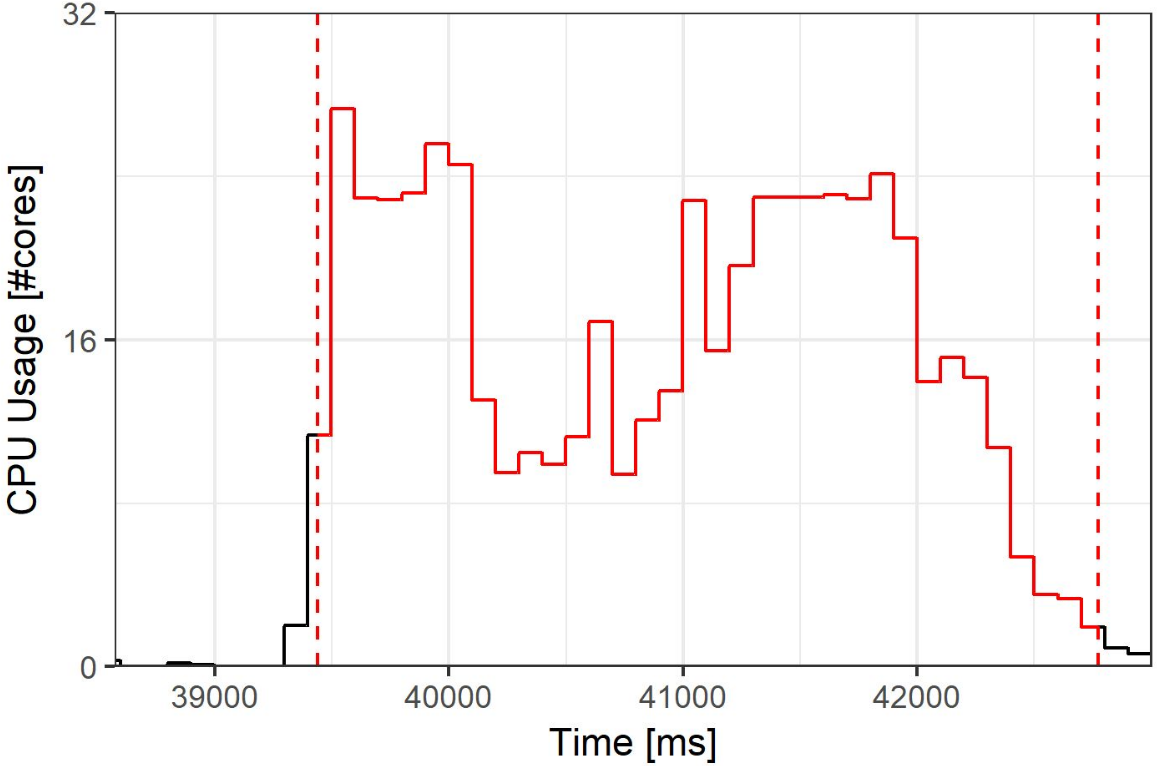
WorkerSuperstep

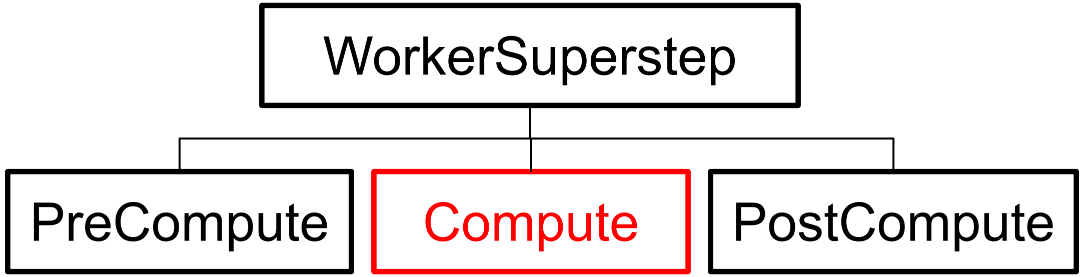
CPU usage < 32 cores (100%), so no bottleneck

