MASSIVIZING COMPUTER SYSTEMS

Co-sponsored by:

NWO Netherlands Organisation for Scientific Research

STW connecting innovators

= MAKING COMPUTER SYSTEMS SCALABLE, RELIABLE, PERFORMANT, ETC., YET ABLE TO FORM AN EFFICIENT ECOSYSTEM

@Large Research
Massivizing Computer Systems

http://atlarge.science

VU VRIJE UNIVERSITEIT AMSTERDAM

@Alosup

Prof. dr. ir. Alexandru Iosup

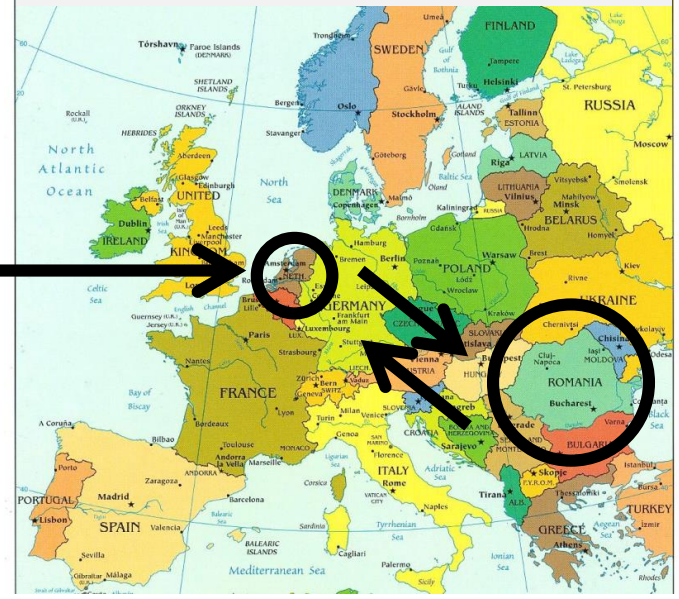
VU AMSTERDAM < SCHIPHOL < THE NETHERLANDS < EUROPE > ROMANIA > OTOPENI > BUCUREȘTI > POLITEHNICA

Amsterdam
founded 10th century
pop: 850,000

VU
founded 1880
pop: 23,500

The Netherlands

Europe



VU UNIVERSITY AMSTERDAM

POLITEHNICA DIN BUCUREȘTI
UNIVERSITATEA
UPB
1818

NOI RADACINI .nl — DRUMUL SPRE SUCCES AL ROMÂNILOR DIN OLANDA

DE CLAUDIA MARCU ȘI ALEXANDRU IOSUP, PLUS O ECHIPĂ DE VOLUNTARI

Mihai Netea

http://bit.ly/NR-Mihai

Ana Maria Oprescu

http://bit.ly/NR-AnaMaria

Teodor Cătăniciu

http://bit.ly/NR-Teodor

VU

MASSIVIZING COMPUTER SYSTEMS: OUR MISSION

1. Improve the lives of millions through impactful research.

2. Educate the new generation of top-quality, socially responsible professionals.

3. Make innovation available to society and industry.

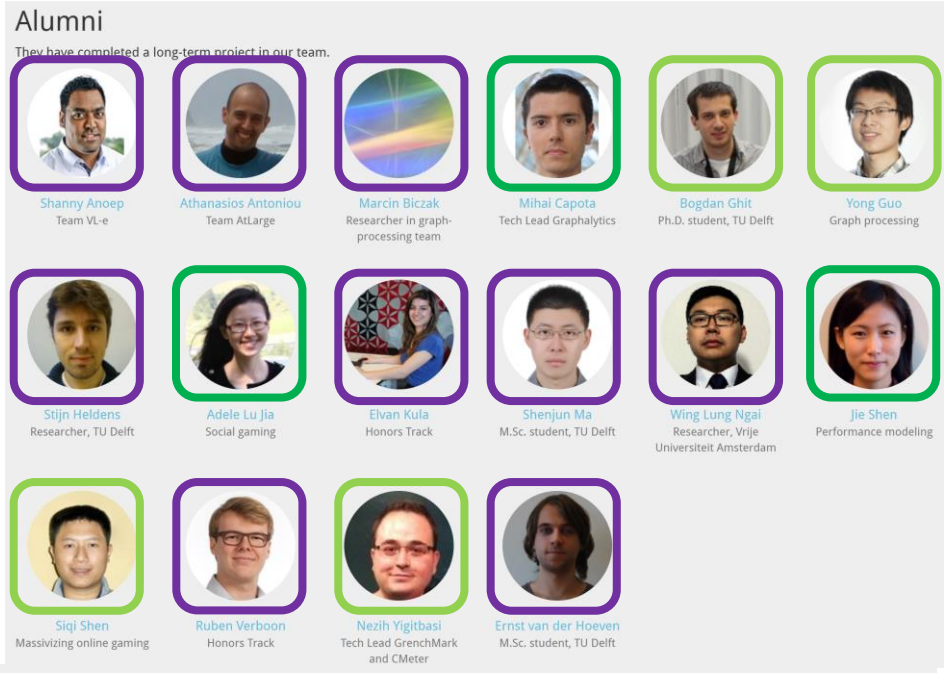
VU VRIJE UNIVERSITEIT AMSTERDAM

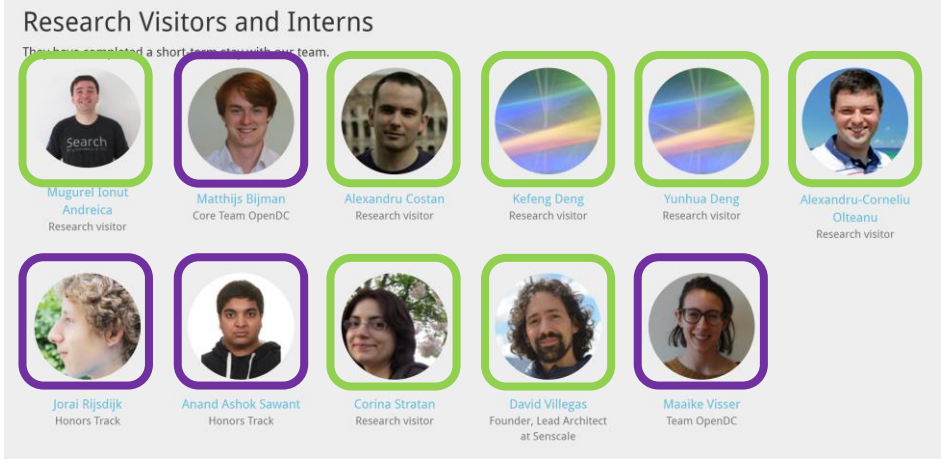
http://atlarge.science/about.html

TU Delft

ATLARGE RESEARCH: OUR TEAM

- Professor
- Assistant Prof.
- Teacher
- Post-doc
- Ph.D. student
- Scientist





http://atlarge.science/people.html

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

- Education:
  - Systems Architecture (BSc)
  - Distributed Systems (MSc)
- Research:
  - Massivizing Computer Systems

VU

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

- Education:
  - > Systems Architecture (BSc)
  - > Distributed Systems (MSc)
- Research:
  - > Massivizing Computer Systems

- About me:
  - > Worked in 7 countries, NL since 2004
  - > I like to help… I train people in need
  - > VU University Research Chair
  - > NL ICT Researcher of the Year
  - > NL Higher-Education Teacher of the Year
  - > NL KNAW Royal Young Academy

VU

~40’

Massivizing Computer Systems
A Structured Discussion

~5’ — About the Massivizing Computer Systems Group & Our Team

~10’ — The Golden Age of Massive Systems …Yet We Are in a Crisis

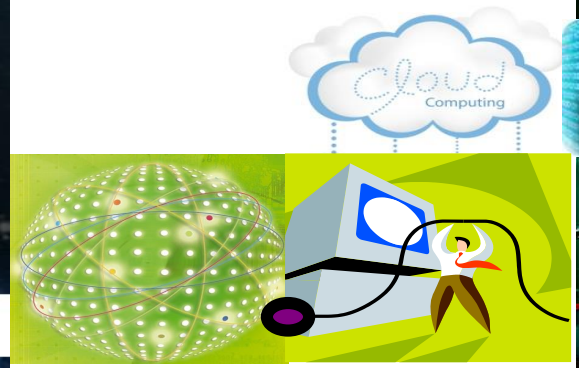
- The main challenges
- How we address them

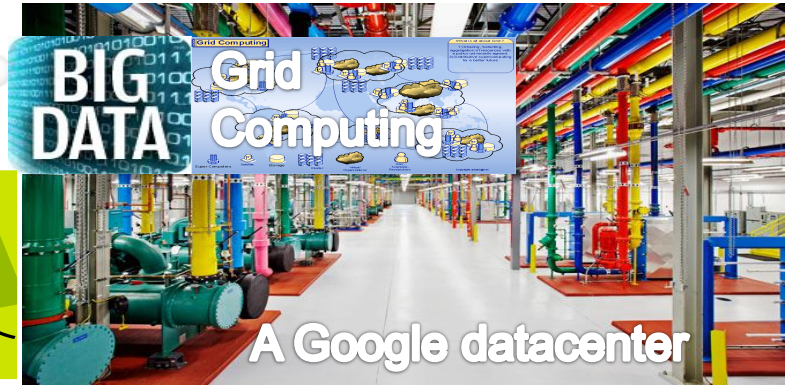
~20’ — Massivizing Computer Systems: Examples

1. The Ecosystem Navigation Challenge
2. The New World Challenge
3. The Scheduling Challenge

~2’ — Conclusion

WHAT DOES OUR SOCIETY NEED? THE QUADRUPLE HELIX, ICT-BASED





prosperous society + blooming economy + inventive academia + wise governance

- Enable data access & processing as a fundamental right in Europe (2018: GDPR v.1.0)
- Enable Industry 4.0 and big science (2020: €100 bn., 1 mil. jobs)
- Sustainability, dependability, and privacy, but with energy footprint <5%
- etc. etc. etc.

