

Massivizing Computer Systems = Making Computer Systems Scalable, Reliable, High-Performance, etc., Yet Form an Efficient Ecosystem



This Is the Golden Age of Computer Systems



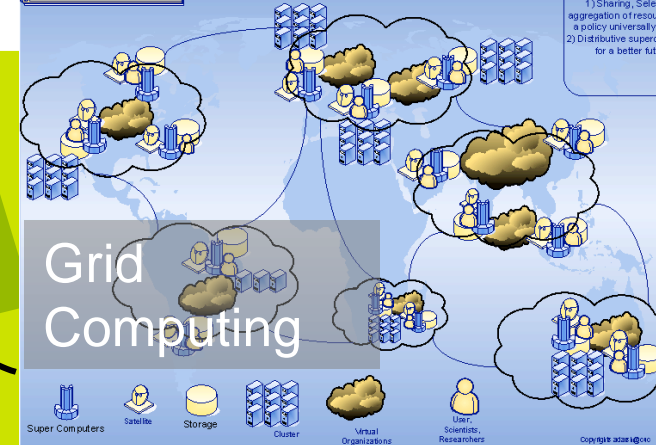
Education for
Everyone (Online)



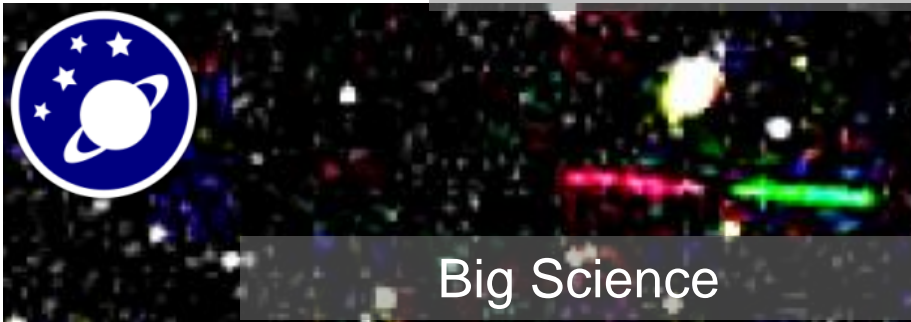
Business
Services



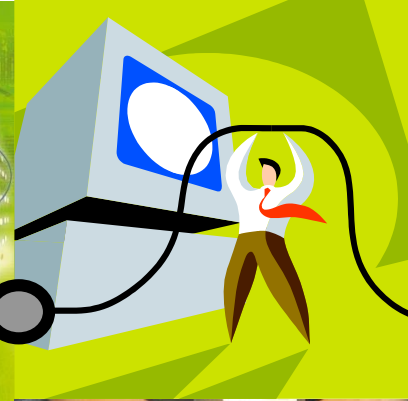
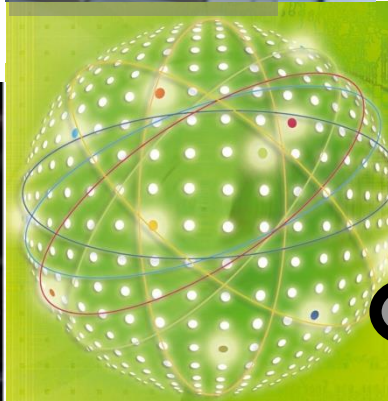
Grid Computing