... yet

Hegeman et al., Grade10: A Framework for Performance

Grade10 Result: Analysing a Giraph Job




WorkerSuperstep

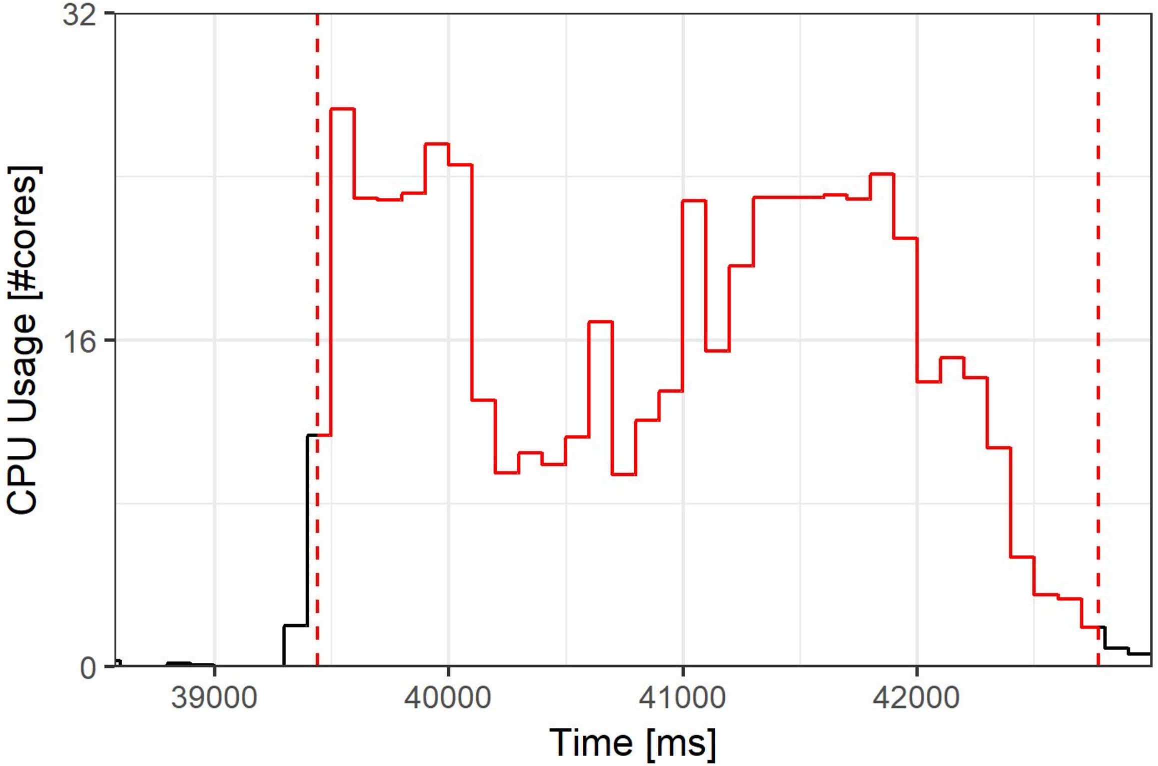
PreCompute

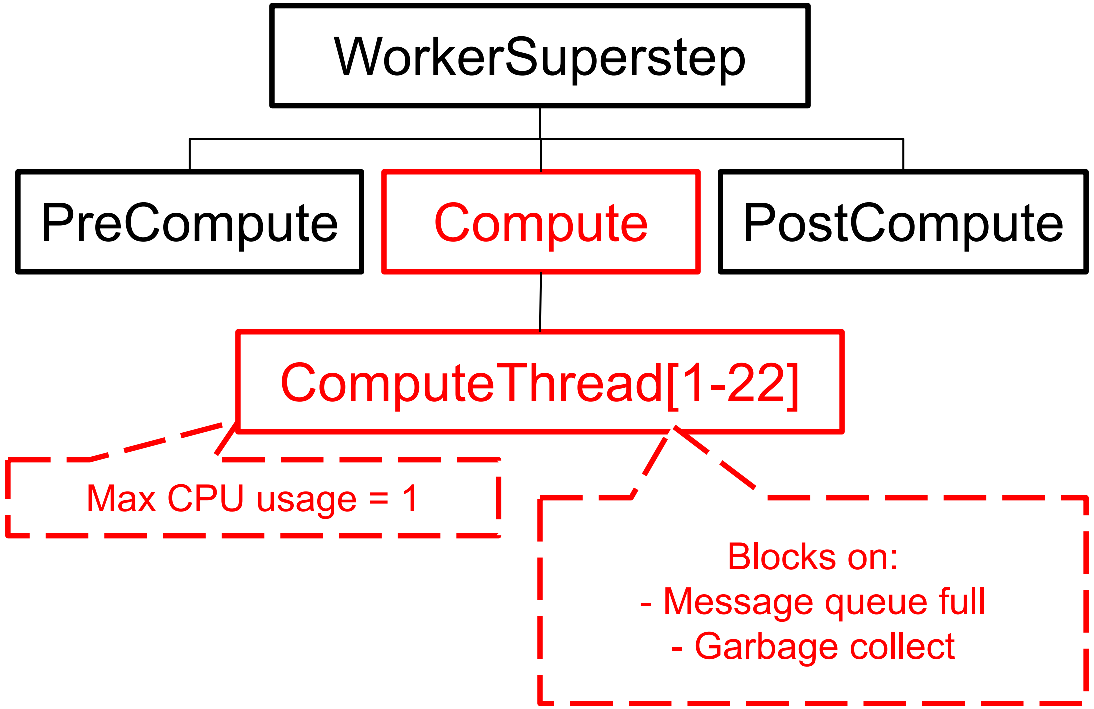
Compute

PostCompute

Hegeman et al., Grade10: A Framework for Performance

Grade10 Result: Analysing a Giraph Job



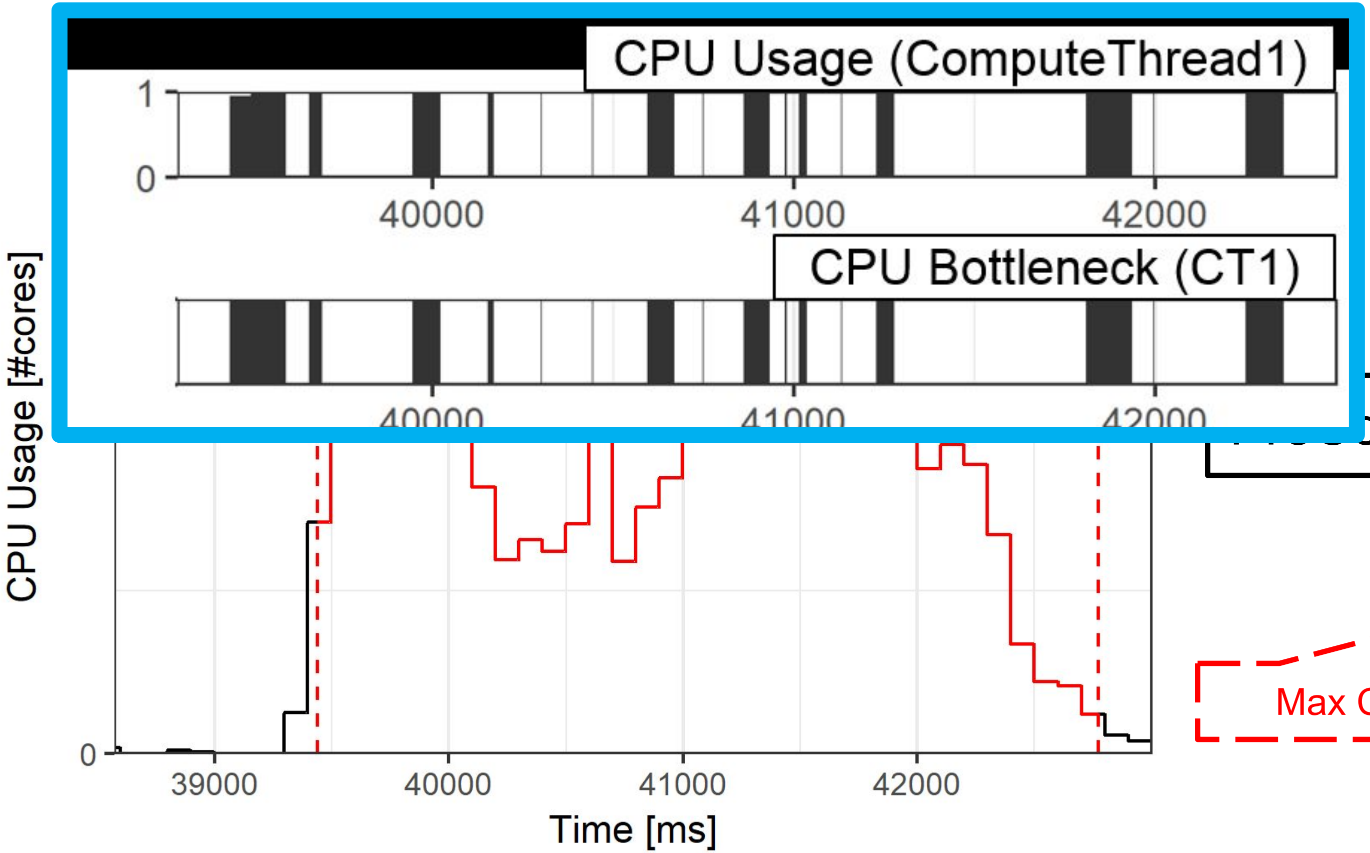


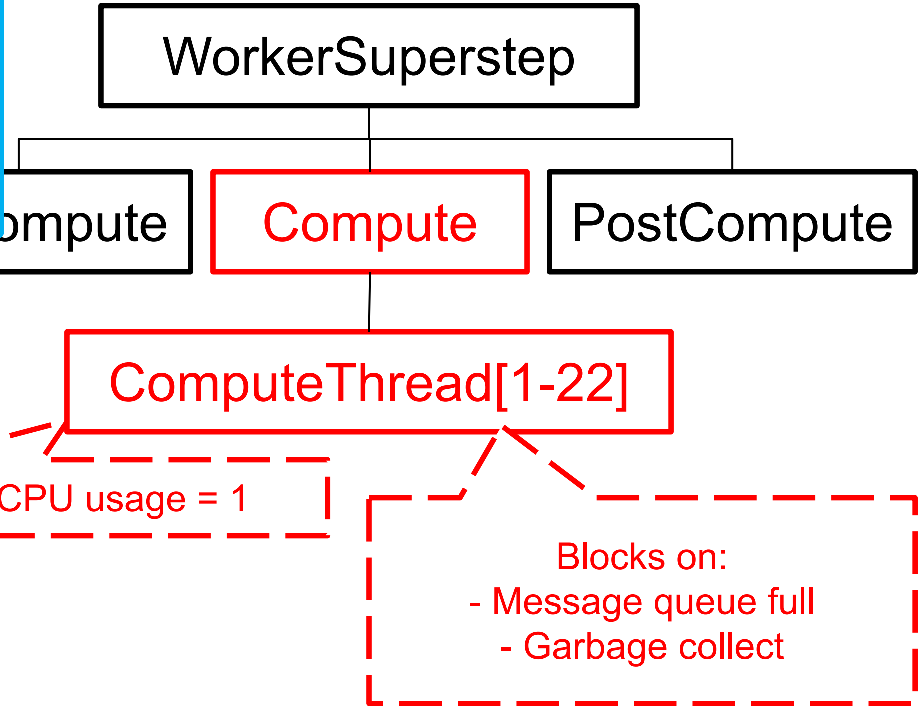
Max CPU usage = 1

Blocks on:
- Message queue full
- Garbage collect

Hegeman et al., Grade10: A Framework for Performance

Grade10 Result: Analysing a Giraph Job



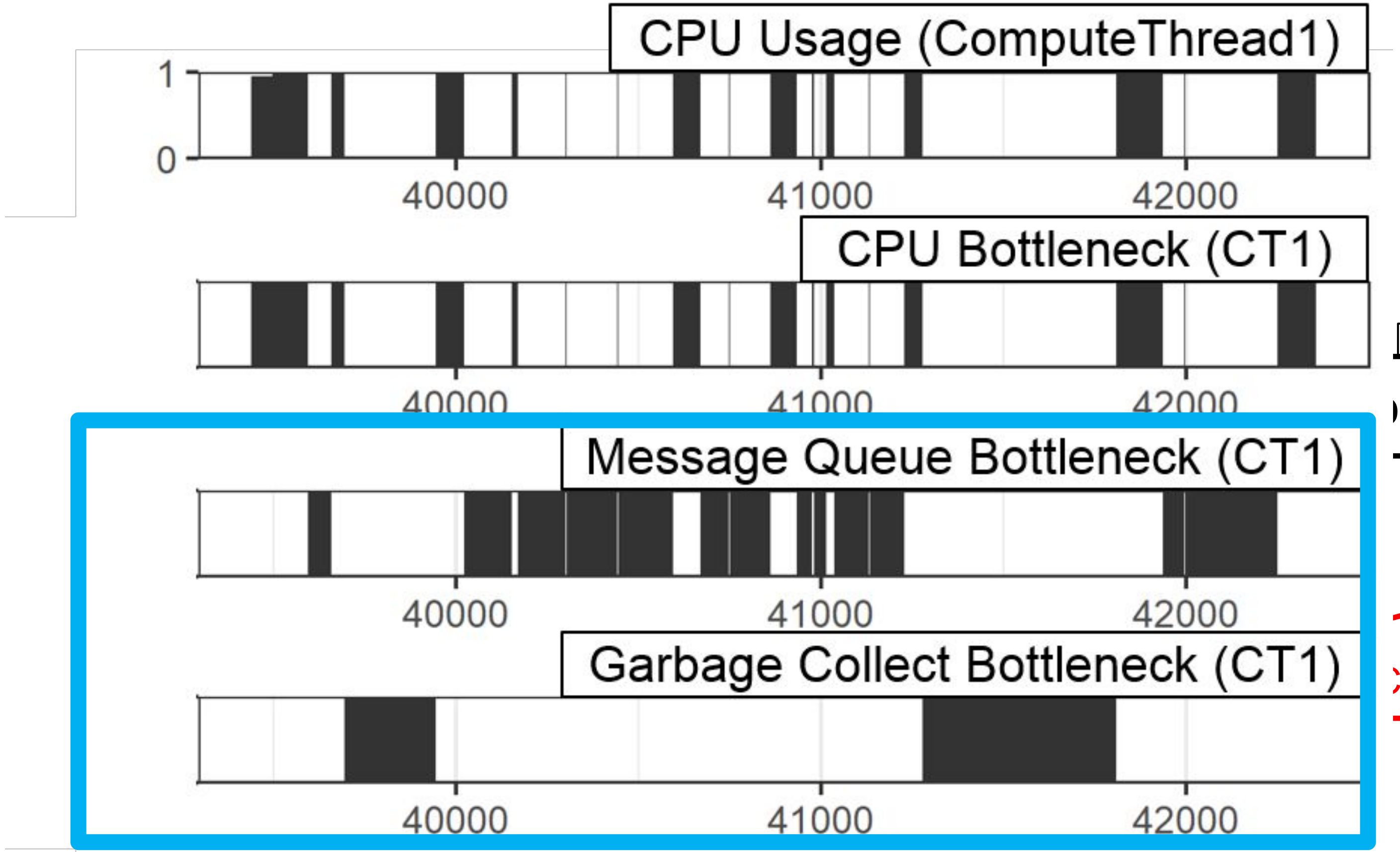


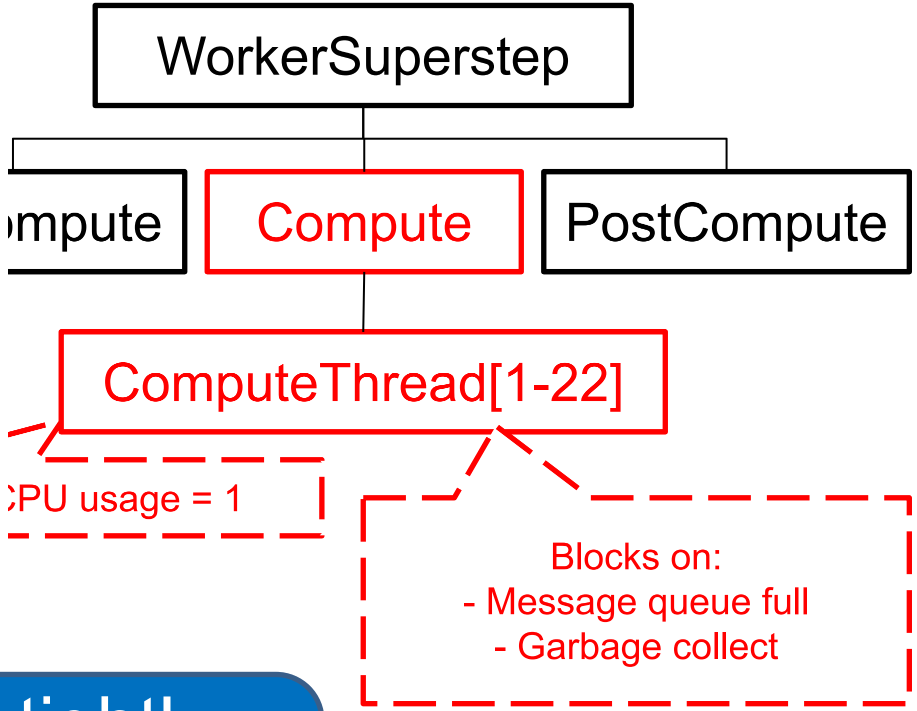
Max CPU usage = 1

Blocks on:
- Message queue full
- Garbage collect

Hegeman et al., Grade10: A
Framework for Performance

Grade10 Result: Analysing a Giraph Job





WorkerSuperstep

mpute

Compute

PostCompute

ComputeThread[1-22]

PU usage = 1

Blocks on:
- Message queue full
- Garbage collect

V

For one class of eco-systems, under tightly controlled experimental conditions, with careful analysis □ bottlenecks!

The interplay of experimental science, design, and engineering leads to important results (end Part I)

1. Science, design, and engineering intertwined
2. Experimentation in vivo, in vitro, (in silico)
3. Benchmarking reveals graph processing has large design space & complex trade-offs
4. Remarkable finding: performance as function of hardware and software platform, data, algorithm
5. New systems through design and engineering, but there must exist a better way than bottleneck by bottleneck

VRIJE UNIVERSITEIT AMSTERDAM

A new science, of complex, smart computer ecosystems

(operational simplicity for the user)

4

This talk <<

High-level approaches >>

AN ANALOGY: MASSIVIZING CLIMATE SCIENCE

TAKE A HOLISTIC VIEW, BASED ON COUPLED NATURAL SYSTEMS

Can be understood only with coupled models

Climatologist Bjorn Stevens. Source: HPCWire

* In climate science, issues are often linked.
The same occurs in massive computer (eco)systems.

What is a Distributed Ecosystem?
Our definition

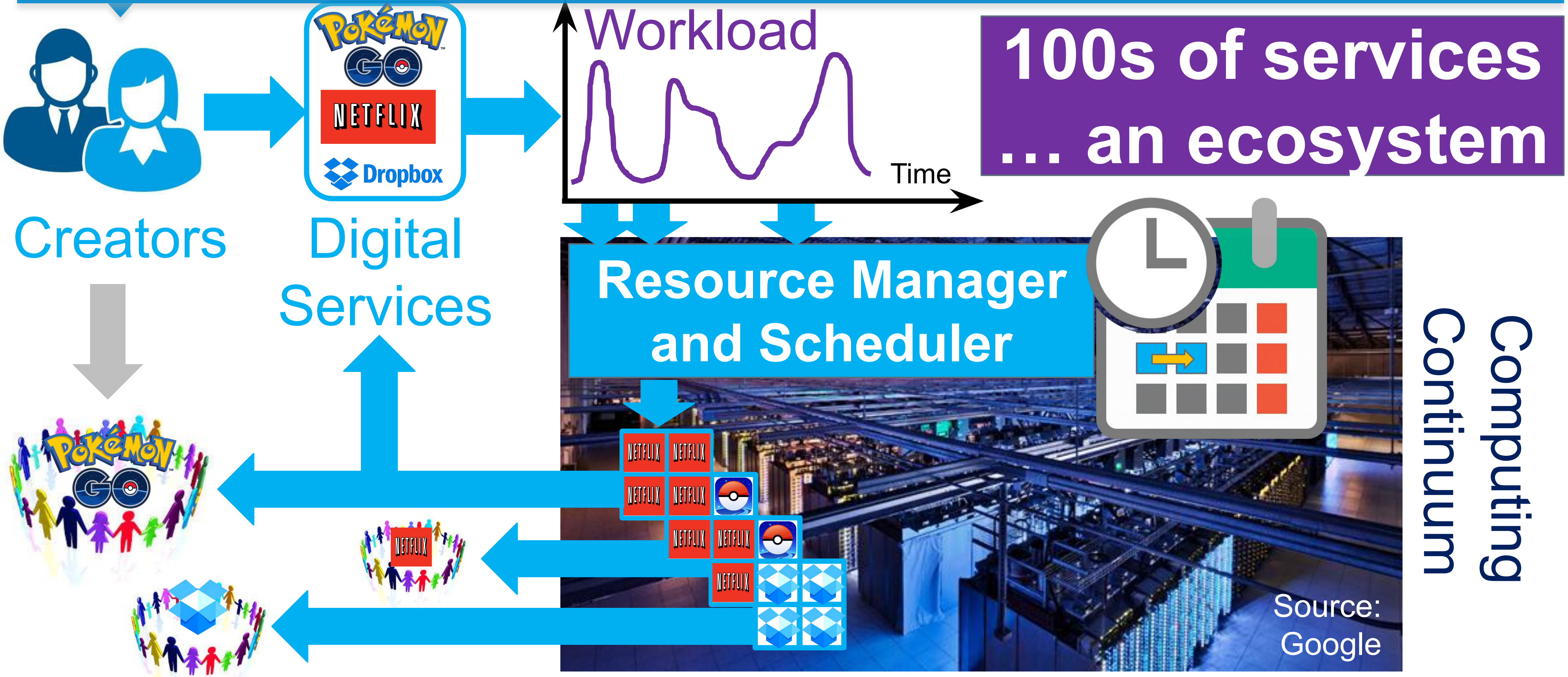
DISTRIBUTED ECOSYSTEMS, OUR DEFINITION

1. Set of 2+ constituents, often heterogeneous
2. Each constituent is a system or an ecosystem (recursively)
3. Constituents are autonomous, cooperative or in competition
4. Ecosystem structure and organization ensure responsibility
    1. Completing functions and providing services
    2. Providing desirable non-functional properties
    3. Fulfill agreements with both operators and clients, clients in the loop
5. Long and short-term dynamics occur in the ecosystem

Iosup et al., Lecture Notes in Distributed Systems, Section 1.1.1

Iosup et al., Massivizing Computer Systems, ICDCS 2018. [Online]

ECOSYSTEM = SERVICES + COMPUTING + SMARTS + GOALS



100s of services … an ecosystem

Source: Google

Extreme Automation, Performance, Dependability, Sustainability

High-level, theoretical approaches

5

(our theories are often frameworks, designs, etc.)