VU

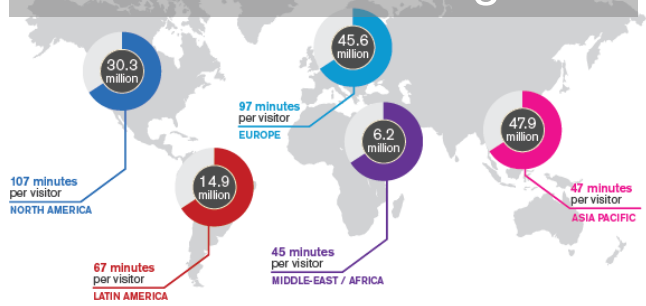
THIS IS THE GOLDEN AGE OF DISTRIBUTED COMPUTER SYSTEMS

Education for Everyone (Online)

Business Services

Big Science

Online Gaming



ABN·AMRO

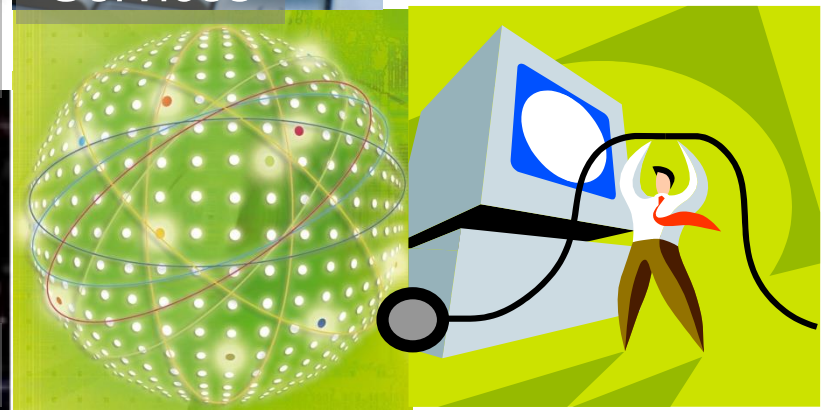
Daily Life

THIS IS THE GOLDEN AGE OF DISTRIBUTED COMPUTER SYSTEMS

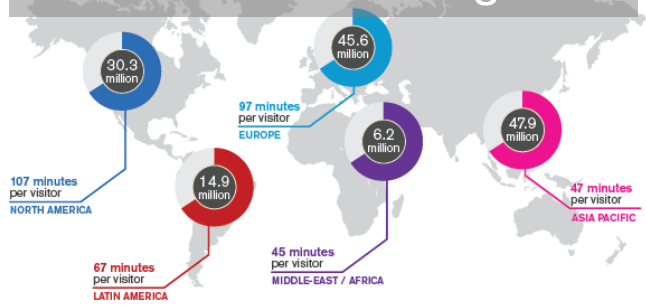
Education for Everyone (Online)

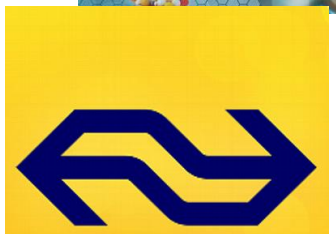
Business Services

Big Science



Online Gaming

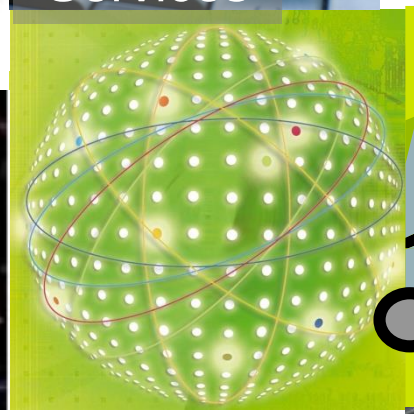


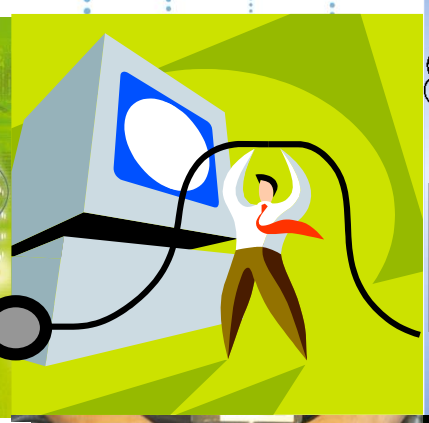


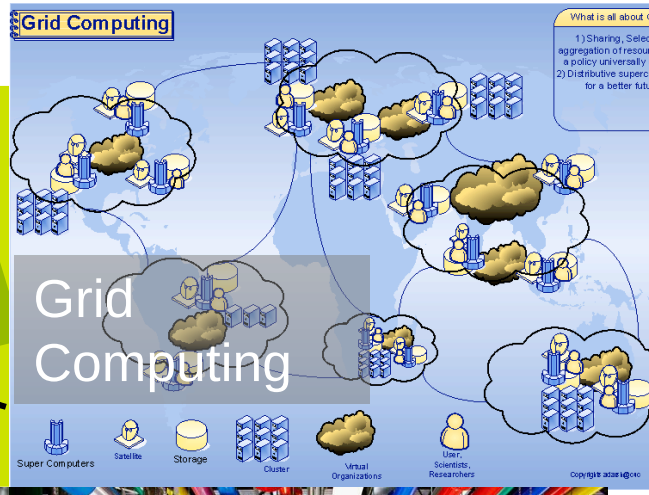
ABN·AMRO

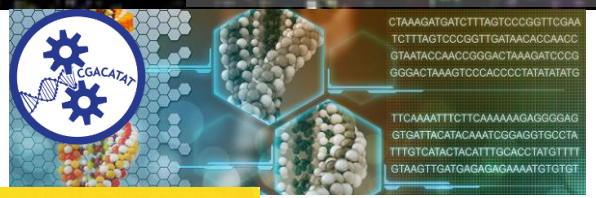
Daily Life

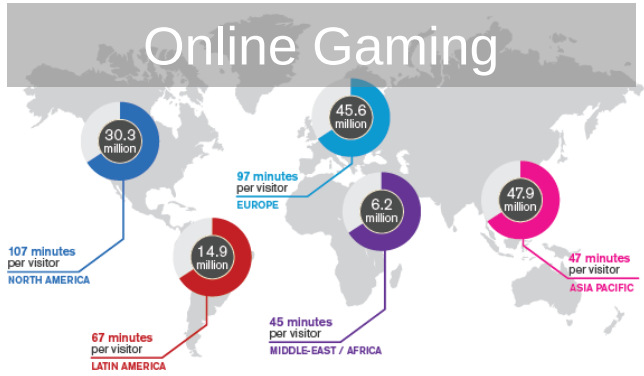
THIS IS THE GOLDEN AGE OF DISTRIBUTED COMPUTER SYSTEMS

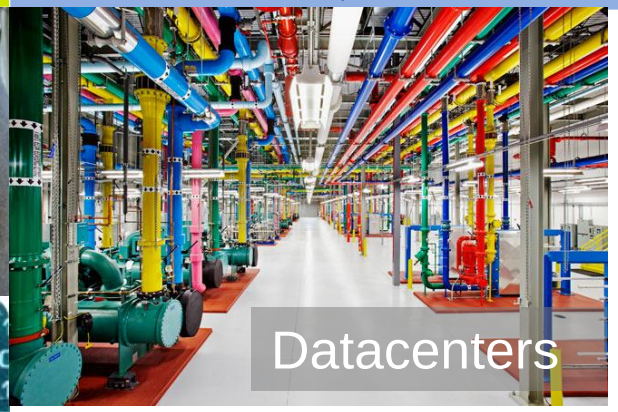






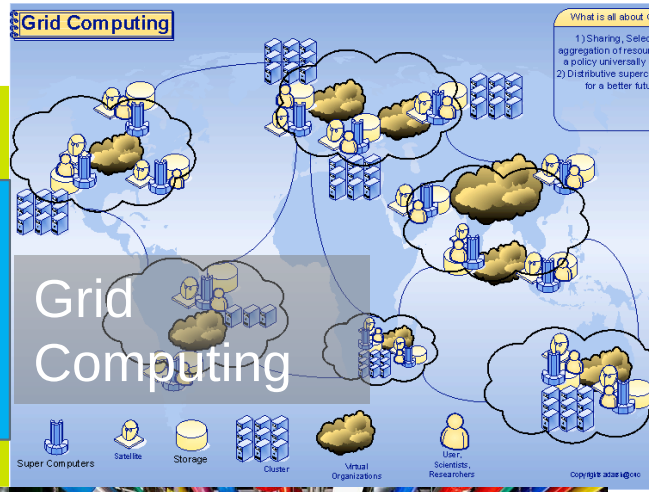


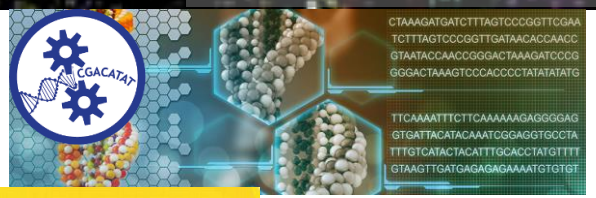


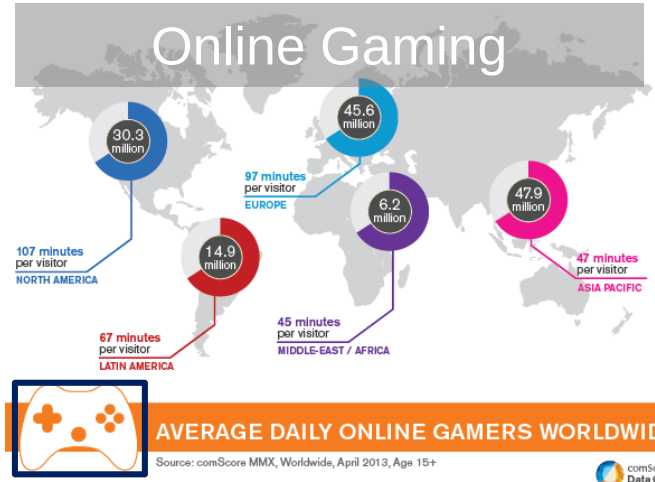


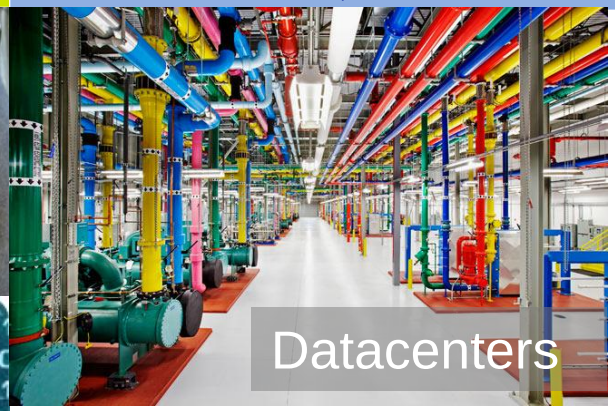
THIS IS THE GOLDEN AGE OF DISTRIBUTED COMPUTER SYSTEMS

Here is how this works…





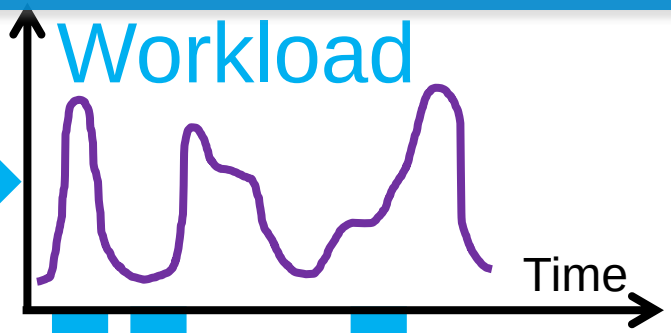


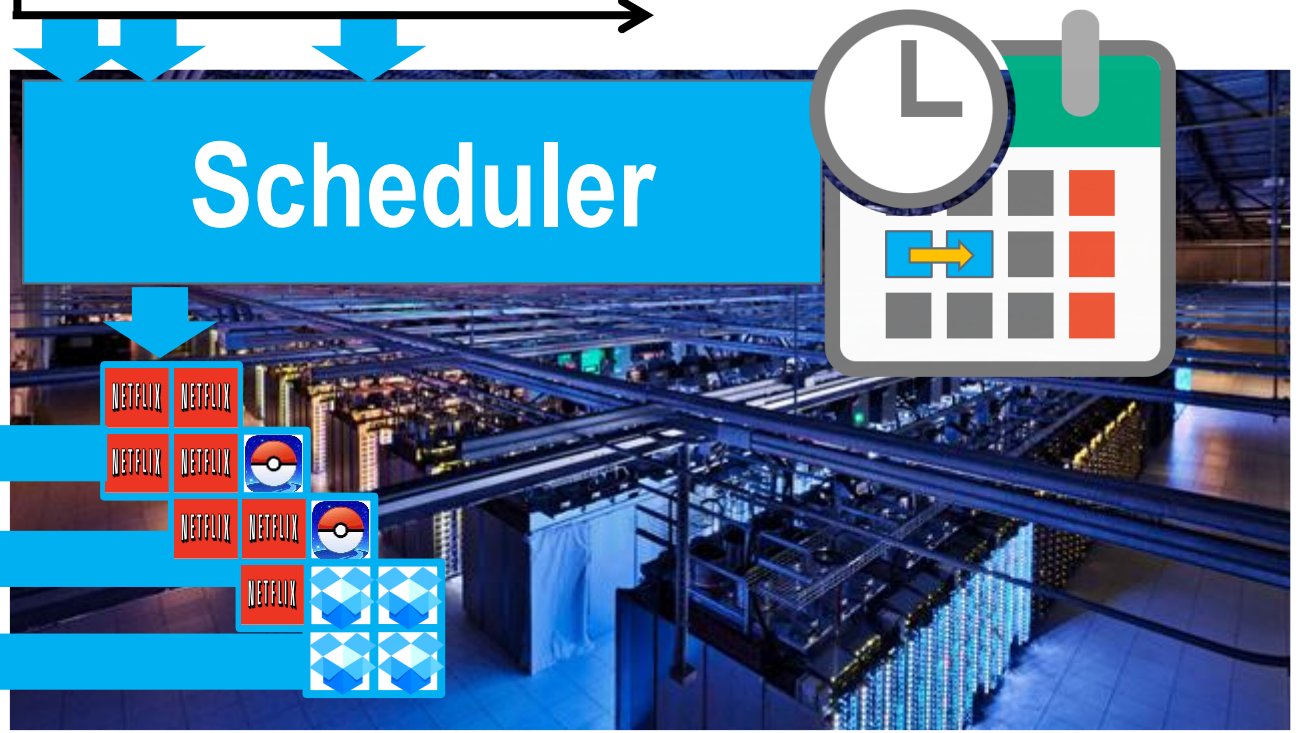


THE CURRENT TECHNOLOGY STACK: DATACENTER, SCHEDULER

Creators

Digital Services





Datacenter (full-/micro-/nano-)

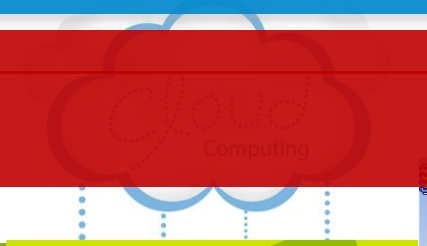
Performance, Dependability, Efficiency

THIS IS THE GOLDEN AGE OF DISTRIBUTED COMPUTER SYSTEMS

YET WE ARE IN A CRISIS

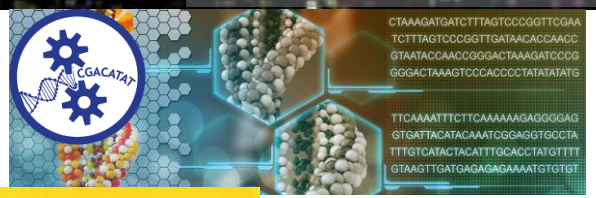
Education for Everyone (Online)

Services

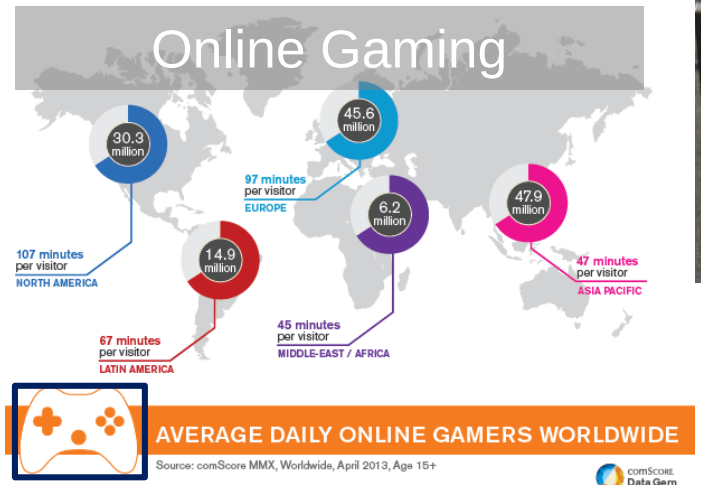


Big Science

A Crisis? What crisis?!

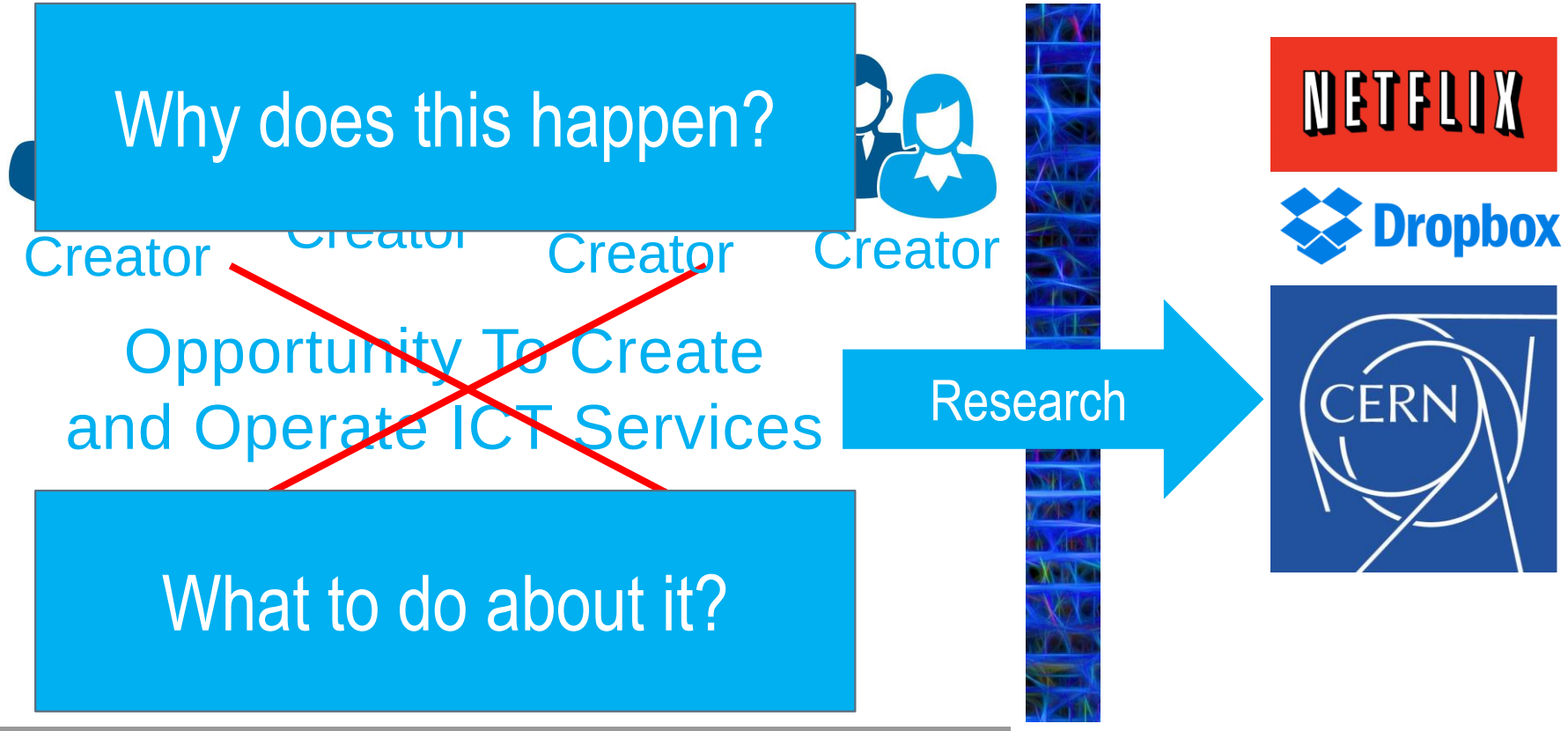


Daily Life



Datacenters

THE CRISIS: IN THIS DIGITAL ECONOMY, FEW CAN BE SUCCESSFUL!



"ICT is vital for SMEs", "SMEs are 60% GDP"

Sources: Eurostat'15,EC Digital Agenda,IDC'14

VU UNIVERSITY AMSTERDAM Faculty of Sciences

THE CRISIS: IN THIS DIGITAL ECONOMY, FEW CAN BE SUCCESSFUL!

THE COMPLEXITY CHALLENGE

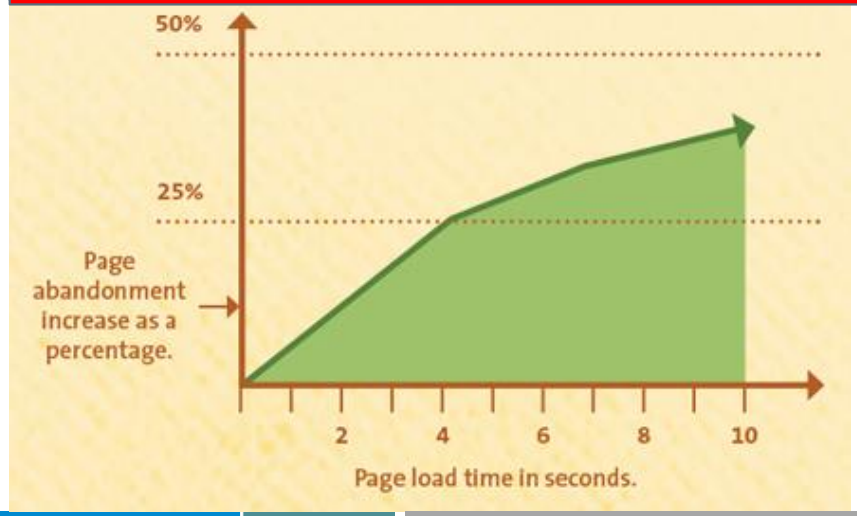
1. We Build and Test
Isolated Computer Systems, Yet
Everything Works in Ecosystems

<<1% OF BIG DATA
BY MATT TURK (2017)

THE CRISIS: IN THIS DIGITAL ECONOMY, FEW CAN BE SUCCESSFUL!

PERFORMANCE, DEPENDABILITY, AND OTHER NON-FUNCTIONAL CHALLENGES

2. We Cannot Even Maintain the Ecosystems we Have Built (and Tested, and Validated)



Google goes dark for 2 minutes, kills 40% of world's net traffic www.theregister.co.uk/2013/08/17/google_outage/

Systemwide outage knocks every service offline

Sources: https://www.fastcompany.com/1825005/how-one-second-could-cost-amazon-16-billion-sales

THE CRISIS: IN THIS DIGITAL ECONOMY, FEW CAN BE SUCCESSFUL!