Ref Archi >

Ref Archi, Sched >

RM&S Framework >

Metrics Framework >

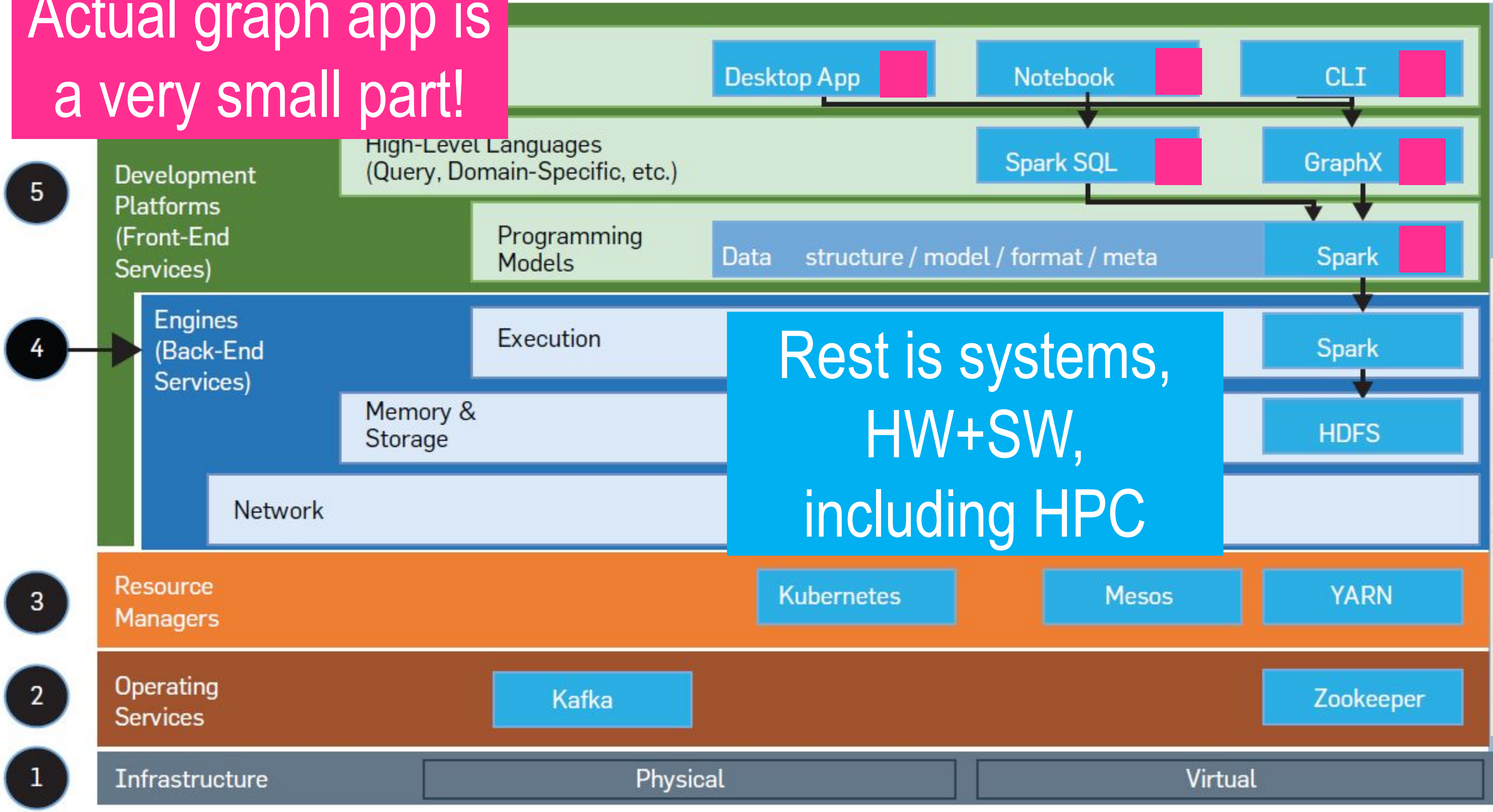
CWL Framework >

A new science <<

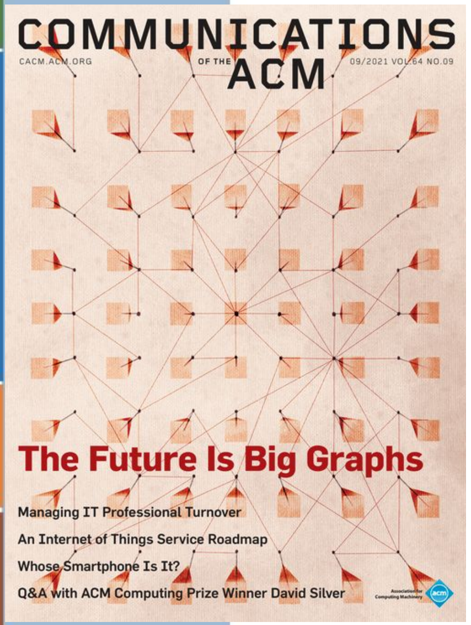
Detailed approaches >>

A Reference Architecture for Big Graph Processing

Actual graph app is a very small part!



Rest is systems, HW+SW, including HPC


COMMUNICATIONS OF THE ACM

CACM.ACM.ORG 09/2021 VOL.64 NO.09

The Future Is Big Graphs

Managing IT Professional Turnover
An Internet of Things Service Roadmap
Whose Smartphone Is It?
Q&A with ACM Computing Prize Winner David Silver

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

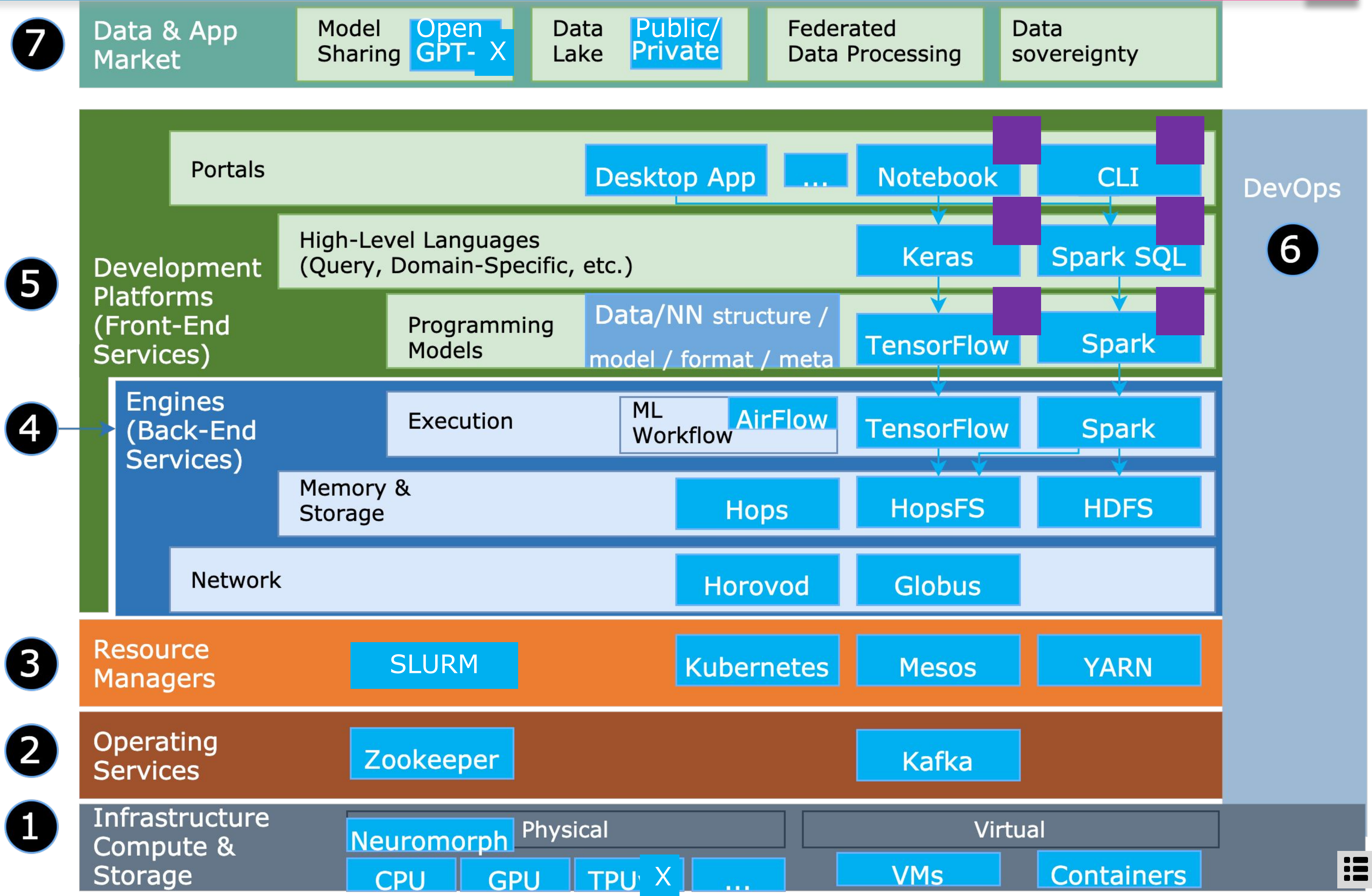
AUTOMATED AI/ML/DL OPERATIONS

^

ISSUES: COMPLEXITY, NON-TECHNICAL

Actual ML app is a very small part!

Rest is systems, HW+SW, including HPC



7 Data & App Market

Model Sharing — Open GPT- X

Data Lake — Public/ Private

Federated Data Processing

Data sovereignty

5 Development Platforms (Front-End Services)

Portals — Desktop App, …, Notebook, CLI

High-Level Languages (Query, Domain-Specific, etc.) — Keras, Spark SQL

Programming Models — Data/NN structure / model / format / meta — TensorFlow, Spark

4 Engines (Back-End Services)

Execution — ML Workflow: AirFlow — TensorFlow, Spark

Memory & Storage — Hops, HopsFS, HDFS

Network — Horovod, Globus

3 Resource Managers — SLURM, Kubernetes, Mesos, YARN

2 Operating Services — Zookeeper, Kafka

1 Infrastructure Compute & Storage

Physical — Neuromorph, CPU, GPU, TPU, X, …

Virtual — VMs, Containers

DevOps 6

Adapted from:

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

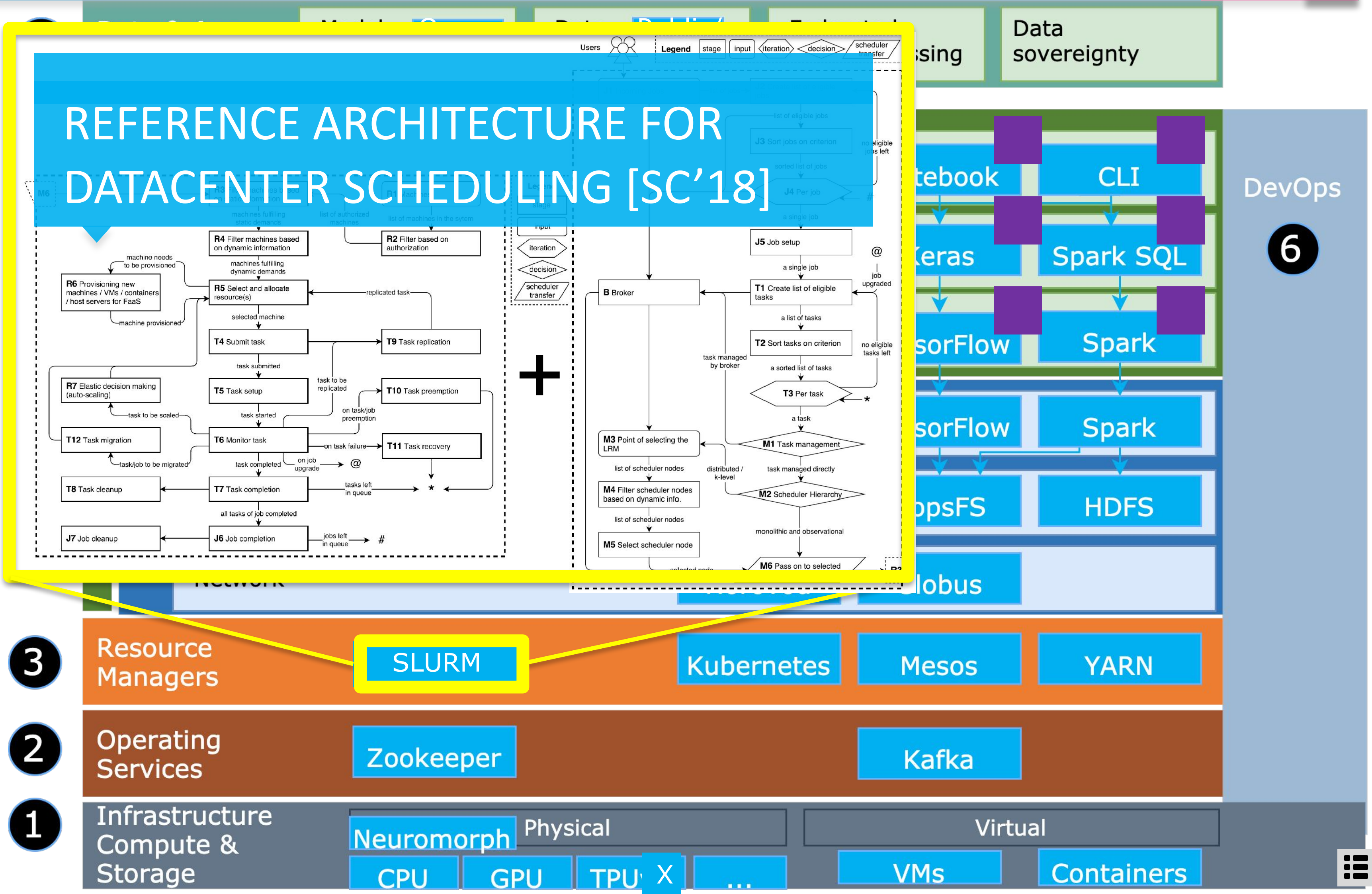
AUTOMATED AI/ML/DL OPERATIONS

^

ISSUES: COMPLEXITY, NON-TECHNICAL

Actual ML app is a very small part!