THE RESOURCE MANAGEMENT CHALLENGE

Based on Jay Walker's recent TED talk.

3. Need To Be Much More Efficient,

5. Need to Also Be Ethical, and to Also Educate Our Customers

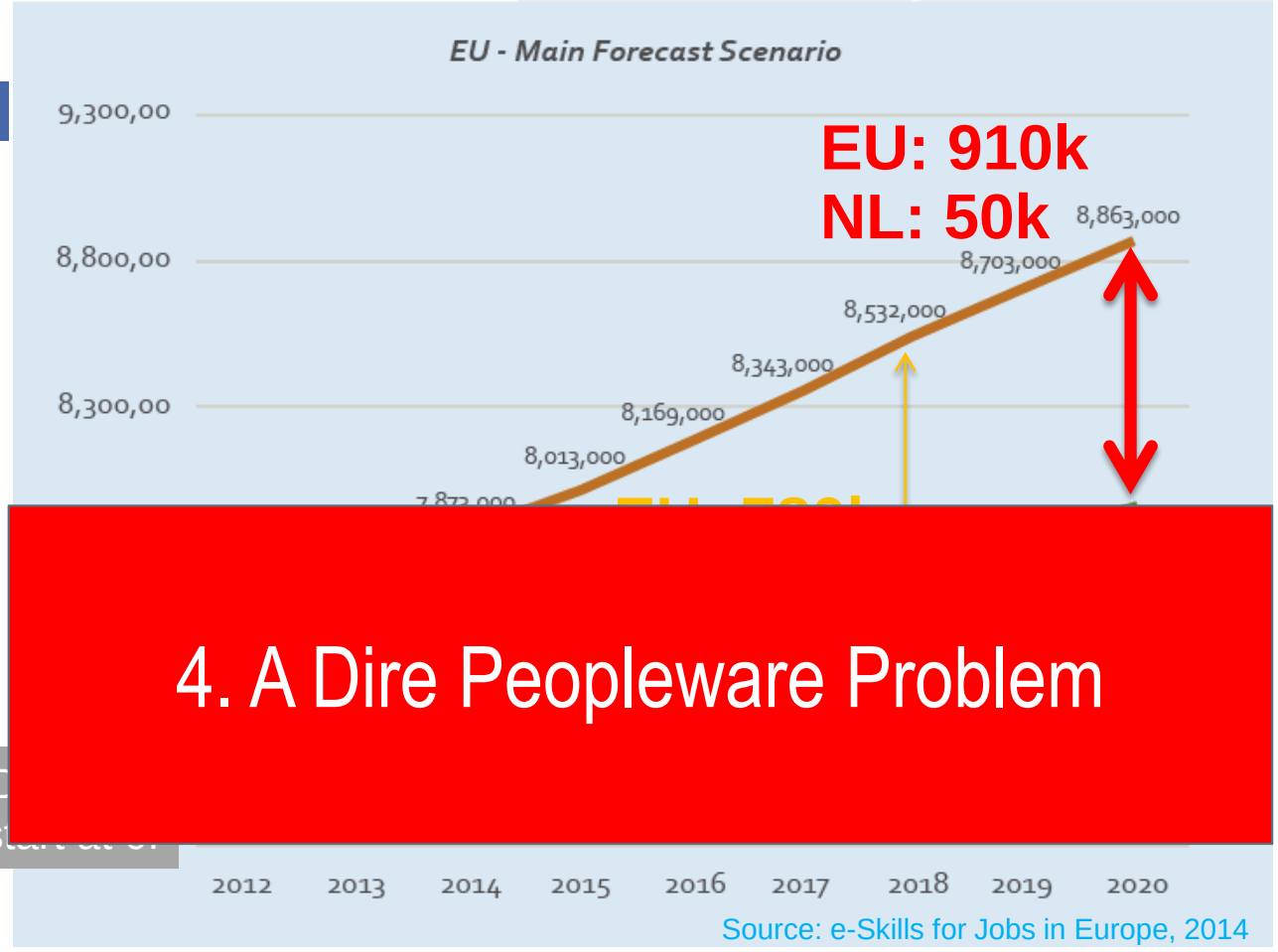
PSY Gangnam consumed ~500GWh

= more than entire countries* in a year (*41 countries),
= over 50MW of 24/7/365 diesel, 135M liters of oil,
= 100,000 cars running for a year, ...

Source: Ian Bitterlin and Jon Summers, UoL, UK, Jul 2013.
Note: Psy has >3.5 billion views (May 2018).

THE CRISIS: IN THIS DIGITAL ECONOMY, FEW CAN BE SUCCESSFUL!

THE WORKFORCE GAP, IN THE NETHERLANDS & IN EUROPE



Source: e-Skills for Jobs in Europe, 2014

4. A Dire Peopleware Problem

THIS IS THE GOLDEN AGE OF DISTRIBUTED COMPUTER SYSTEMS

YET WE ARE IN A CRISIS – 5 CORE PROBLEMS

1. The Current Laws and Theories Are Built and Tested for Isolated Computer Systems

TRADITIONAL DISTRIBUTED SYSTEMS COURSES TEACH YOU ALL ABOUT THIS

2. Need to Understand How to Maintain Ecosystems

3. Need to Understand How to Make Ecosystems Automated, Efficient (Smarter)

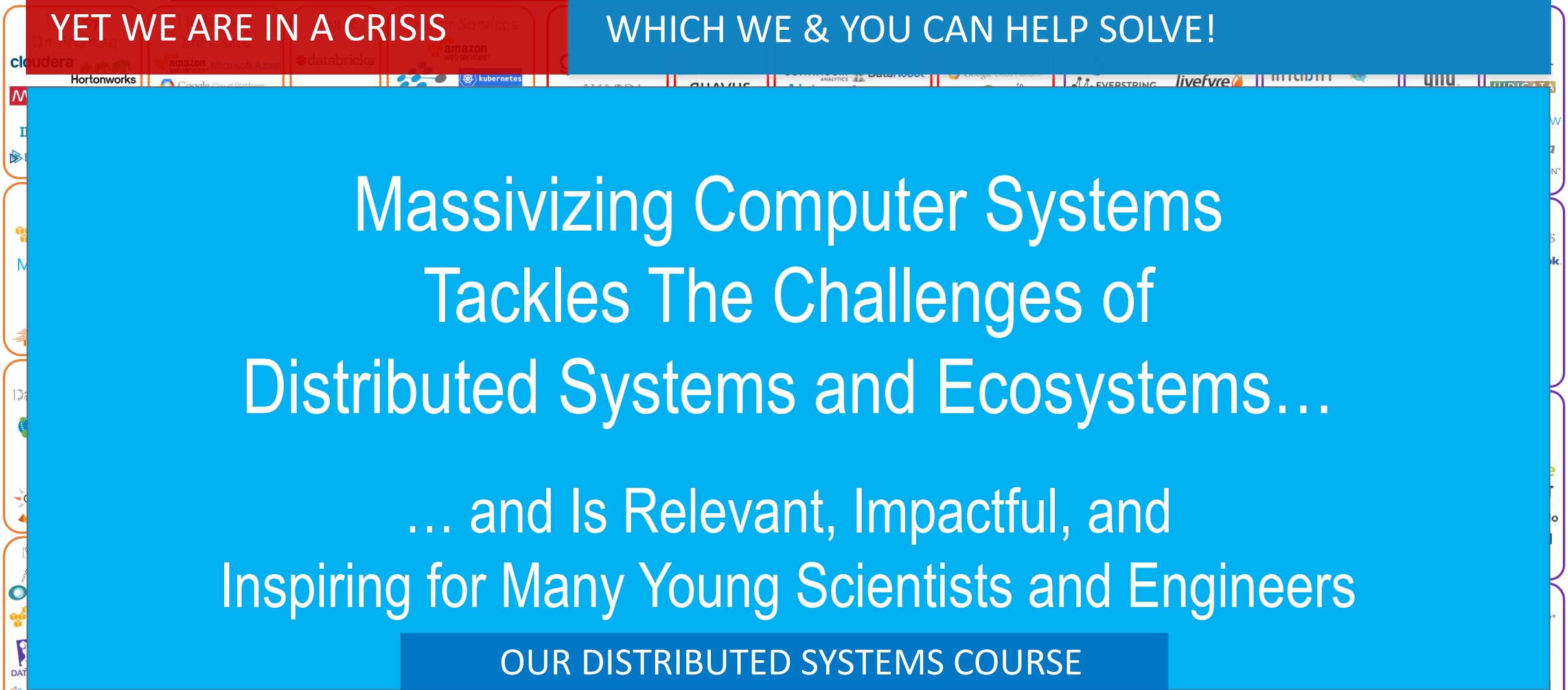
4. Need to Address the Peopleware Problems

5. Beyond Tech: Need to Also Be Ethical

THIS IS THE GOLDEN AGE OF DISTRIBUTED COMPUTER SYSTEMS

YET WE ARE IN A CRISIS

WHICH WE & YOU CAN HELP SOLVE!



Massivizing Computer Systems Tackles The Challenges of Distributed Systems and Ecosystems…

… and Is Relevant, Impactful, and Inspiring for Many Young Scientists and Engineers

OUR DISTRIBUTED SYSTEMS COURSE

THIS IS THE GOLDEN AGE OF DISTRIBUTED COMPUTER SYSTEMS

MASSIVIZING COMPUTER SYSTEMS IN A NUTSHELL

WHO?	SCIENTISTS, ENGINEERS, DESIGNERS, MANAGERS, ETC.
WHAT? MAIN GOAL	UNDERSTAND AND CONTROL DISTRIBUTED ECOSYSTEMS, TO TURN THEM INTO EFFICIENT, AUTOMATED UTILITIES
HOW? CENTRAL PARADIGM	ECOSYSTEM OPERATION AND CHARACTERISTICS DERIVES NON-TRIVIALLY FROM ITS SYSTEMS AND USERS (RECURSIVELY)
WHICH APPROACH?	MODERN DISTRIBUTED SYSTEMS AND PROBLEM-SOLVING

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

THE SCIENCE, DESIGN, AND ENGINEERING OF MASSIVIZING

Scheduling
Serverless
Workflows
Portfolio, Auto-scaling*

Dependability
Performance & Failure Analysis*
Space-/Time-Correlation
Availability-On-Demand

New World+
Workload Modeling
Business-Critical
Online Gaming

Ecosystem Navigator+
Performance Variability
Grid*, Cloud, Big Data
Benchmarking*
Longitudinal Studies

Scalability/Elasticity+
Delegated Matchmaking*
BTWorld*, POGGI*, AoS
Auto-Scalers
Heterogeneous Systems

Socially Aware+
Collaborative Downloads*
Groups in Online Gaming
Toxicity Detection*
Interaction Graphs

Education
Social Gamification*

Software Artifacts
Graphalytics, OpenDC

Data Artifacts
Distributed Systems Memex*

Fundamental Problems/Research Lines
Our Contribution So Far
Competitive personal grants
+ Please ask for a definition
* Award-level work

MASSIVIZING COMPUTER SYSTEMS

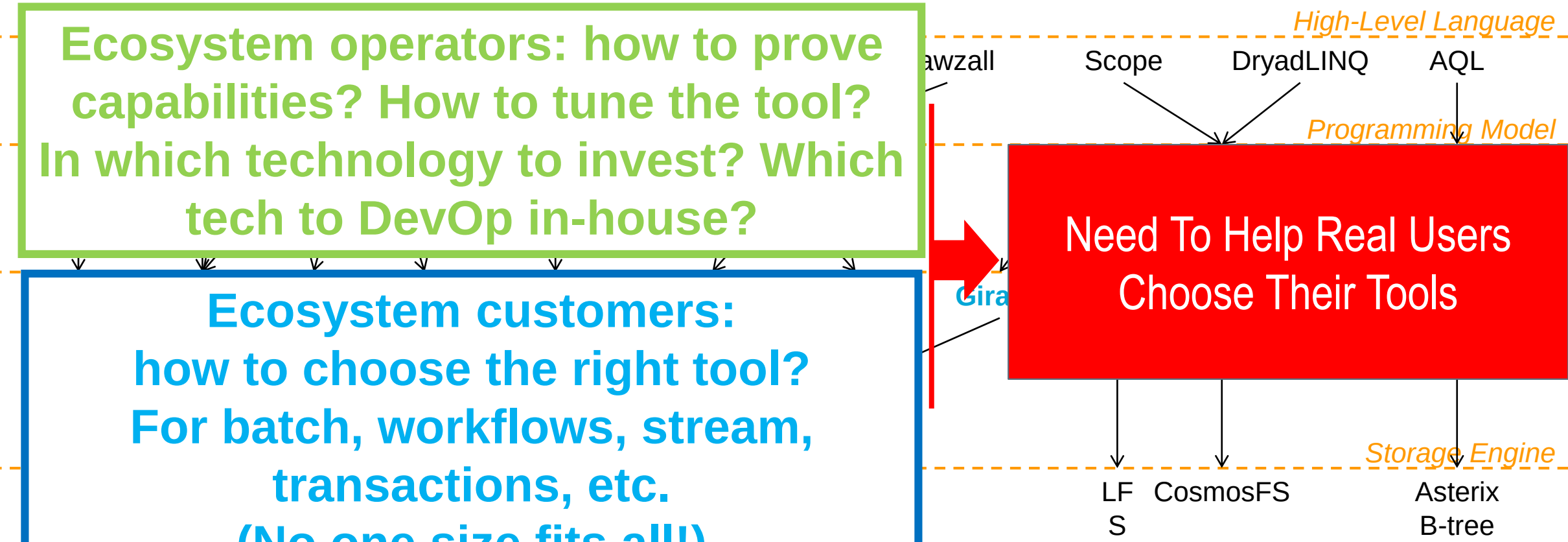
WHICH APPROACH?	MODERN DISTRIBUTED SYSTEMS AND PROBLEM-SOLVING
HOW TAUGHT?	AWARD-WINNING METHOD, BASED ON GAMIFICATION

- Choose your own path:
  - > The Ecosystem Navigation Challenge (Understanding + Exp.)
  - > The New World Challenge (Abstraction + Design)
  - > The Scheduling Challenge (Design + Operation)

The Ecosystem Navigation Challenge

Ecosystem operators: how to prove capabilities? How to tune the tool? In which technology to invest? Which tech to DevOp in-house?

Ecosystem customers: how to choose the right tool? For batch, workflows, stream, transactions, etc. (No one size fits all!)



Batch data processing ecosystem in 2011. Our latest update covers ecosystems until 2018.

Alexandru Iosup
Chair

Nikolas Herbst
Vice-Chair

The SPEC RG Cloud Group

Methodology, Benchmarking, and Performance Analysis of Cloud Systems and Applications

"A broad approach, **relevant for both academia and industry**, to cloud benchmarking, quantitative evaluation, and experimental analysis."
"To develop new **methodological elements** for gaining deeper understanding not only of **cloud performance**, but also of **cloud operation and behavior"**
"… through diverse quantitative evaluation tools"

http://research.spec.org/working-groups/rg-cloud-working-group.html

A General Approach for Ecosystem Benchmarking (Works for Cloud and Big Data Benchmarking)

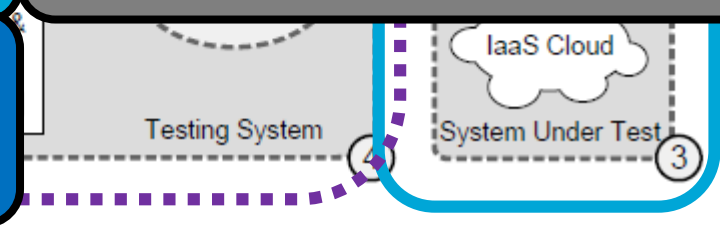
Q1: What is the performance of production IaaS cloud services?

Q4: What is the performance of graph-processing systems?

Q2: How variable is the performance of widely used production cloud services?

+ many other questions…

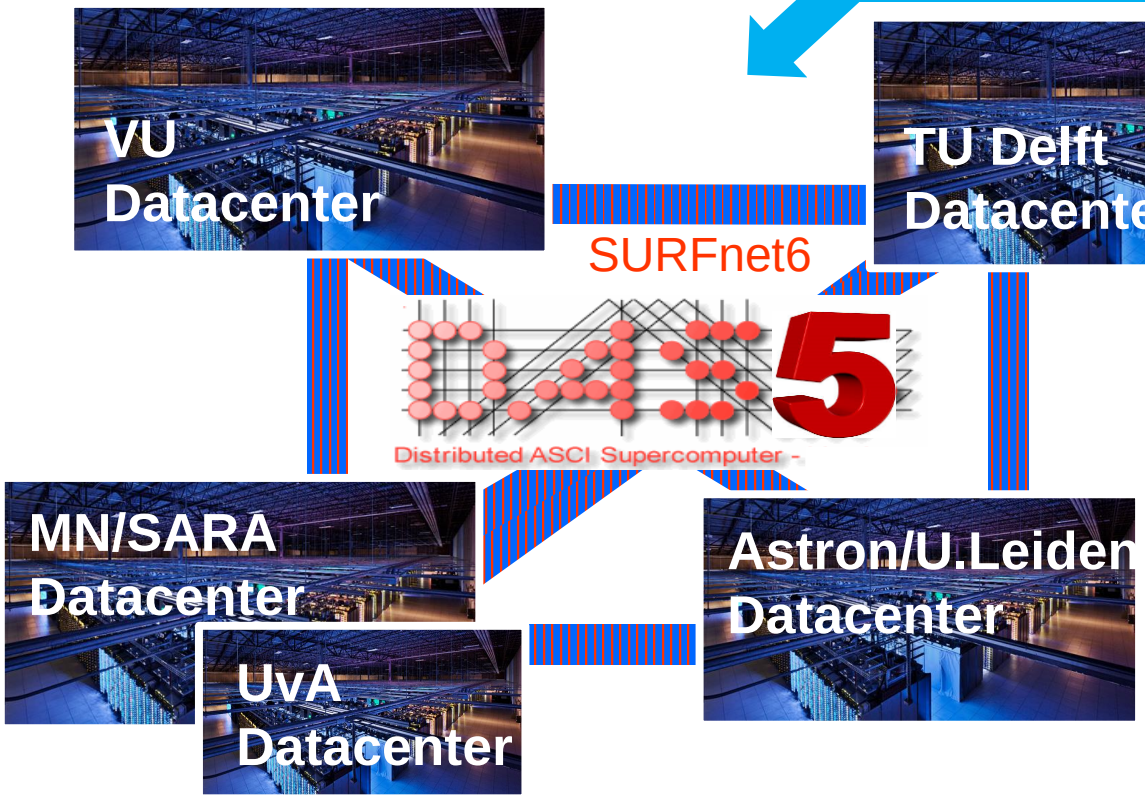
Q3: What is the performance of IaaS-cloud auto-scalers?



Iosup, Prodan, Epema. IaaS Cloud Benchmarking: Approaches, Challenges, and Experience. Cloud Computing for Data-Intensive Applications 2014: 83-104

VU VRIJE UNIVERSITEIT AMSTERDAM

Experimental Research Methodology
Our Main Scientific Instrument: DAS-5



\+ amazon web services™

We also use IaaS clouds

\+ OpenDC

And simulators

300+ scientists as users

We won IEEE Scale Challenge 2014

Our Method

Q1

- General performance technique, adapted to clouds: model performance of individual components; system performance is performance of workload + model [Saavedra and Smith, ACM TOCS'96]

```
Iosup et al., Performance Analysis of Cloud Computing Services for
  Many Tasks Scientific Computing, IEEE TPDS 2011.
  Highest cited article in the best journal of the field (2009-2015).
```

Q2

- Performance traces from CloudStatus
  - All Amazon AWS and Google GAE services
  - Periodic performance probes, Sampling rate under 2 minutes
- Simulations to assess impact of performance variability
  - Based on traces collected from other applications

```
Iosup, Yigitbasi, Epema. On the Performance Variability of
  Production Cloud Services, IEEE CCgrid 2011.
  Highest cited study on performance variability in clouds.
```

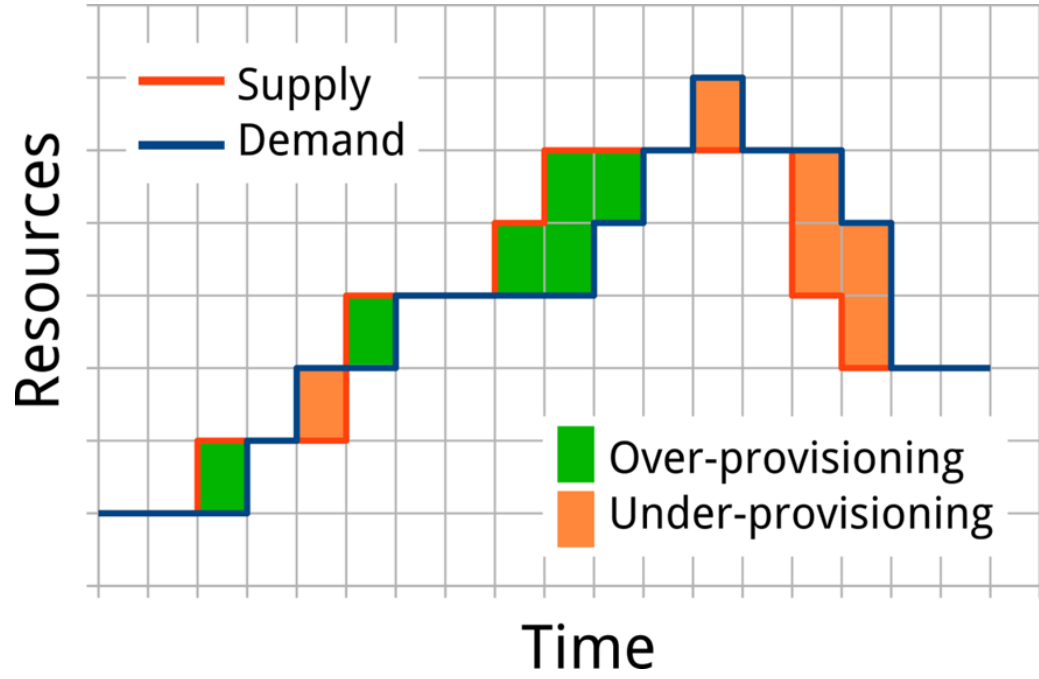
VU VRIJE UNIVERSITEIT AMSTERDAM

Our Method

Q3

Ilyushkin et al. An Experimental Performance Evaluation of Autoscaling Policies for Complex Workflows. ICPE 2017.
Nominated for Best Paper Award.