Rest is systems, HW+SW, including HPC


REFERENCE ARCHITECTURE FOR DATACENTER SCHEDULING [SC'18]

Adapted from:

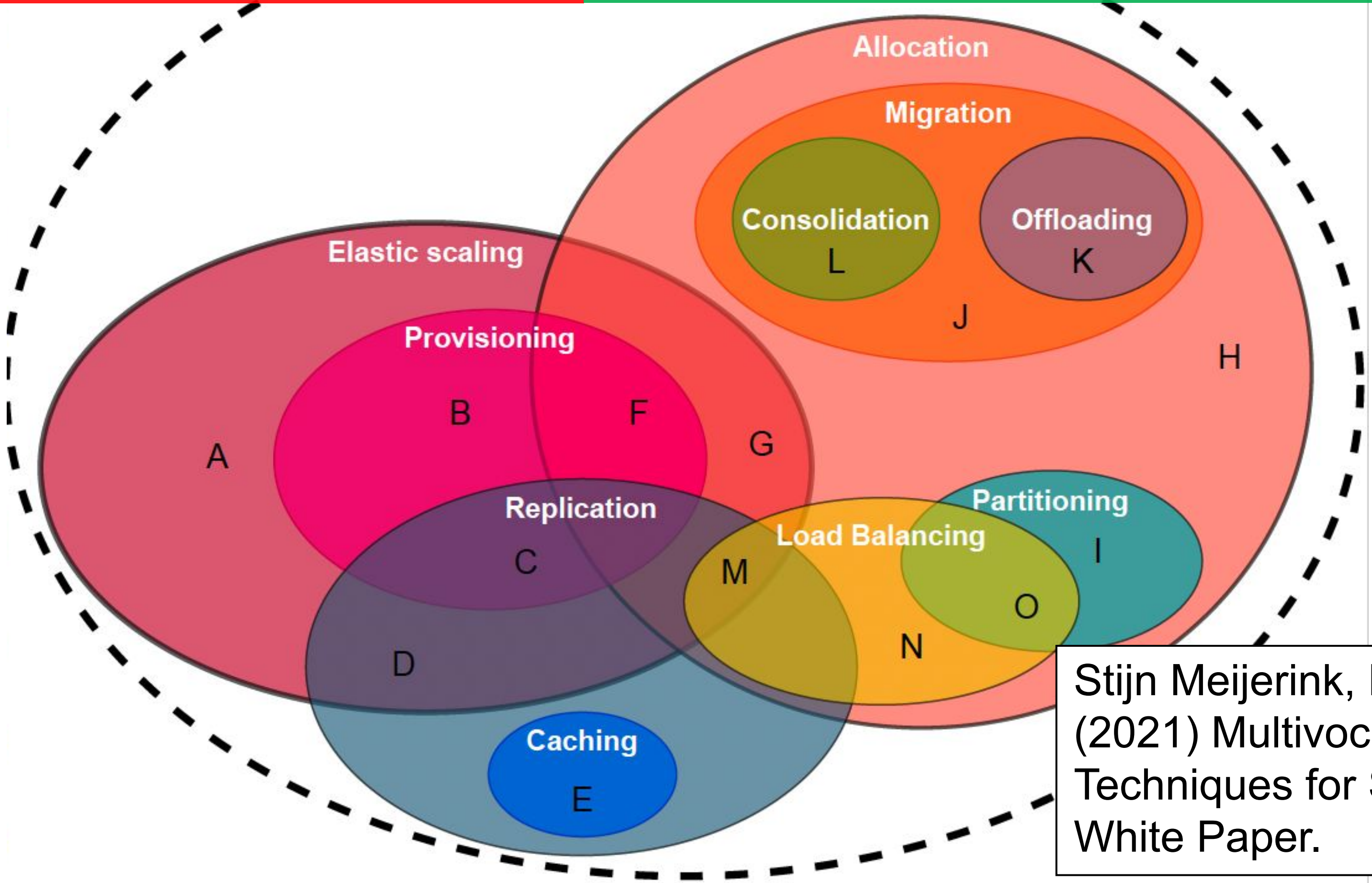
Sakr, Bonifati, Voigt, Iosup, et al. (2021) The Future Is Big Graphs! CACM

HOW TO DO RM&S ACROSS THE ECOSYSTEM?

^

IT'S OPERATIONS!

REFERENCE VIEW ON OPERATIONAL TECHNIQUES

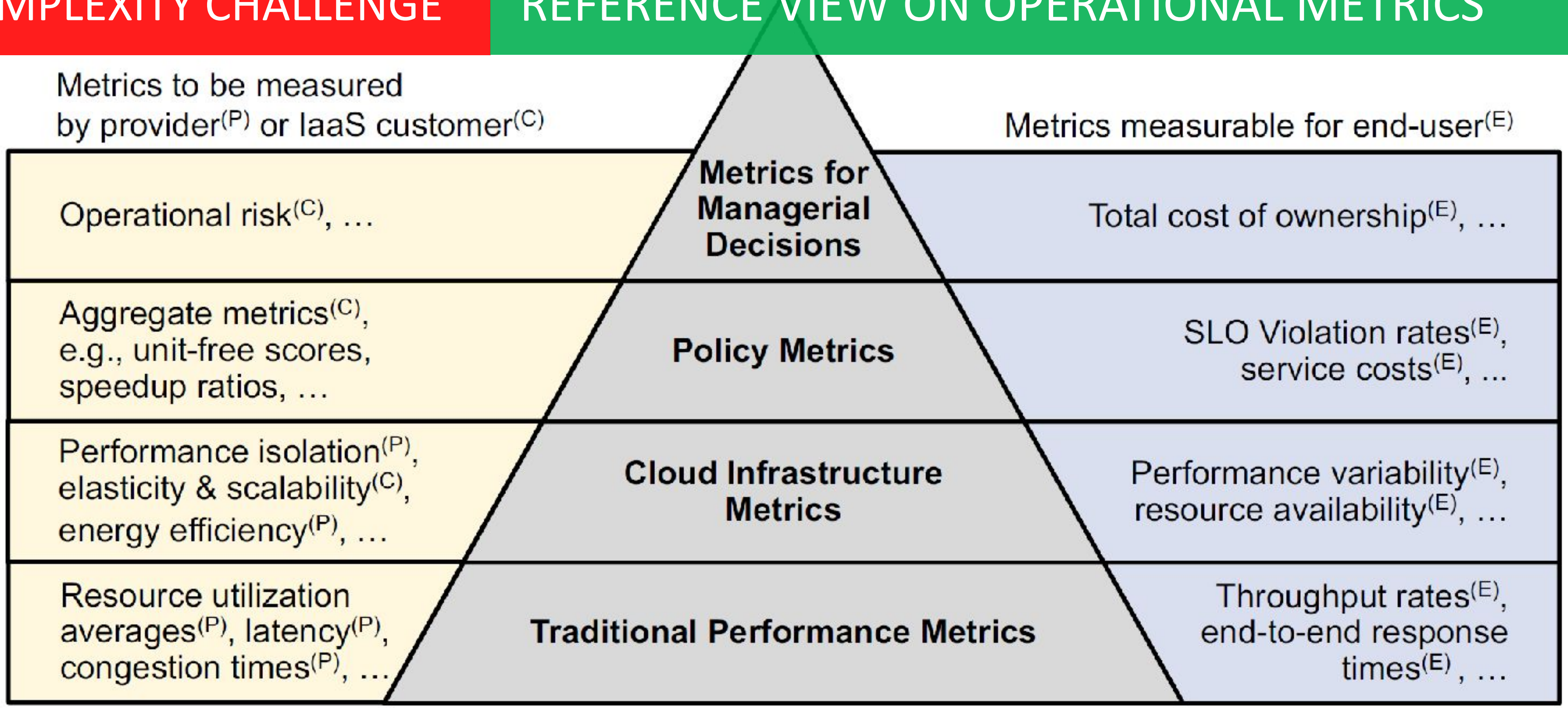


Stijn Meijerink, Erwin van Eyk, Alexandru Iosup (2021) Multivocal Survey of Operational Techniques for Serverless Computing. White Paper.

HOW TO REPORT PERFORMANCE?

THE COMPLEXITY CHALLENGE

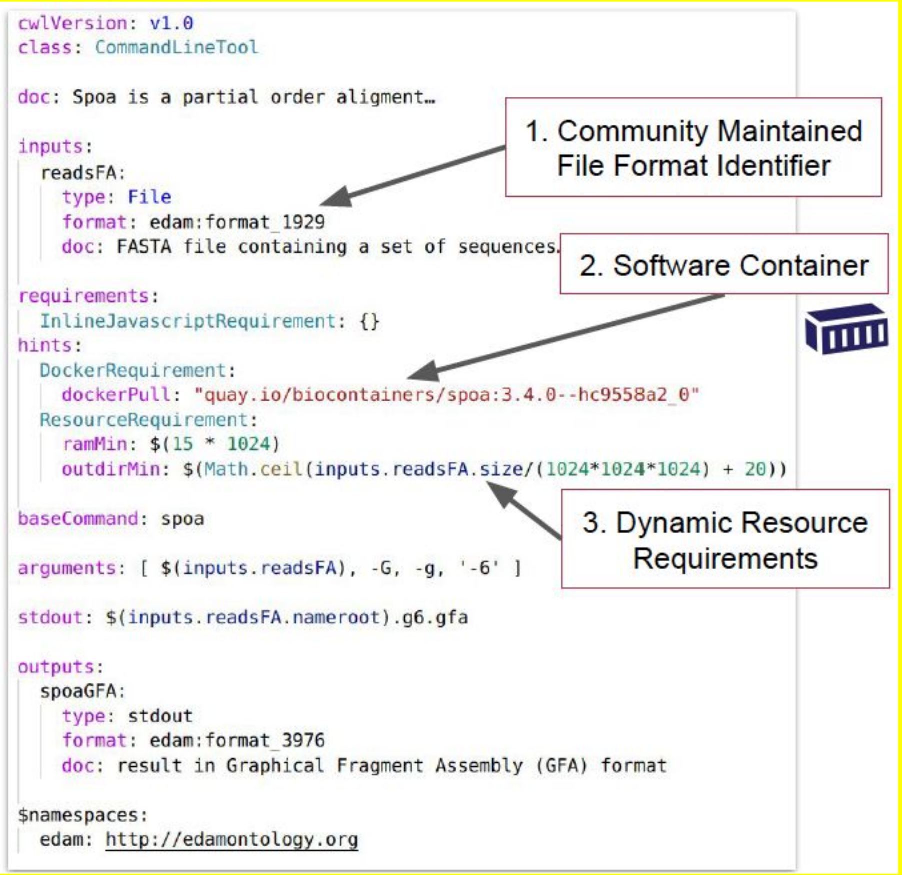
REFERENCE VIEW ON OPERATIONAL METRICS



Metrics to be measured by provider$^{(P)}$ or IaaS customer$^{(C)}$		Metrics measurable for end-user$^{(E)}$
Operational risk$^{(C)}$, …	**Metrics for Managerial Decisions**	Total cost of ownership$^{(E)}$, …
Aggregate metrics$^{(C)}$, e.g., unit-free scores, speedup ratios, …	**Policy Metrics**	SLO Violation rates$^{(E)}$, service costs$^{(E)}$, …
Performance isolation$^{(P)}$, elasticity & scalability$^{(C)}$, energy efficiency$^{(P)}$, …	**Cloud Infrastructure Metrics**	Performance variability$^{(E)}$, resource availability$^{(E)}$, …
Resource utilization averages$^{(P)}$, latency$^{(P)}$, congestion times$^{(P)}$, …	**Traditional Performance Metrics**	Throughput rates$^{(E)}$, end-to-end response times$^{(E)}$, …

N. Herbst, E. Van Eyk, C. L. Abad, A. Iosup, et al. (2018) Quantifying Cloud Performance and Dependability: Taxonomy, Metric Design, and Emerging Challenges. TOMPECS 3(4): 19:1-19:36

REPRODUCIBLE WORKFLOWS



```
cwlVersion: v1.0
class: CommandLineTool

doc: Spoa is a partial order aligment…

inputs:
  readsFA:
    type: File
    format: edam:format_1929
    doc: FASTA file containing a set of sequences.

requirements:
  InlineJavascriptRequirement: {}
hints:
  DockerRequirement:
    dockerPull: "quay.io/biocontainers/spoa:3.4.0--hc9558a2_0"
  ResourceRequirement:
    ramMin: $(15 * 1024)
    outdirMin: $(Math.ceil(inputs.readsFA.size/(1024*1024*1024) + 20))

baseCommand: spoa

arguments: [ $(inputs.readsFA), -G, -g, '-6' ]

stdout: $(inputs.readsFA.nameroot).g6.gfa

outputs:
  spoaGFA:
    type: stdout
    format: edam:format_3976
    doc: result in Graphical Fragment Assembly (GFA) format

$namespaces:
  edam: http://edamontology.org
```

1. Community Maintained File Format Identifier

2. Software Container

3. Dynamic Resource Requirements

p

Model Sharing | Open GPT- X

Data Lake | Public/ Private

Federated Data Processing

Data sovereignty

tals

Desktop App | ... | Notebook | CLI

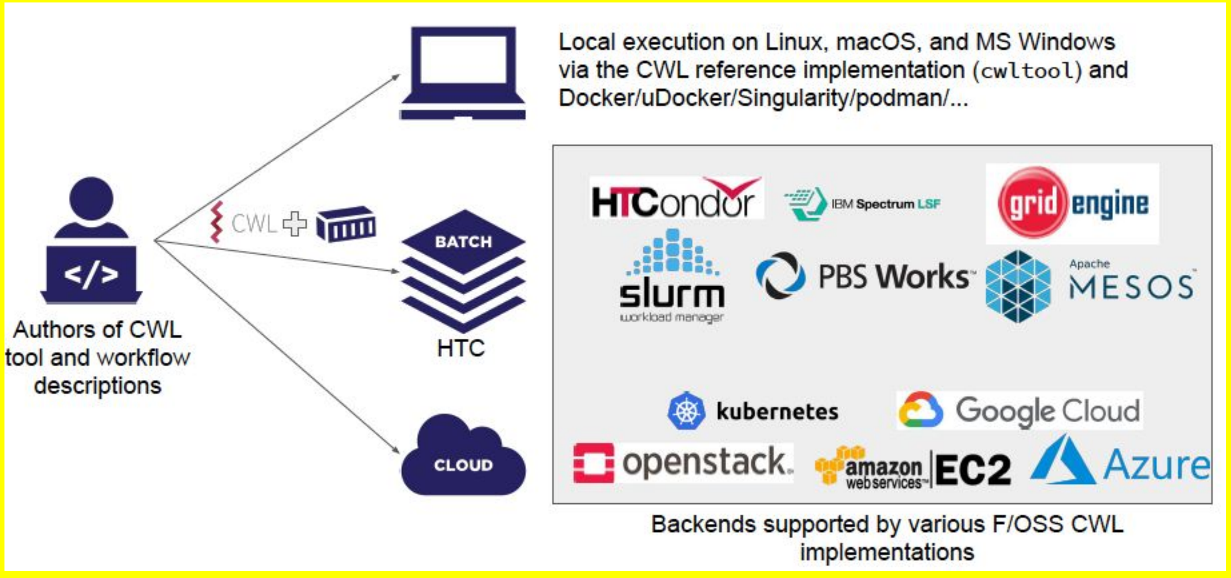
ent

High-Level Languages (Query, Domain-Specific, etc.)

Keras | Spark SQL

DevOps

6



Local execution on Linux, macOS, and MS Windows via the CWL reference implementation (cwltool) and Docker/uDocker/Singularity/podman/...

Authors of CWL tool and workflow descriptions

Backends supported by various F/OSS CWL implementations

Managers

2 Operating Services

1 Infrastructure Compute & Storage

Physical: Neuromorph | CPU | GPU | TPU X | ...