- Real-world experiments (1 workload)
- Later: simulations (more parameters)



Q4

- Many classes of algorithms used in practice
- Diverse real and synthetic datasets
  - Recently, Broido and Clauset found that power-laws are rare in graphs
- Diverse set of experiments representative for practice

Iosup et al. LDBC Graphalytics: A Benchmark for Large-Scale Graph Analysis on Parallel and Distributed Platforms. PVLDB 9(13): 1317-1328 (2016) Collaboration academia-industry.

VU VRIJE UNIVERSITEIT AMSTERDAM

Main Findings

Q1

- Lower performance than theoretical peak in IaaS services
  - Especially CPU (GFLOPS)
  - Not explained by traditional models

Q2

- Performance variability in IaaS and PaaS services
  - Explored in longitudinal study of Amazon Web Services and Google App Engine
  - Not captured in traditional models

- Compared performance of IaaS clouds with many commercial alternatives, such as supercomputers and clusters

Main Findings

Q3

- Explored impact of auto-scaler (+ ecosystem of schedulers) on over 10 facets of elasticity
  - Findings not explained by traditional models

Q4

- The HPAD model for the performance of graph-processing systems ~ replaces previous theories
  - Performance is function of Platform, but also Dataset and Algorithm
  - With configurable many-/multi-cores, the Hardware also is also crucial for performance

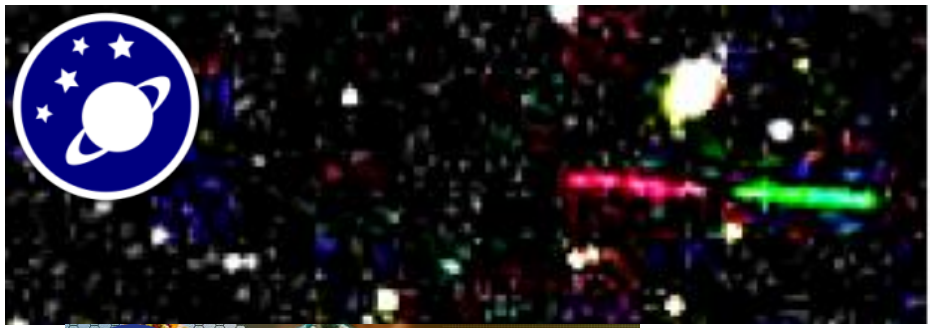
MASSIVIZING COMPUTER SYSTEMS

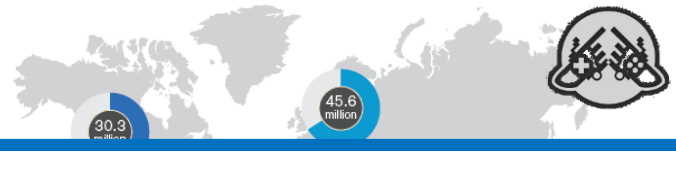
WHICH APPROACH?	MODERN DISTRIBUTED SYSTEMS AND PROBLEM-SOLVING
HOW TAUGHT?	AWARD-WINNING METHOD, BASED ON GAMIFICATION

- Choose your own path:
  - > The Ecosystem Navigation Challenge (Understanding + Exp.)
  - > The New World Challenge (Abstraction + Design)
  - > The Scheduling Challenge (Design + Operation)

The New World Challenge

Ecosystem operators: new value-adding services, new workloads including FaaS, DevOps workloads





Need Operational Models

Ecosystem customers: new apps, new services, micro-services, customers can become operators (recursive value-chain)

Source: comScore MMX, Worldwide, April 2013, Age 15+

comScore Data Gem

Erwin
van Eyk

Alexandru
Iosup

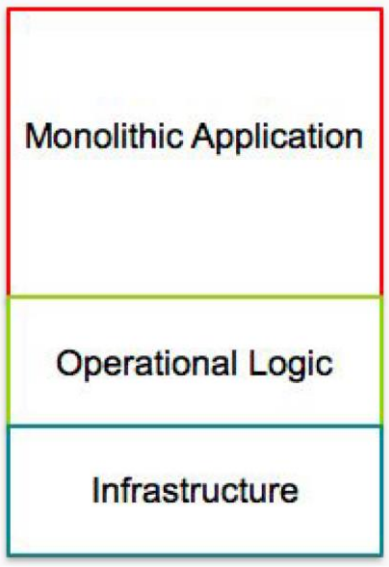
Serverless / FaaS Execution

Vision and Architecture for Serverless Execution in Cloud Environments

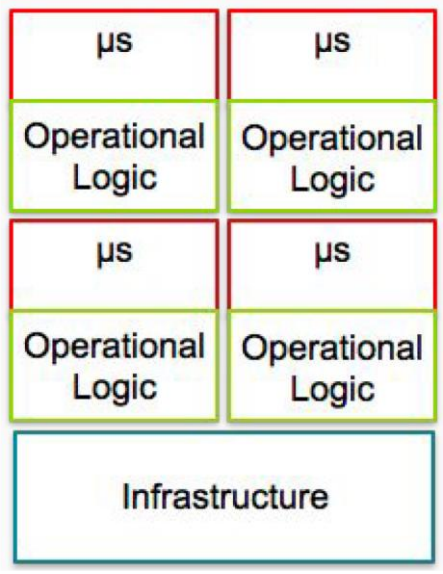
Erwin Van Eyk, Alexandru Iosup, Cristina L. Abad, Johannes Grohmann, Simon Eismann: A SPEC RG Cloud Group's Vision on the Performance Challenges of FaaS Cloud Architectures. ICPE 2018.

Erwin van Eyk, Simon Seif (SAP), Markus Thoemmes (IBM Germany), Alexandru Iosup. The SPEC Cloud Group's Research Vision on FaaS and Serverless Architectures. Workshop on Serverless Computing (WoSC'17), held in conjunction with Middleware'17.

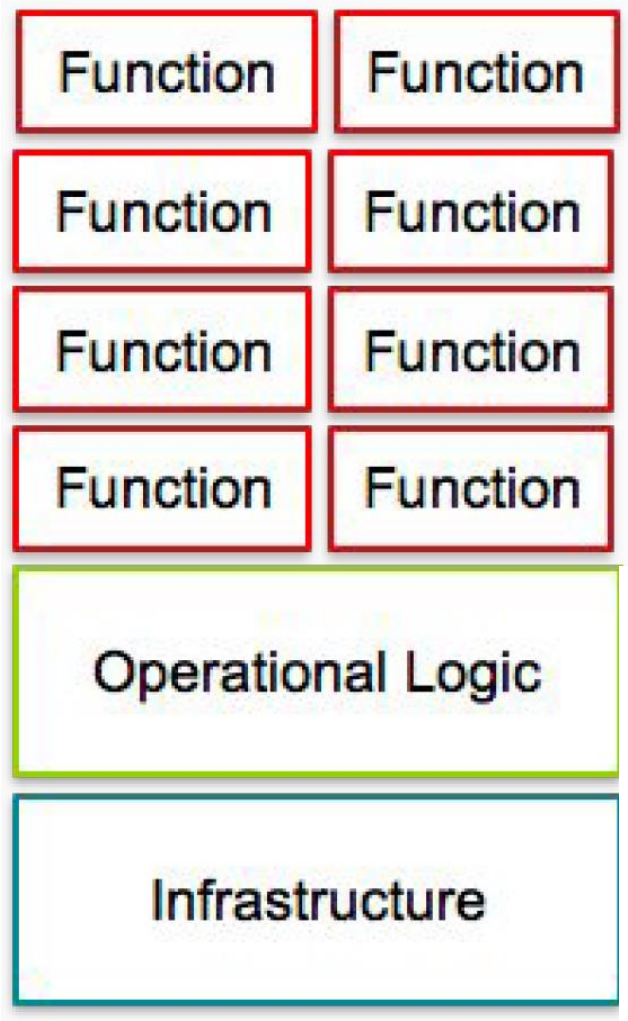
From Monoliths to Microservices to FaaS



- Difficult to Scale
- Infrequent
- Inflexible
- Complex deployment
- Tightly coupled stack



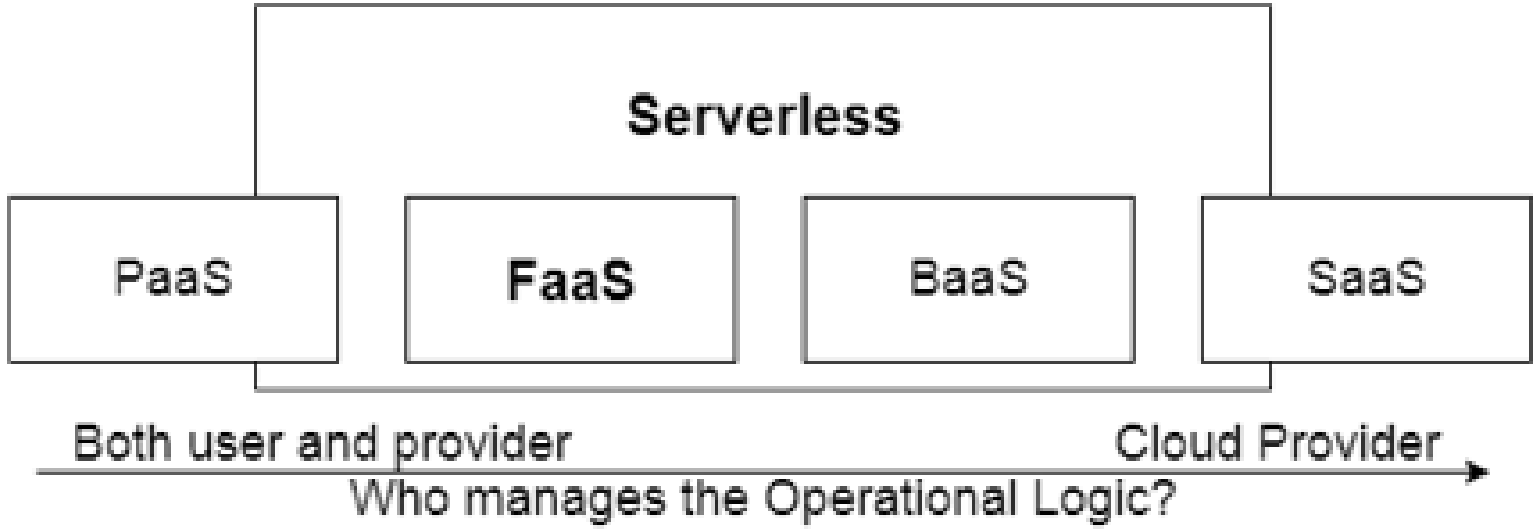
- Scalable
- Frequent
- Flexible
- Complexity: from application logic to operational logic.
- Need for DevOps



- Scalable
- Frequent
- Flexible
- Explicit separation of Business Logic vs. Operational Logic.
- Minimal layer coupling, unit of deployment