Virtual: VMs | Containers

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

Detailed, practical approaches

High-level approaches <<

Take-Home Message >>

6

(our practical approaches are often instruments and datasets)

First distributed & heterogeneous graph analytics system

(we often try unusual ideas, *be the first* philosophy)

Early Graph-Processing Systems: *Either* Distributed *or* Heterogeneous

Distributed CPU-based systems cannot use additional computational power of accelerators

Oracle Labs PGX

APACHE GIRAPH

GraphLab

GraphX

YARN

GPU-enabled systems are (mostly) single-machine systems, cannot handle large-scale graphs

NetSysLab

MEDUSA

medusa-gpu

Medusa: Simplified Graph Processing on GPUs

mapgraph Beta

Massively Parallel Graph processing on GPUs

TOTEM

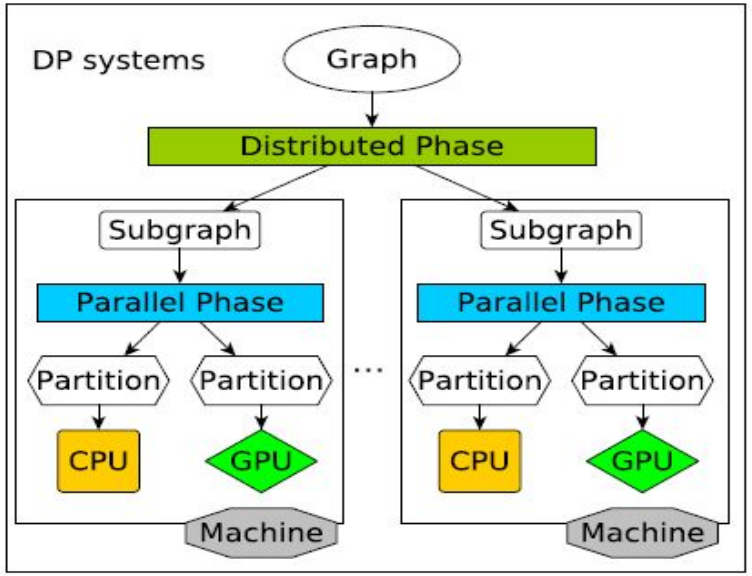
Gunrock

High-performance Graph Primitives on GPU

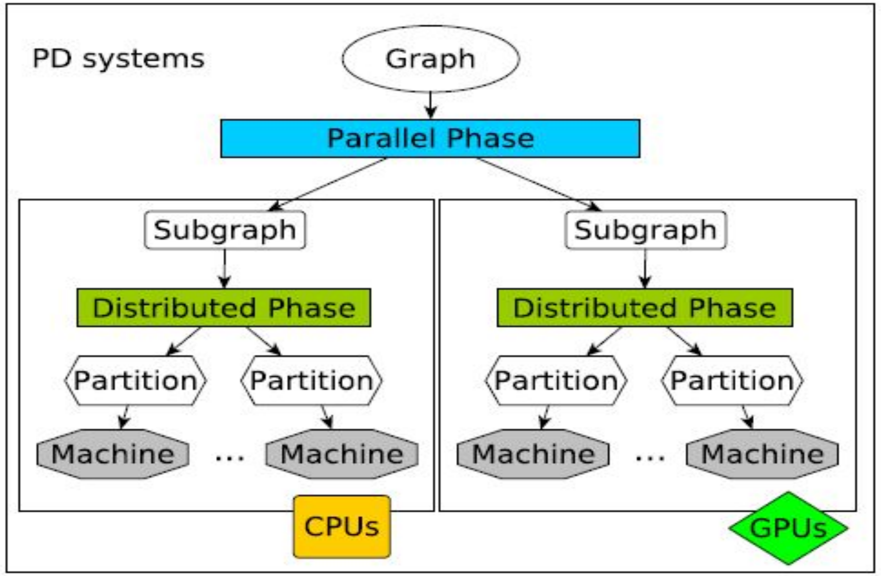
VertexAPI2

Y. Guo, A. L. Varbanescu, D. Epema, and A. Iosup, "Design and Experimental Evaluation of Distributed Heterogeneous Graph-Processing Systems," Submission to *CCGrid, 2016.*

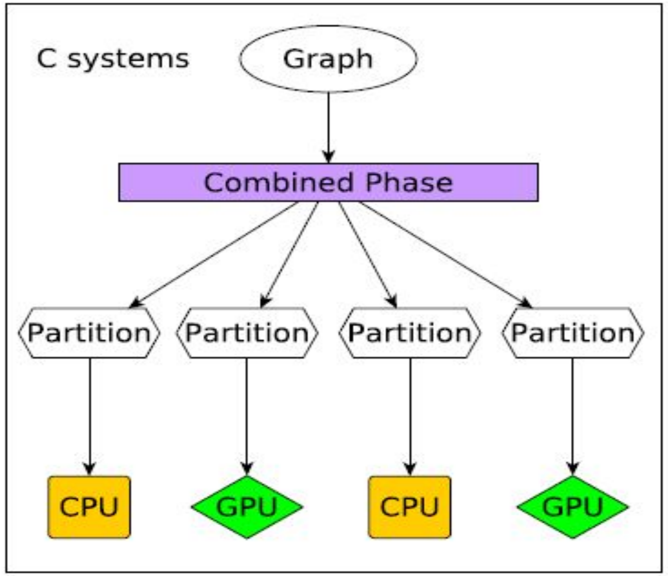
Our approach: 3 Families of Distributed *and* Heterogeneous (CPU+GPU) Graph-Processing Systems



Distributed-then-Parallel (DP) Systems



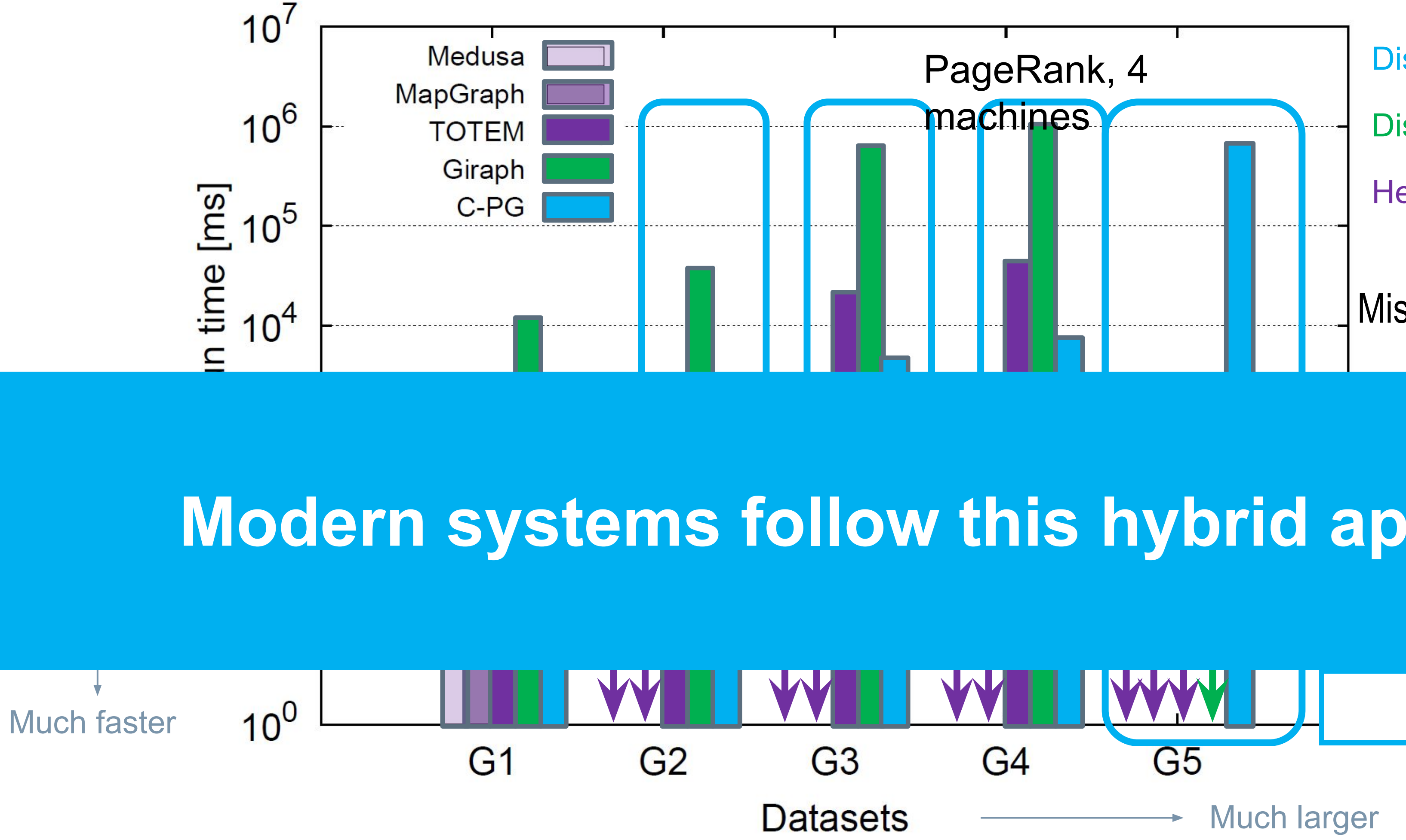
Parallel-then-Distributed (PD) Systems



(Combined Par.-and-Distributed (C) Systems

Y. Guo, A. L. Varbanescu, D. Epema, and A. Iosup, "Design and Experimental Evaluation of Distributed Heterogeneous Graph-Processing Systems," Submission to *CCGrid, 2016.*

Promising Results for Distributed *and* Heterogeneous Graph-Processing Systems



Distributed *and* Heterogeneous

Distributed only

Heterogeneous only

Missing bar = ➤ = system failure

Modern systems follow this hybrid approach

C-PG system scales

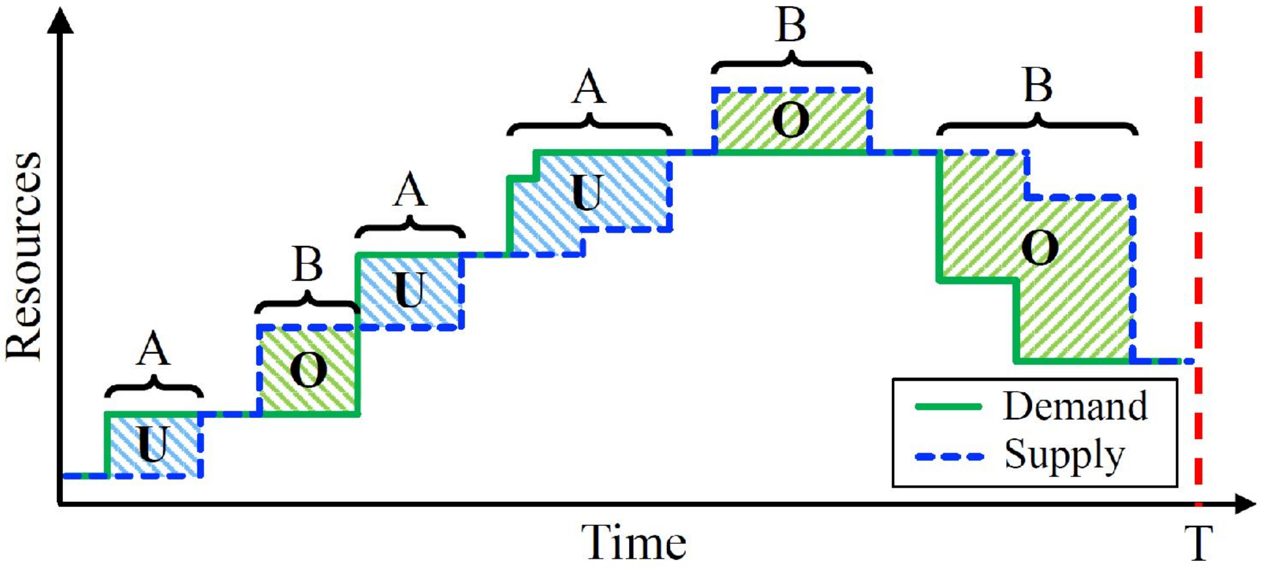
Y. Guo, A. L. Varbanescu, D. Epema, and A. Iosup, "Design and Experimental Evaluation of Distributed Heterogeneous Graph-Processing Systems," Submission to *CCGrid, 2016.*

First elastic graph analytics system

(we often try unusual ideas, *be the first* philosophy)

Modern Metrics: Elasticity

= Degree to which a system adapts to workload changes by provisioning and de-provisioning resources autonomically, s.t. the supply (the provisioned resources) matches the demand



Zone type A:
D > S ~ Underprovisioning

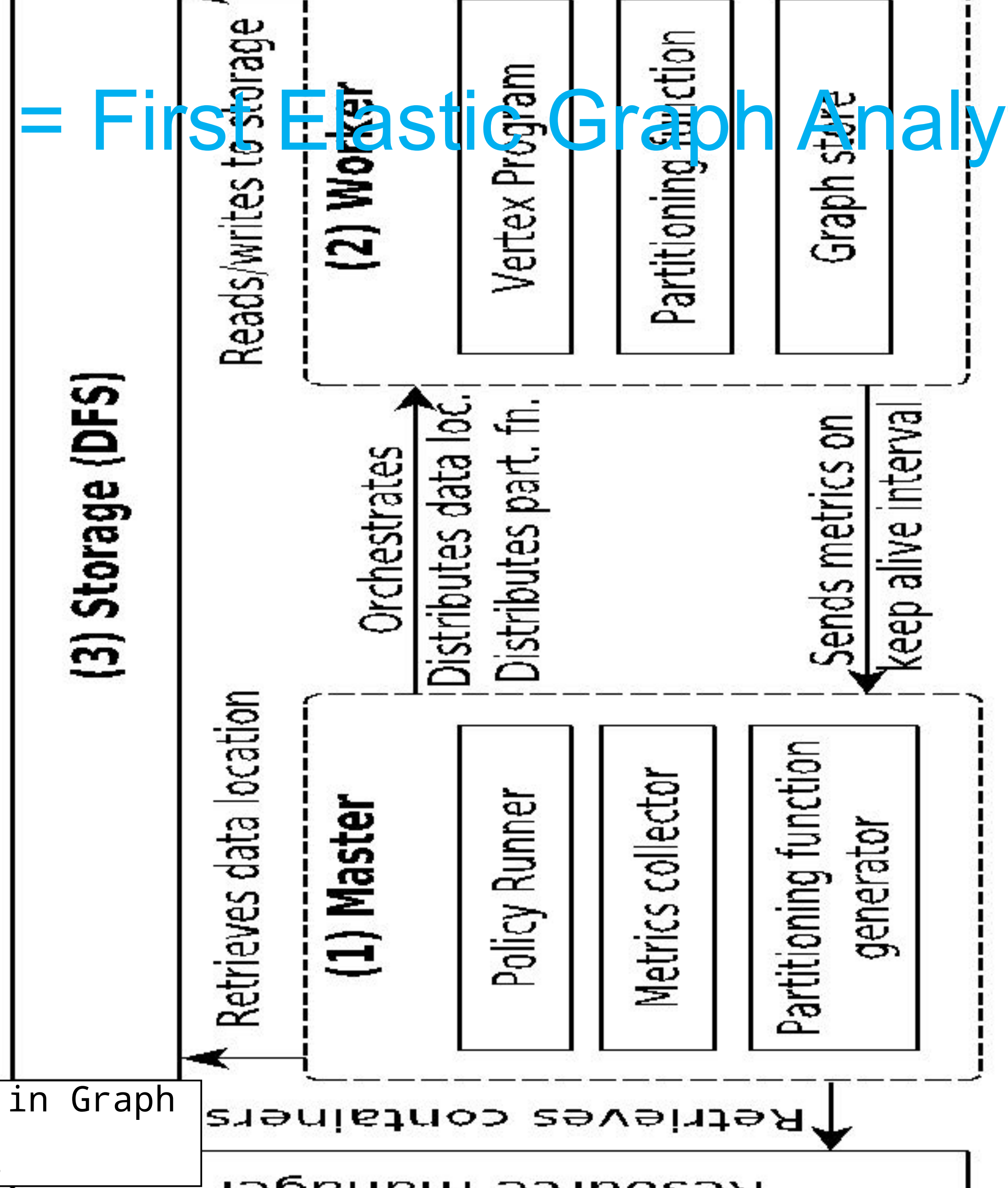
Zone type B:
D < S ~ Overprovisioning

Auto-scaling = system tries to provision exactly as many resources as the user workload demands

A. Ilyushkin, A. Ali-Eldin, N. Herbst, A. Bauer, A. V. Papadopoulos, D. H. J. Epema, A. Iosup (2018) An Experimental Performance Evaluation of Autoscalers for Complex Workflows. TOMPECS 3(2).

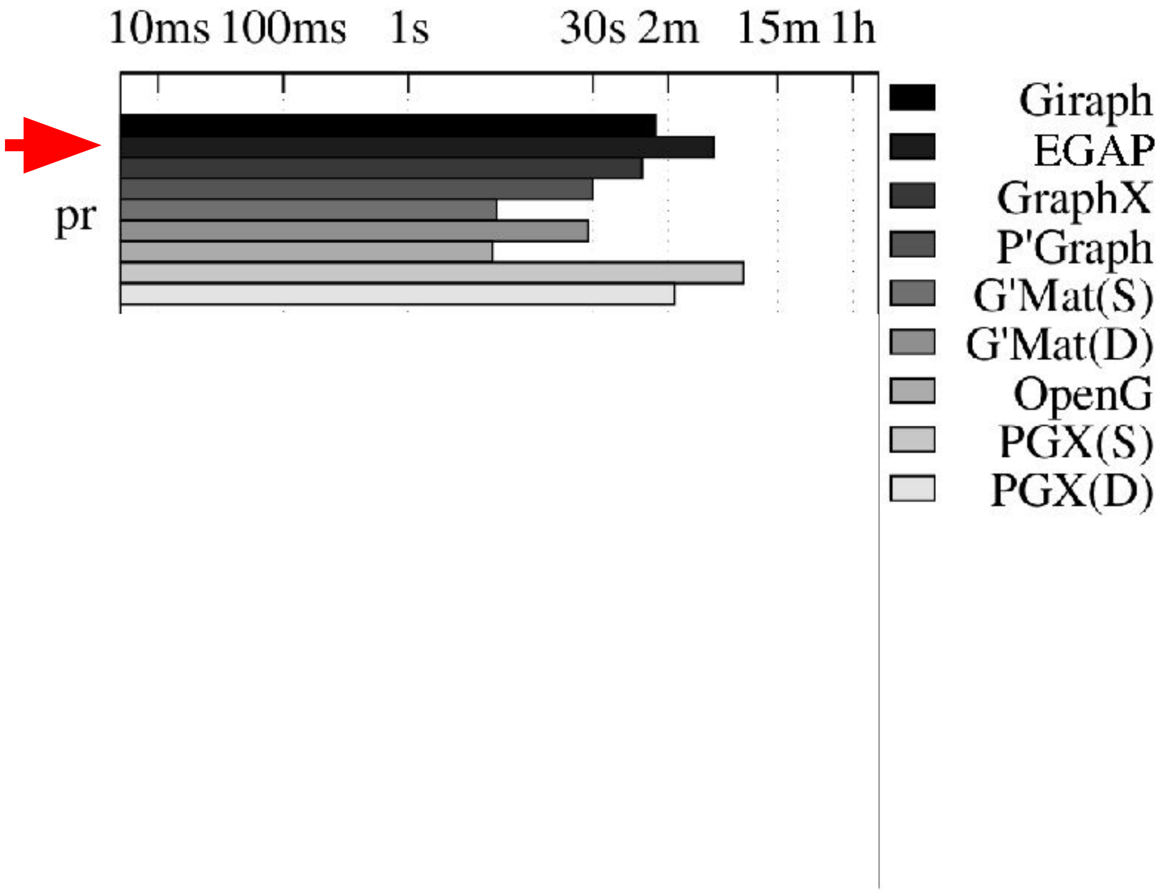
JoyGraph = First Elastic Graph Analytics Platform

Auto-scaling



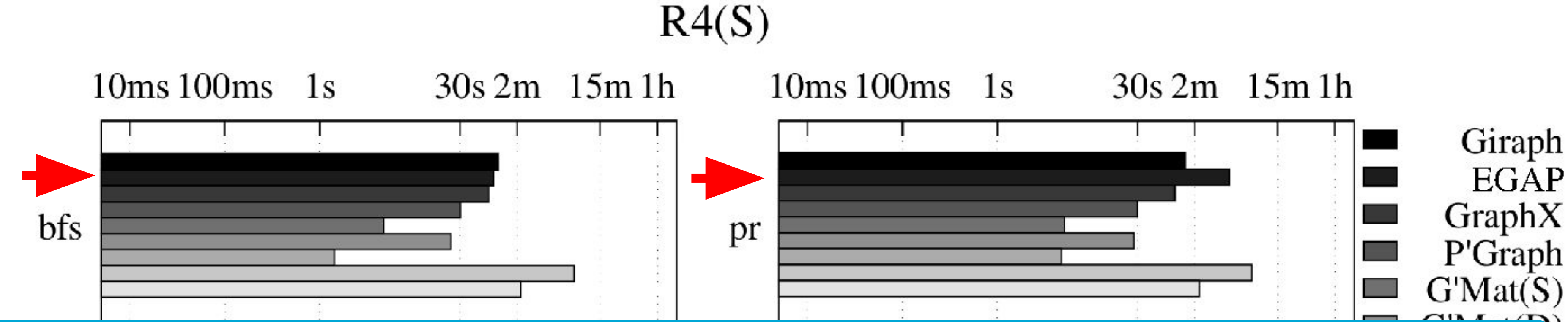
A. Uta et al., Elasticity in Graph Analytics? CLUSTER 2018

JoyGraph EGAP vs. State-of-Art



A. Uta et al., Elasticity in Graph Analytics? CLUSTER 2018

JoyGraph EGAP vs. State-of-Art



EGAP performs on-par with JVM-based platforms (Giraph, GraphX), but is much more efficient!

Modern systems increasingly follow this hybrid approach

A. Uta et al., Elasticity in Graph Analytics? CLUSTER 2018

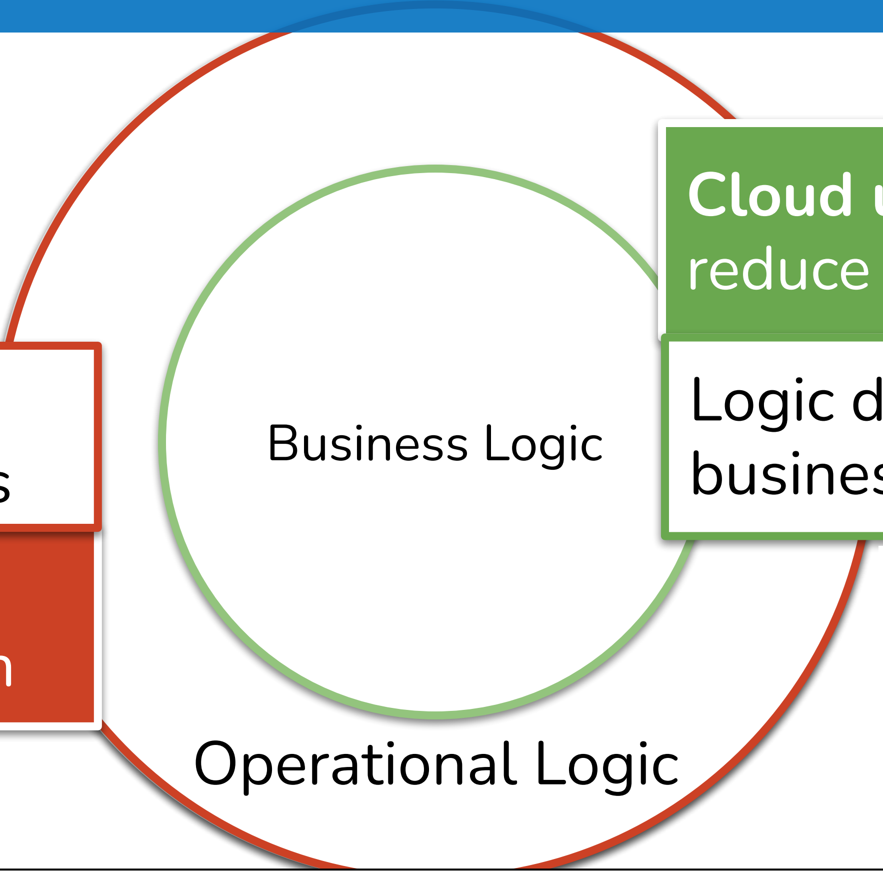
First serverless analytics system

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

(we often try unusual ideas,
be the first philosophy)

SERVERLESS: DECOMPOSING CLOUD APPLICATIONS

VERY EASY TO USE, VERY EFFICIENT OPERATION



Logic related to QoS of the application, system aspects

Cloud providers:
improve resource utilization

Cloud users:
reduce operational logic

Logic directly related to users, business aspects

Motivators

47% Save Costs

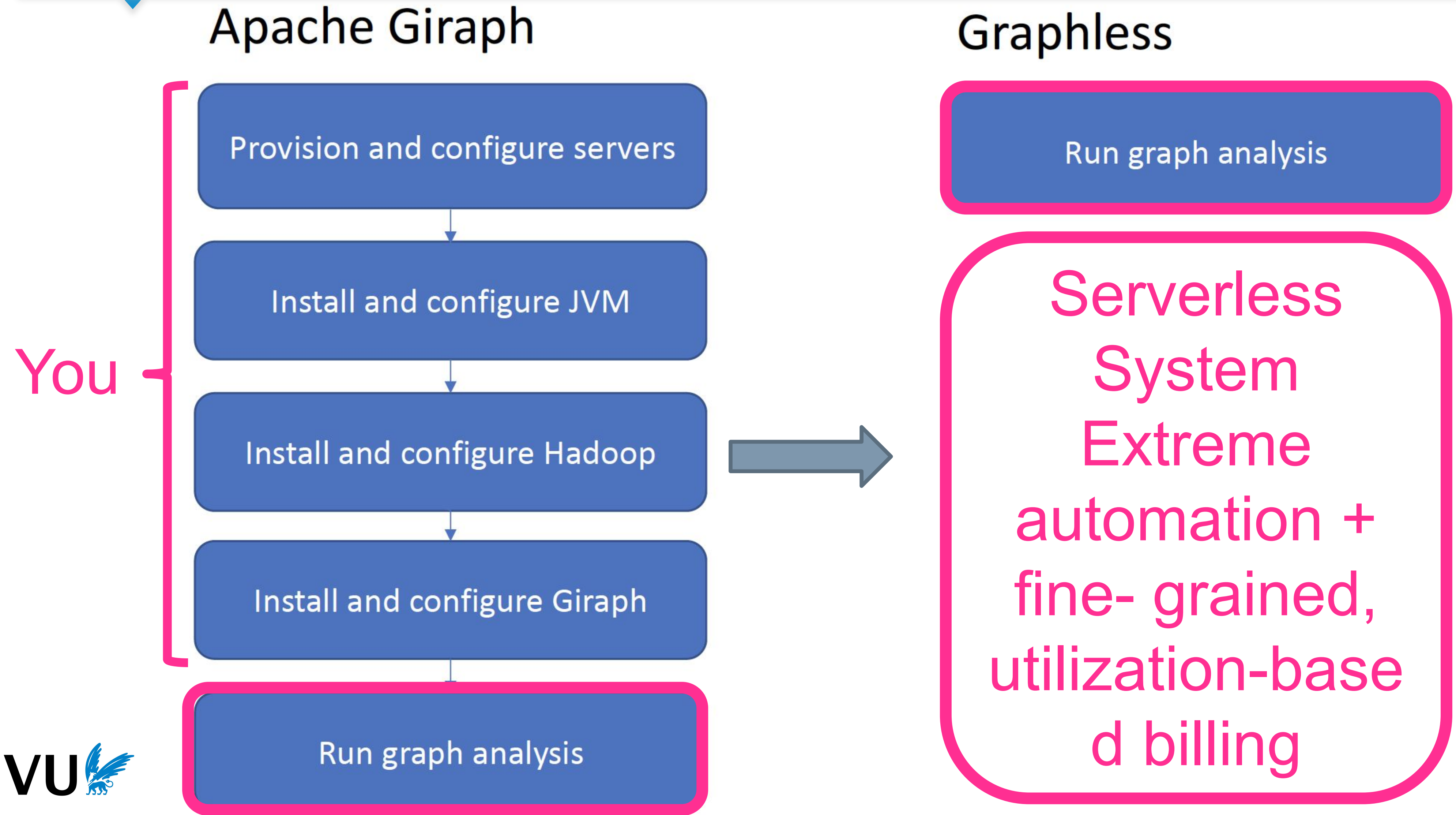
34% Built-in scalability

34% No operations

[Eismann et al. (2020) Serverless Applications: Why, When, and How? IEEE Software]