VU VRIJE UNIVERSITEIT AMSTERDAM

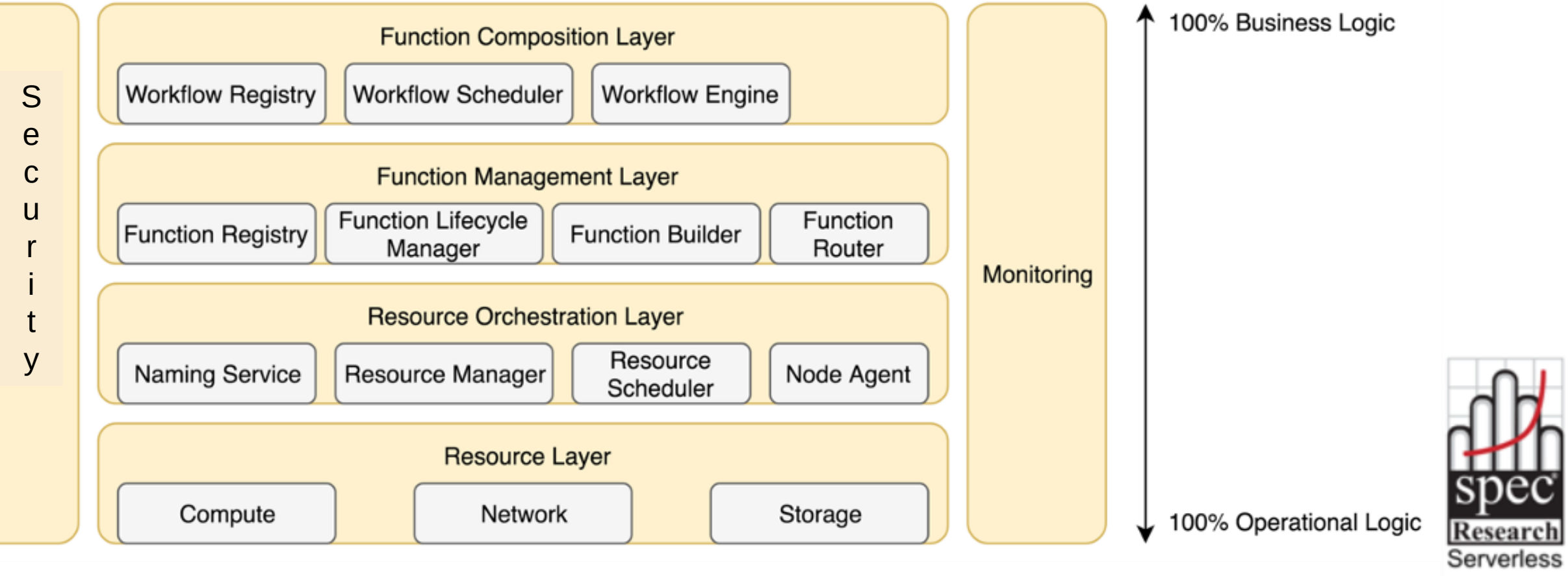
Serverless and FaaS



- **Serverless**
  - (Almost) no operational logic
  - Event-Driven
  - Granular billing
- **FaaS**
  - A form of *serverless computing*
  - User provides a function
  - Function deployed and managed by cloud provider

VU

Reference Architecture for FaaS Management



spec Research Serverless

Workflow Management Architecture in Fission.io

Designed by Erwin van Eyk during internship at Platform9, in collaboration w/ Platform9 team and Alexandru Iosup.

1 Core Function / 2 AI Server

- Exposes all actions through API

3 Event Store / 4 Projector

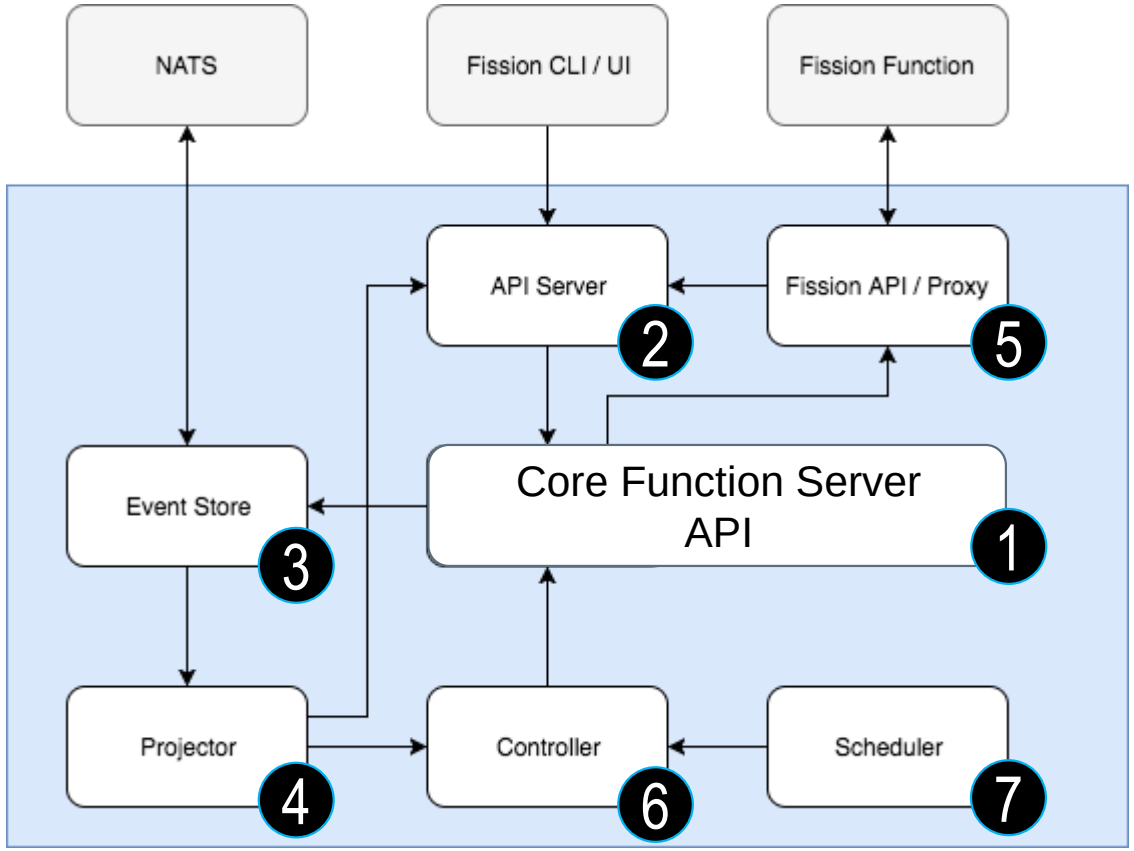
- Events update the workflow
- Store has Pub/Sub functionality
- Projector builds current state

5 Fission Proxy

- API access to Fission FaaS

6 Controller / 7 Scheduler

- Workflow manager



https://github.com/fission/fission-workflows/blob/master/Docs/architecture.md

MASSIVIZING COMPUTER SYSTEMS

WHICH APPROACH?	MODERN DISTRIBUTED SYSTEMS AND PROBLEM-SOLVING
HOW TAUGHT?	AWARD-WINNING METHOD, BASED ON GAMIFICATION

- Choose your own path:
  - > The Ecosystem Navigation Challenge (Understanding + Exp.)
  - > The New World Challenge (Abstraction + Design)
  - > The Scheduling Challenge (Design + Operation)

The Scheduling Challenge

amazon web services™ Windows Azure rackspace

Ecosystem operators:

Which resources to lend?
Where to place? Penalty v reward?

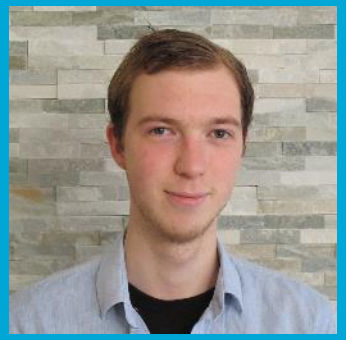
Need scheduling policies for both users and operators

Ecosystem customers:

Which resources to lease?
When? How many? When stop?
Utility functions?

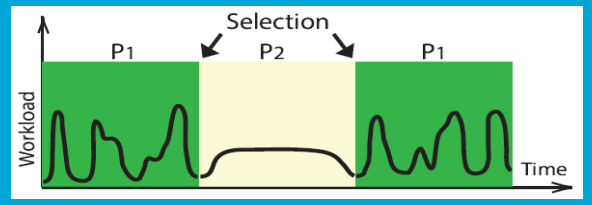
Vincent van Beek

Tim Hegeman



Jesse Donkervliet

Alexandru Iosup



Portfolio Scheduling for DCs

Self-Expressive Management of Business-Critical Workloads in Virtualized Datacenters

van Beek, Donkervliet, Hegeman, Hugtenburg, Iosup. Self-Expressive Management of Business-Critical Workloads in Virtualized Datacenters. IEEE Computer 48(7): 46-54 (2015)

Deng, Song, Ren, Iosup. Exploring portfolio scheduling for long-term execution of scientific workloads in IaaS clouds. SC 2013: 55:1-55:12

Massivizing Distributed Systems

Scheduling
Bags-Of-Tasks
Workflow
Mixed-Workload
Portfolio

Dependability
Failure Analysis*
Space-/Time-Correlation
Availability-On-Demand

New World
Workload Modeling
Interaction Graphs
Business-Critical
Online Gaming

1st time in DCs

Ecosystem Navigation
Performance Variability
Grid*, Cloud, Big Data
Benchmarking
Longitudinal Studies

Scalability/Elasticity
Delegated Matchmaking*
POGGI*
Area-Of-Simulation
BTWorld*
Auto-Scalers

Socially Aware Techniques
Collaborative Downloads*
Groups in Online Gaming
Toxicity Detection*

Software Artifacts
Graphalytics, etc.