GRAPHLESS = SERVERLESS GRAPH PROCESSING

Lucian Toader



Apache Giraph

You

Provision and configure servers

Install and configure JVM

Install and configure Hadoop

Install and configure Giraph

Run graph analysis

Graphless

Run graph analysis

Serverless System Extreme automation + fine- grained, utilization-base d billing

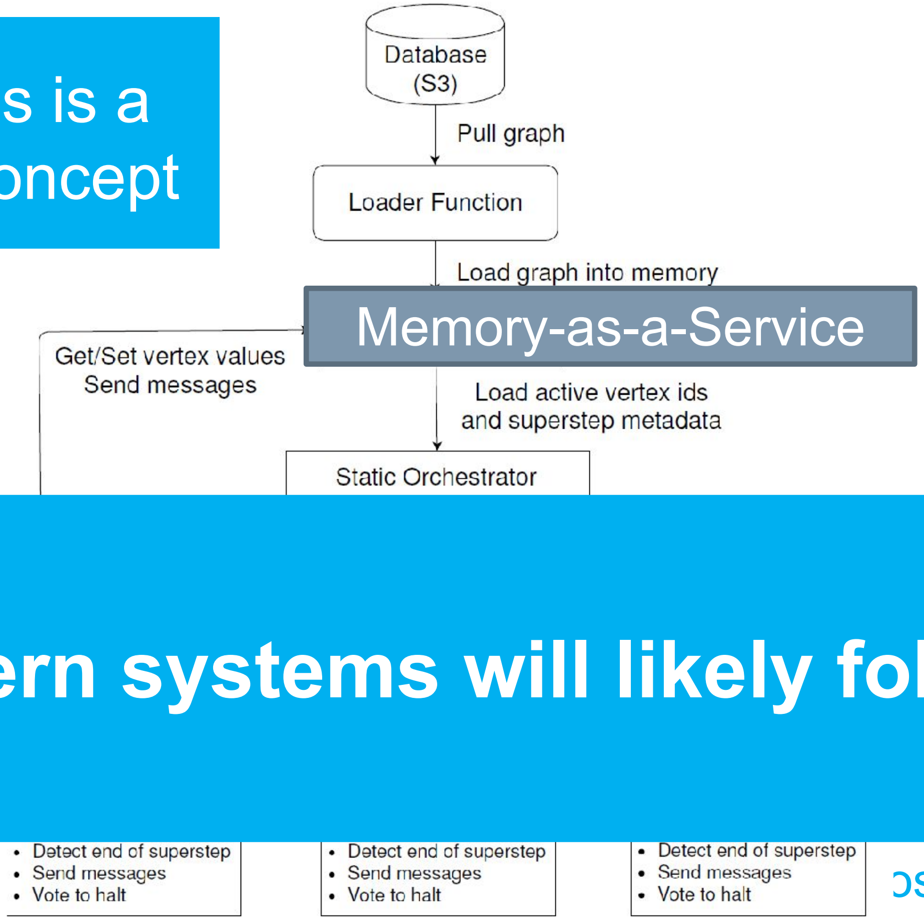
[Toader et al. '19]

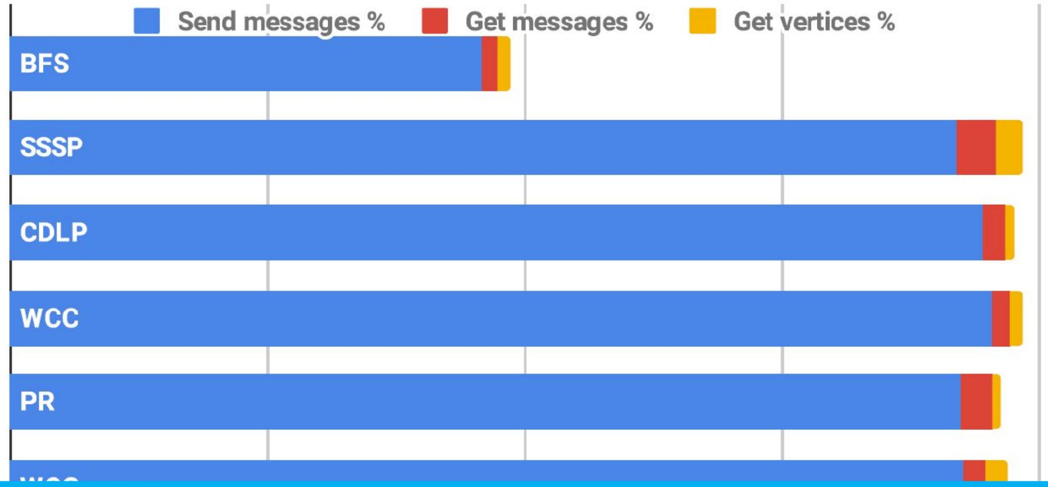
GRAPHLESS = SERVERLESS GRAPH PROCESSING

Lucian Toader

SERVERLESS ARCHITECTURE, API, SCHEDULER

Graphless is a proof-of-concept





Modern systems will likely follow this hybrid approach

ɔsup. All rights reserve

[Toader et al. '19]

GraphMassivizer

High-level approaches <<

Take-Home Message >>

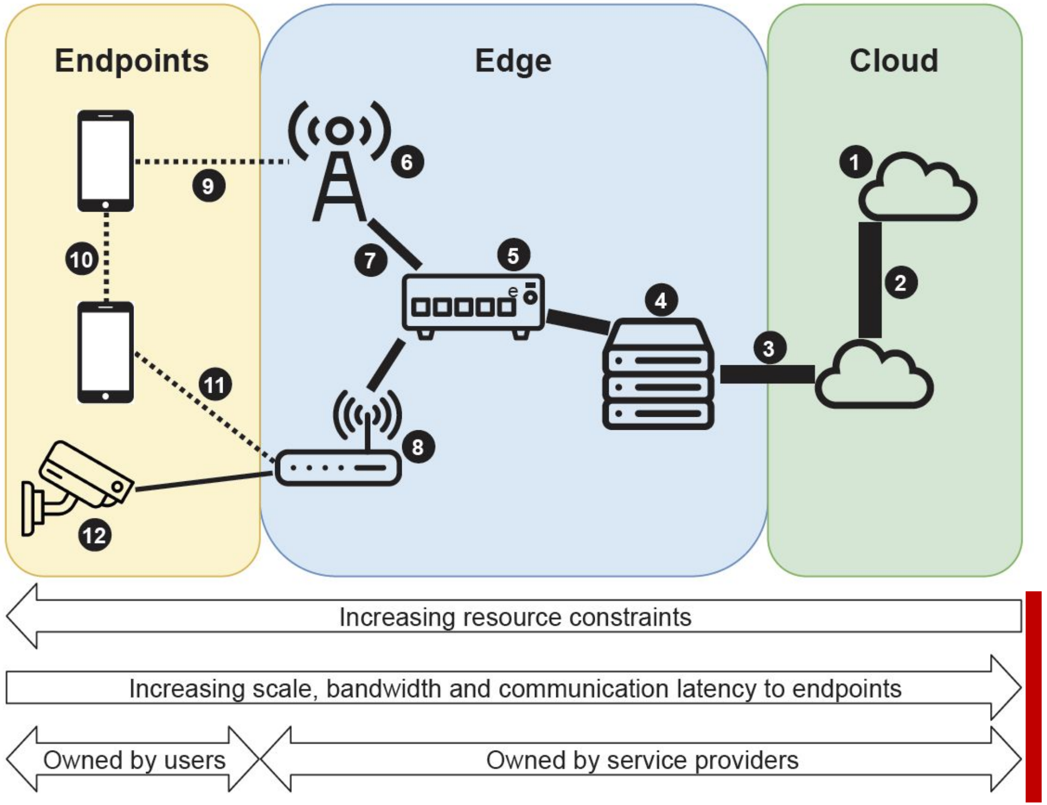
7

(all our knowledge and capability in one project)

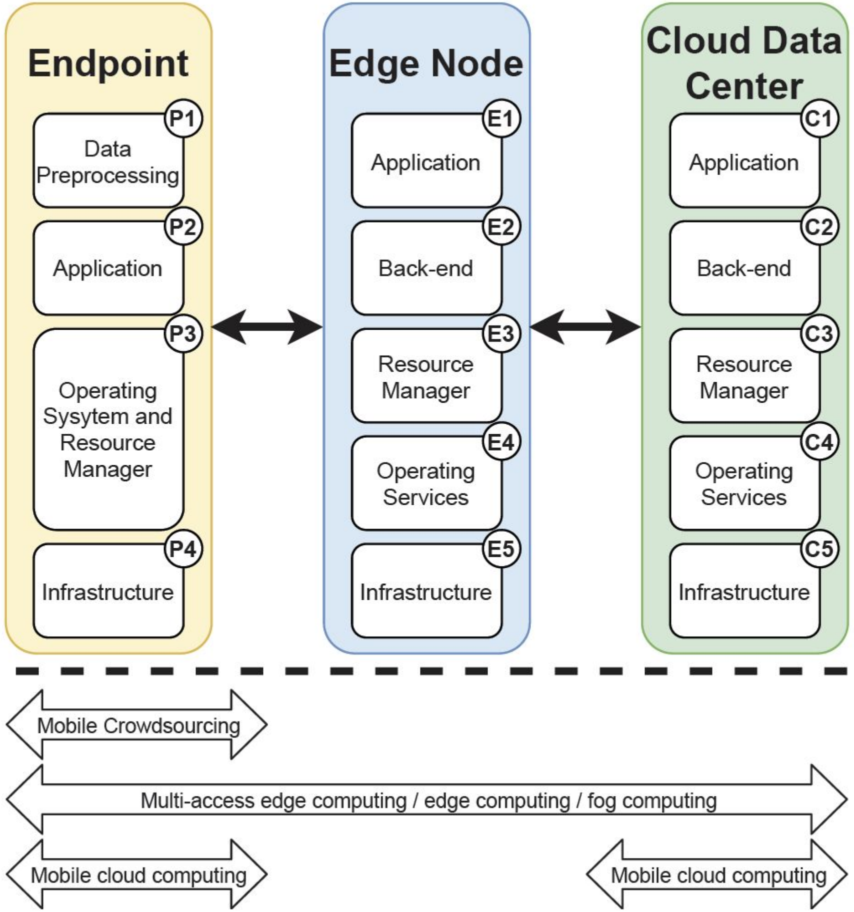
Computational Continuum DC → Endpoint-edge-cloud

Vision: Massivizing computer systems approaches enable holistic understanding and management in the computational continuum

APPROACH: WITH BROAD PARTICIPATION, UNIFY CURRENT COMPUTING MODELS, FORM A COMPUTATIONAL CONTINUUM



Endpoints
Edge
Cloud
1
2
3
4
5
6
7
8
9
10
11
12
Increasing resource constraints
Increasing scale, bandwidth and communication latency to endpoints
Owned by users
Owned by service providers



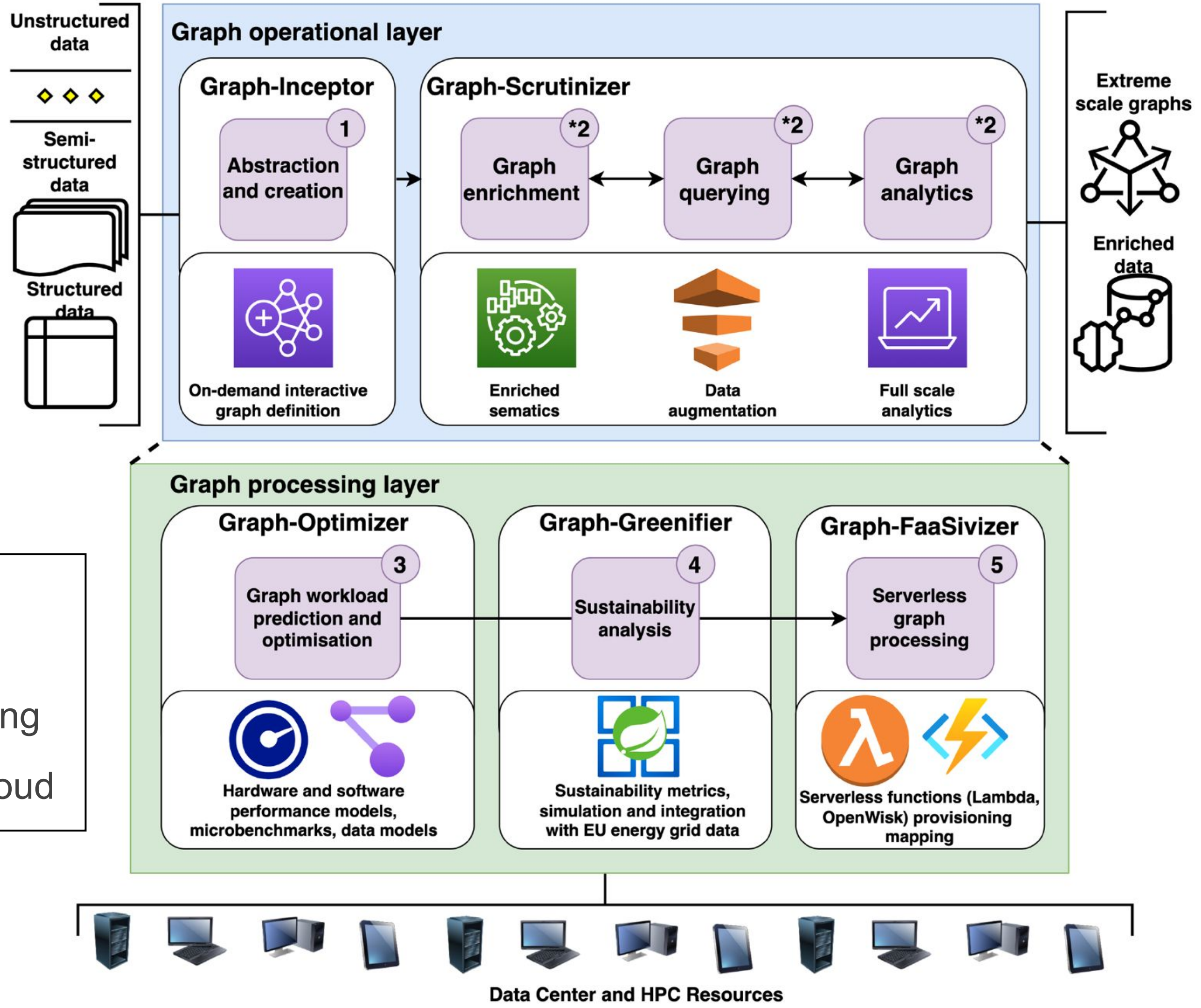
Trivedi, Wang, Bal, Iosup (2021) Sharing and Caring of Data at the Edge. HotEdge.

Jansen et al. (2022) The SPEC-RG Reference Architecture for the Edge Continuum. CoRR abs/2207.04159

ONE PROJECT TO MENTION…

GraphMassiziver

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



Unstructured data

Semi-structured data

Structured data

Graph operational layer

Graph-Inceptor

1 Abstraction and creation

On-demand interactive graph definition

Graph-Scrutinizer

*2 Graph enrichment

*2 Graph querying

*2 Graph analytics

Enriched sematics

Data augmentation

Full scale analytics

Extreme scale graphs

Enriched data

Graph processing layer

Graph-Optimizer

3 Graph workload prediction and optimisation

Hardware and software performance models, microbenchmarks, data models

Graph-Greenifier

4 Sustainability analysis

Sustainability metrics, simulation and integration with EU energy grid data

Graph-FaaSivizer

5 Serverless graph processing

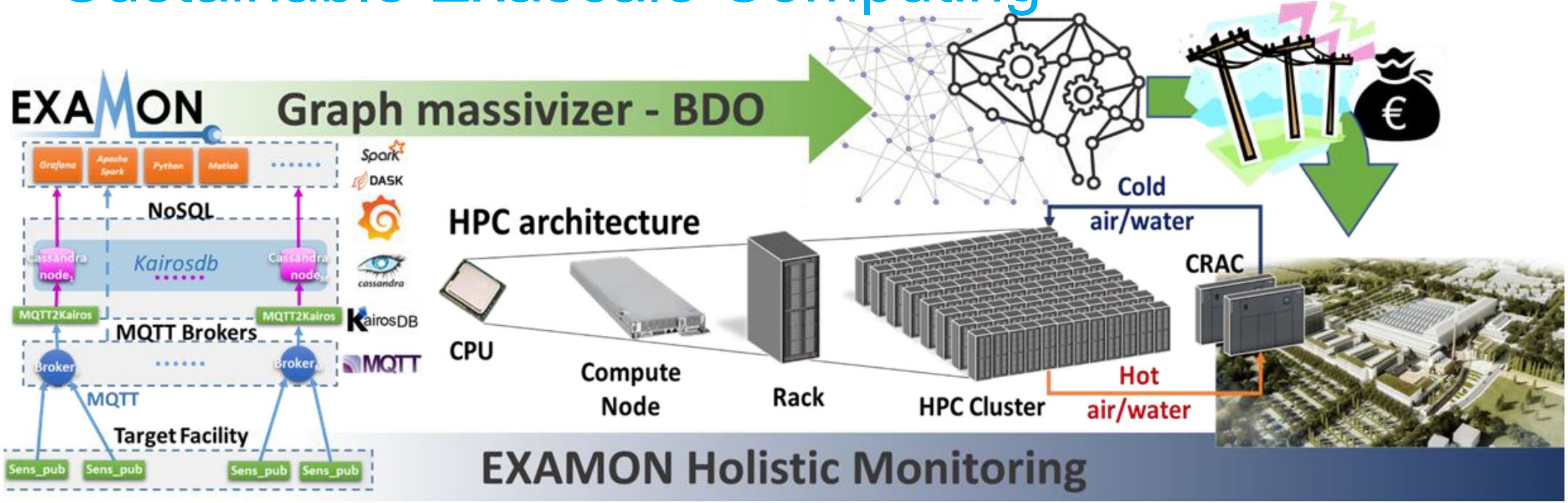
Serverless functions (Lambda, OpenWisk) provisioning mapping

Data Center and HPC Resources

Prodan et al (2022) The GraphMassivizer project: Towards Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe. IEEE Cloud Summit.

ONE APPLICATION TO MENTION…

Data Centre Digital Twin for Sustainable Exascale Computing



Goal: Design massive datacenter graph-baseds models capturing spatiotemporal dependencies between computation, nodes, and cooling equipment. Conduct analytics to predict, e.g., the impact of the spatial power distribution on cooling efficiency and cost.

TAKE-HOME MESSAGE

Massivizing □ computer ecosystems with good functional and non-functional properties, for all

The graph processing ecosystem is vast, challenging: many apps, many platforms, may goals. We lead many (first) approaches related to performance, RM&S.

Many modern, open challenges: resource management and scheduling, telemetry, analysis, simulation, experimentation, etc. GraphMassivizer to help.

About our team <<

Golden Age <<

…but <<

This talk <<

A new science <<

High-level approaches <<

Detailed approaches >>

MASSIVIZING COMPUTER SYSTEMS

FURTHER READING

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

1. Iosup et al. Massivizing Computer Systems. ICDCS 2018 ← start here
2. Andreadis et al. A Reference Architecture for Datacenter Scheduling, SC18
3. Van Eyk et al. Serverless is More: From PaaS to Present Cloud Computing, IEEE IC Sep/Oct 2018
4. Uta et al. Exploring HPC and Big Data Convergence: A Graph Processing Study on Intel Knights Landing, IEEE Cluster 2018
5. Talluri et al. Big Data Storage Workload in the Cloud. ACM/SPEC ICPE 2019.
6. Toader et al. Graphless. IEEE ISPDC’19.
7. Jiang et al. Mirror. CCPE 2018.
8. Ilyushkin et al. Autoscalers. TOMPECS 2018.
9. Versluis et al. Autoscaling Workflows. CCGRID’18.
10. Uta et al. Elasticity in Graph Analytics? IEEE Cluster 2018.
11. Herbst et al. Ready for rain? TOMPECS 2018.
12. Guo et al. Streaming Graph-partitioning. JPDC’18.
13. Iosup et al. The OpenDC Vision. ISPDC’17.
14. Iosup et al. Self-Aware Computing Systems book.
15. Iosup et al. LDBC Graphalytics. PVLDB 2016.

Etc.

MASSIVIZING COMPUTER SYSTEMS

FURTHER READING

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

1. Crusoe, Iosup, et al. (2022) Methods Included: CWL. CACM
2. Sakr, Bonifati, Voigt, Iosup, et al. (2021) The Future Is Big Graphs! CACM
3. Andreadis et al. (2021) Capelin: Data-Driven Capacity Procurement for Cloud Datacenters using Portfolios of Scenarios. TPDS, under review.
4. Versluis et al. The Workflow Trace Archive: Open-Access Data From Public and Private Computing Infrastructures. TPDS 2020.

5. Uta et al. (2020) Beneath the SURFace: An MRI-like View into the Life of a 21st-Century Datacenter. login USENIX
6. Iosup, Hegeman, et al. (2020) The LDBC Graphalytics Benchmark. CoRR. https://arxiv.org/abs/2011.15028
7. Hegeman et al. (2021) GradeML. HotCloudPerf.

8. Abad, Iosup, et al. An Analysis of Distributed Systems Syllabi With a Focus on Performance-Related Topics. WEPPE 2021. https://arxiv.org/abs/2103.01858

Etc.

MASSIVIZING COMPUTER SYSTEMS

FURTHER READING

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

1. Iosup et al. The AtLarge Vision on the Design of Distributed Systems and Ecosystems. ICDCS 2019 ← Start here
2. Uta et al. Is big data performance reproducible in modern cloud networks? NSDI 2020
3. Van Eyk et al. The SPEC-RG Reference Architecture for FaaS: From Microservices and Containers to Serverless Platforms, IEEE IC 2019
4. Papadopoulos et al. Methodological Principles for Reproducible Performance Evaluation in Cloud Computing. TSE 2019 and (journal-first) ICSE 2020
5. van Beek et al. Portfolio Scheduling for Managing Operational and Disaster-Recovery Risks in Virtualized Datacenters Hosting Business-Critical Workloads. ISPDC 2019
6. van Beek et al. A CPU Contention Predictor for Business-Critical Workloads in Cloud Datacenters. HotCloudPerf19

+ Iyushkin et al. Performance-Feedback Autoscaling with Budget Constraints for Cloud-based Workloads of Workflows. Under submission

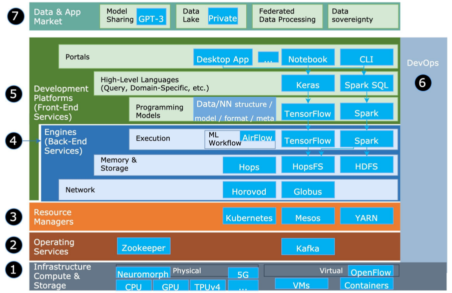
Etc.

EXTRAS

Massivizing □□ Grand Challenges

MASSIVIZING COMPUTER SYSTEMS

A LARGER VISION OF HOW COMPUTING WILL HELP OUR SOCIETY


7 Data & App Market: Model Sharing (GPT-3), Data Lake (Private), Federated Data Processing, Data sovereignty

5 Development Platforms (Front-End Services)
- Portals: Desktop App, ..., Notebook, CLI
- High-Level Languages (Query, Domain-Specific, etc.): Keras, Spark SQL
- Programming Models: Data/NN structure / model / format / meta, TensorFlow, Spark

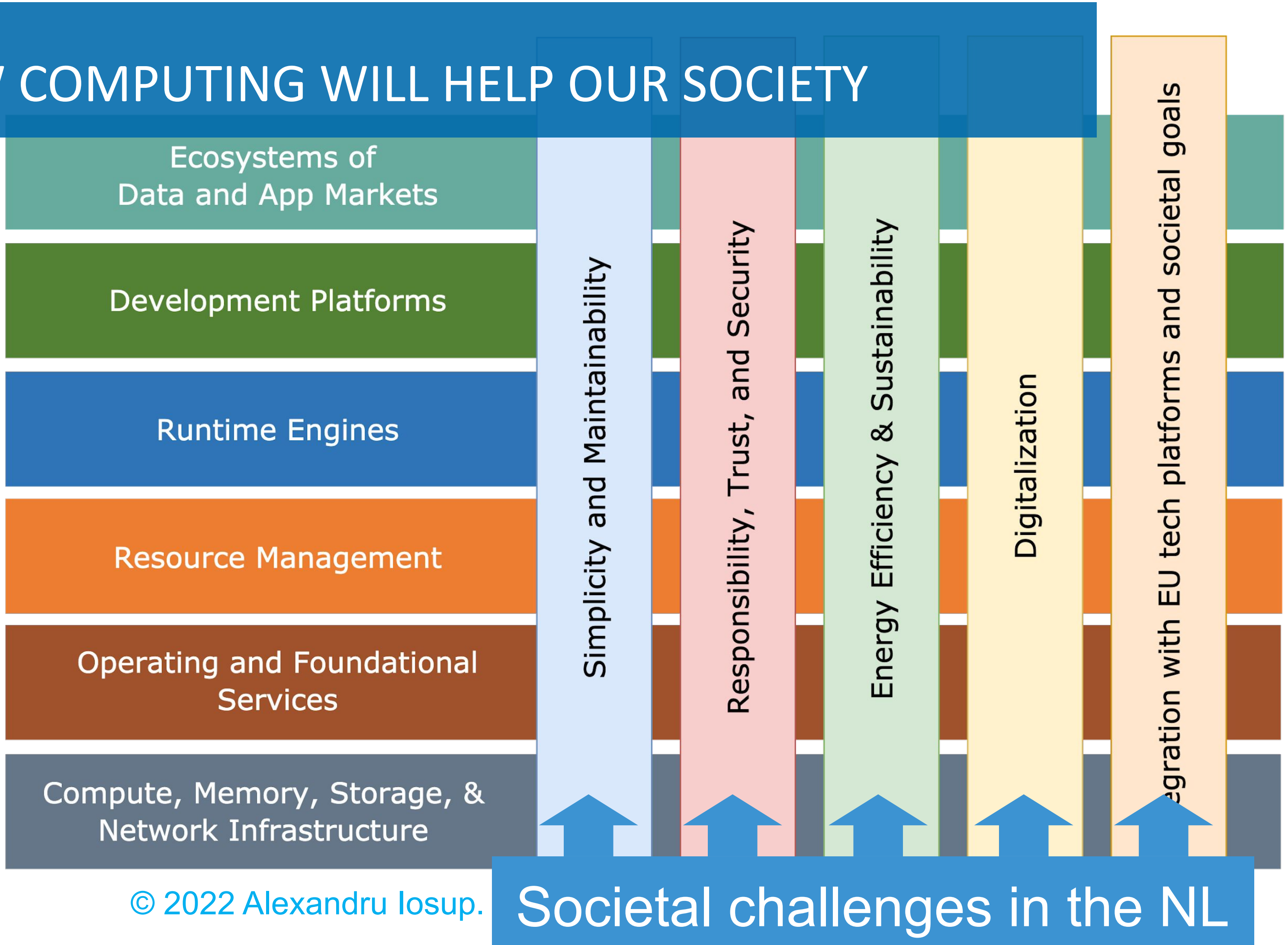
4 Engines (Back-End Services)
- Execution: ML Workflow (AirFlow), TensorFlow, Spark
- Memory & Storage: Hops, HopsFS, HDFS
- Network: Horovod, Globus

3 Resource Managers: Kubernetes, Mesos, YARN

2 Operating Services: Zookeeper, Kafka

1 Infrastructure Compute & Storage: Physical (Neuromorph, 5G, CPU, GPU, TPUv4, ...); Virtual (OpenFlow, VMs, Containers)

6 DevOps


Ecosystems of Data and App Markets

Development Platforms

Runtime Engines

Resource Management

Operating and Foundational Services

Compute, Memory, Storage, & Network Infrastructure

Simplicity and Maintainability

Responsibility, Trust, and Security

Energy Efficiency & Sustainability

Digitalization

egration with EU tech platforms and societal goals

Societal challenges in the NL

A.Iosup@vu.nl
http://atlarge.science

© 2022 Alexandru Iosup.