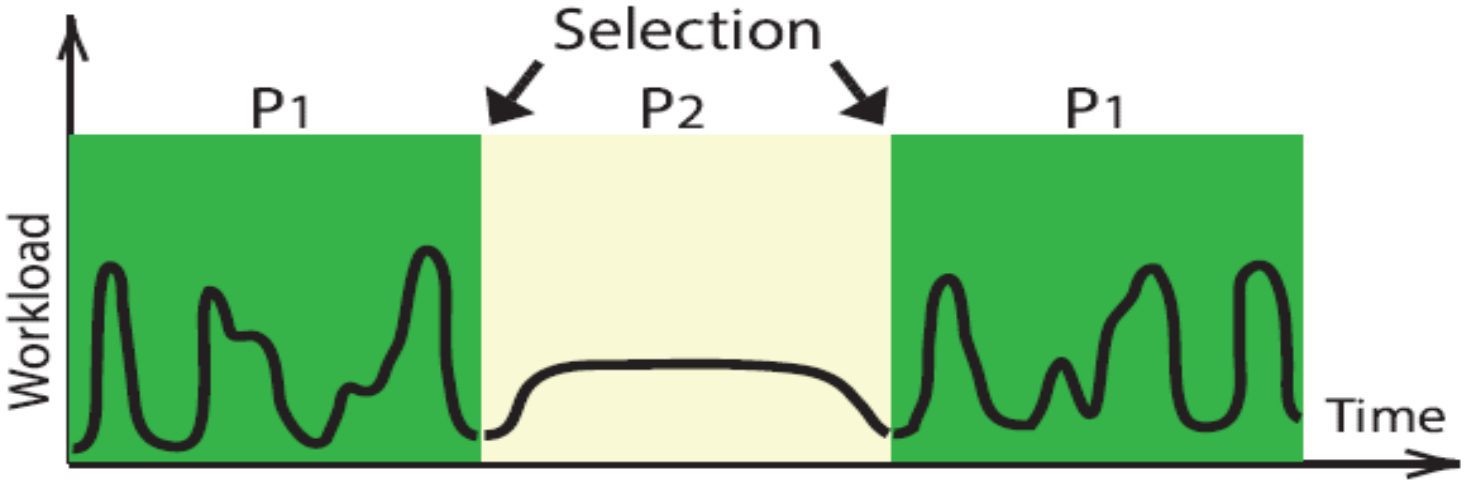
Data Artifacts
A Distributed Systems Memex*

Fundamental Problems

Our Contribution So Far (* Award-winning)

Portfolio Scheduling, In A Nutshell

- Datacenters cannot work without one or even several schedulers
- Instead of ephemeral, risky schedulers, we propose to



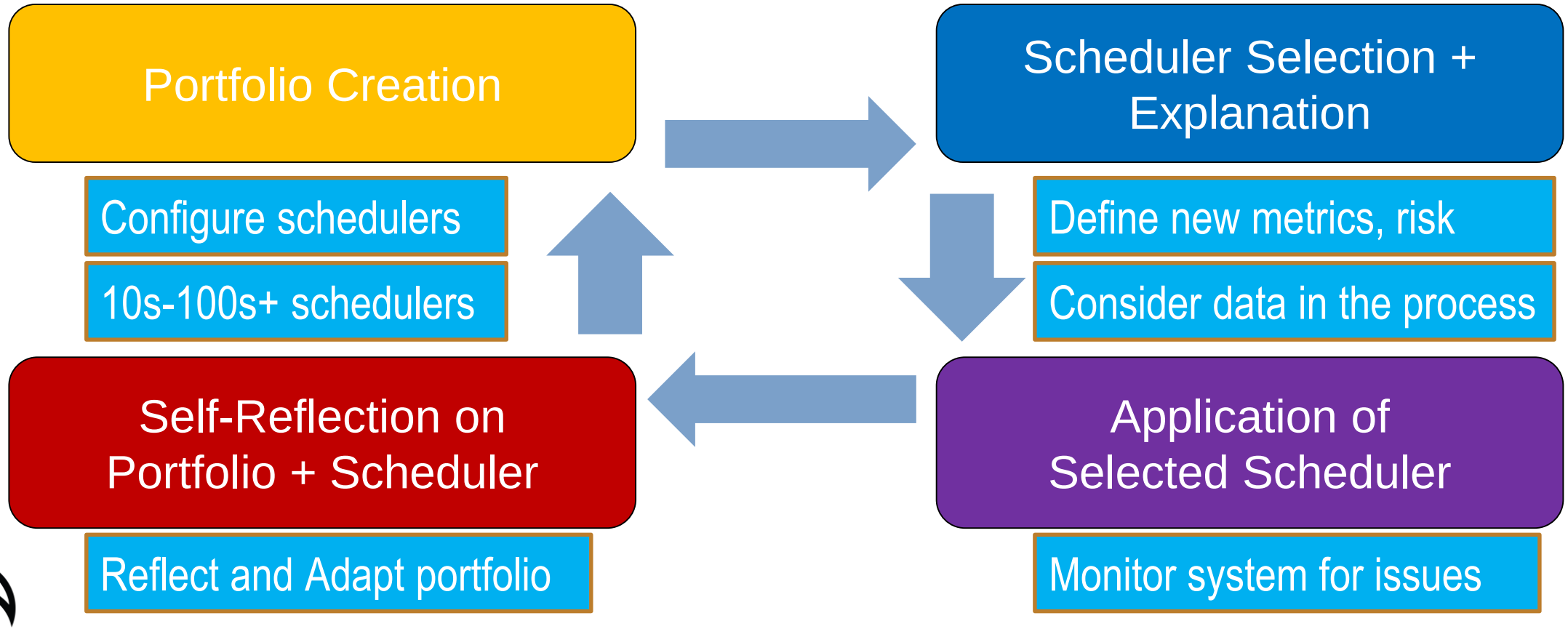
1. Create a set of schedulers (resource provisioning and allocation policies)
2. Select active scheduler online, apply for the next period, analyze results

(Repeat)

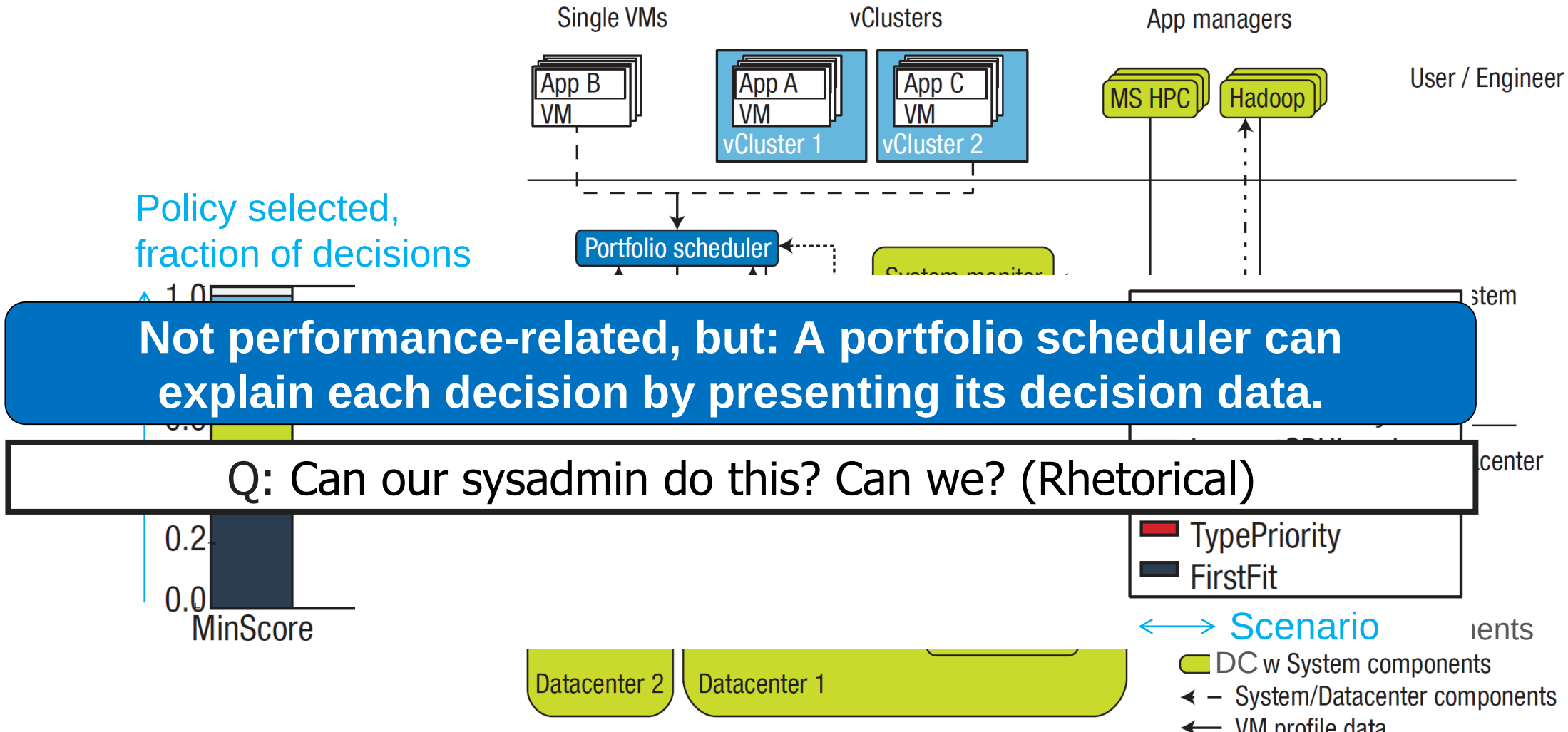
K. Deng et al. Exploring portfolio scheduling for long-term execution of scientific workloads in IaaS clouds. SC|13

Portfolio Scheduling

Portfolio Scheduling for Computer Systems



Portfolio Scheduling in Massive Datacenters



Not performance-related, but: A portfolio scheduler can explain each decision by presenting its decision data.

Q: Can our sysadmin do this? Can we? (Rhetorical)

V. van Beek et al. Mnemos: Self-Expressive Management of Business-Critical Workloads in Virtualized Datacenters. IEEE Computer 2015

MASSIVIZING COMPUTER SYSTEMS

= MAKING COMPUTER SYSTEMS SCALABLE, RELIABLE, PERFORMANT, ETC., YET ABLE TO FORM AN EFFICIENT ECOSYSTEM

- Golden Age of Distributed Systems
- Massivizing Computer Systems means modern distributed systems
  - Think Ecosystems
  - Methods to address key challenges in science, design, and engineering
  - Teaching facilitated by award-winning method
- Plenty of challenges
  - You can help! You can make a career, in science / industry.

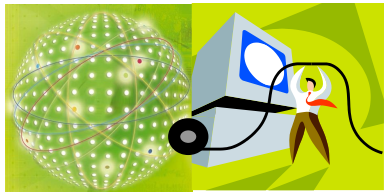
VU VRIJE UNIVERSITEIT AMSTERDAM

MASSIVIZING COMPUTER SYSTEMS

FURTHER READING

1. Iosup et al. Massivizing Computer Systems. ICDCS 2018 (in print)
2. Ilyushkin et al. An Experimental Performance Evaluation of Autoscaling Policies for Complex Workflows. ICPE 2017.
3. Iosup et al. LDBC Graphalytics: A Benchmark for Large-Scale Graph Analysis on Parallel and Distributed Platforms. PVLDB 2016.
4. Guo et al.: Design and Experimental Evaluation of Distributed Heterogeneous Graph-Processing Systems. CCGrid 2016.
5. van Beek et al.: Self-Expressive Management of Business-Critical Workloads in Virtualized Datacenters. IEEE Computer 2015.
6. Jia et al.: Socializing by Gaming: Revealing Social Relationships in Multiplayer Online Games. TKDD 2015.
7. Ghit et al. Balanced resource allocations across multiple dynamic MapReduce clusters. SIGMETRICS 2014.
8. Iosup and Epema: Grid Computing Workloads. IEEE Internet Computing 2011.
9. Iosup et al.: On the Performance Variability of Production Cloud Services. CCGRID 2011.
10. Iosup et al.: Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing. IEEE TPDS 2011.

Contact Me or Our Team



Collaboration or discussion about Massivizing Computer Systems:
Understanding, designing, deploying, tuning, analyzing, benchmarking distributed systems and ecosystems, including cloud computing and big data systems. Other topics in large-scale distributed systems and performance engineering are welcome.

A.Iosup@vu.nl

+31-20 59 89468 (Amsterdam)

@AIosup

https://atlarge-research.com/aiosup/

https://www.linkedin.com/in/aiosup

VU University, Faculty FEW/building W&N, Room P4.14
De Boelelaan 1081, 1081HV Amsterdam,
The Netherlands

VU VRIJE UNIVERSITEIT AMSTERDAM