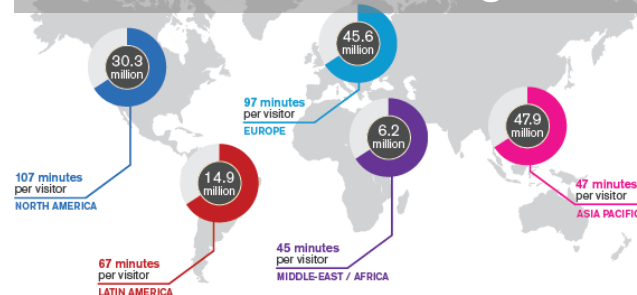
Grid
Computing



Big Science



Online Gaming



AVERAGE DAILY ONLINE GAMERS WORLDWIDE

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



BIG
DATA



Datacenters



Daily Life



This Is the Golden Age of Computer Systems

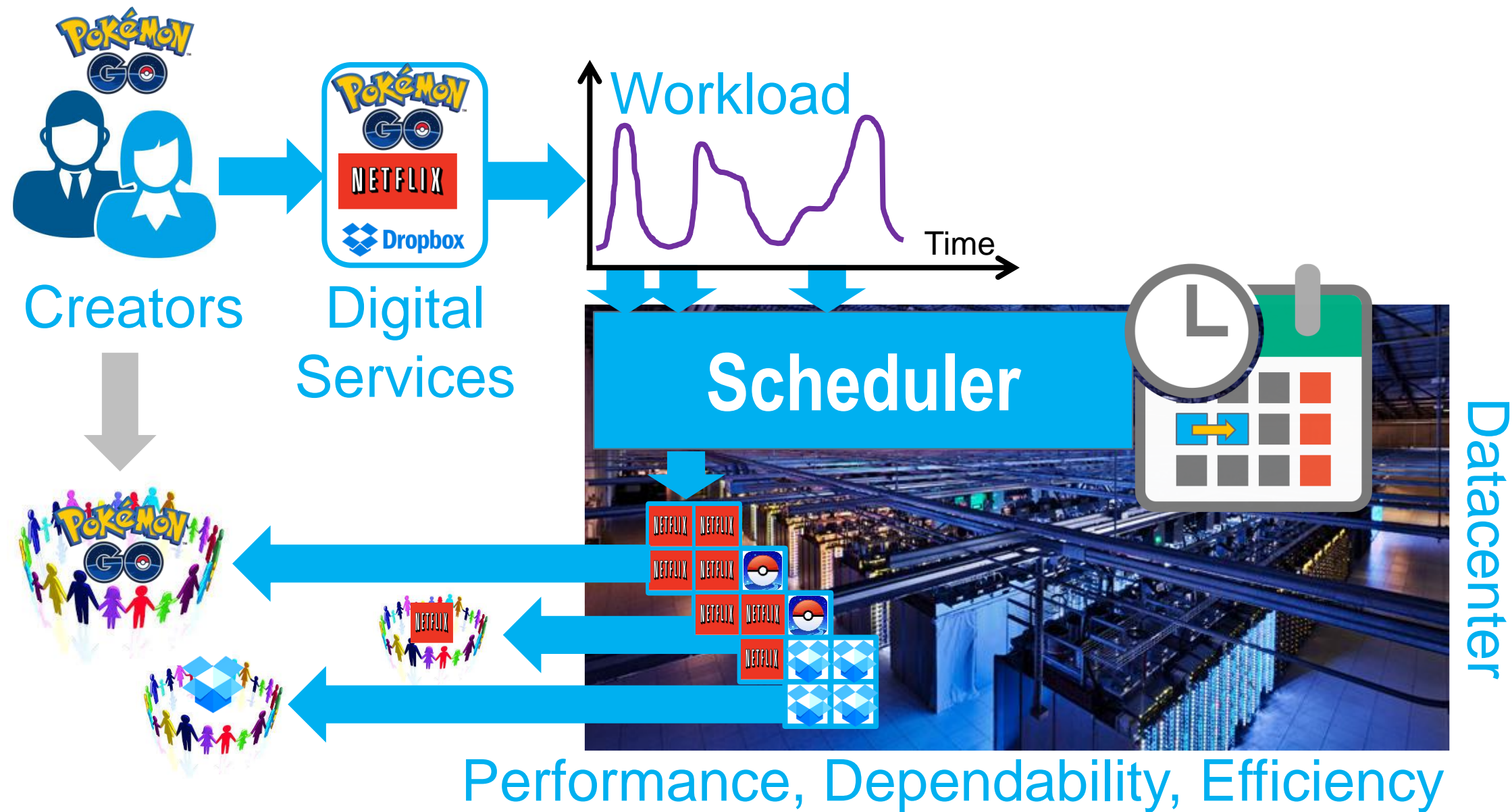
Do you recognize this App?



Pokemon Go ~ 10% NL for 3 months

Here is how it operates...

Current Technology: Scheduler? Datacenter? Etc.



The Golden Age of Computer Systems

... Yet We Are in a Crisis

Is 56% uptime good? 66%? 96%?

Why does this* happen?

What to do about it*?

* In modern computer systems, several or all issues may be linked. Thus, looking at any single issue in isolation is no longer sufficient.

This Is the Golden Age of Computer Systems and We Have Many Tools... Yet We Are in a Crisis

Need to Help Real Users
Choose Their Tools

Need Smarter Schedulers

Need Dependable Systems

Need to Address
“Data Deluge”,
“Ecosystem Navi”, etc.

Need to Be Much More
Efficient, But Also Ethical

... but the Current Laws and Theories Were
Built For Isolated Computer Systems

Need to Understand Operational Laws
when Massivizing Computer Systems

Need to Create Theories on how to
Massivize Computer Systems
while Ensuring Wanted Properties

Need to Build, to Massivize Computer
Systems with Wanted Properties

This Is the Golden Age of Computer Systems ... Yet We Are in a Crisis

Massivizing Computer Systems
Tackles All These Challenges...

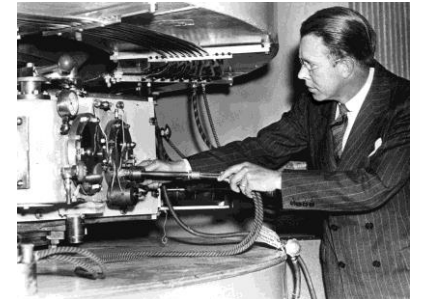
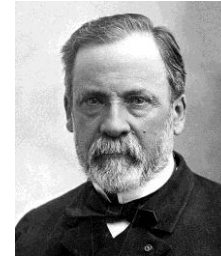
... and Is Relevant, Impactful, and I Believe Inspiring
for Many Young Scientists

My Story From Now On...

Massivizing Computer Systems

In Pasteur's Quadrant+:

- Fundamental research
 - Inspired by real use
 - Experimental in nature
- ~ Big Science as management



Fundamental Research in Massivizing Comp. Sys.

Scheduling

Bags-Of-Tasks

Workflows !

Portfolio !

Dependability

Failure Analysis*

Space-/Time-Correlation

Availability-On-Demand

New World+

Workload Modeling

Business-Critical

Online Gaming

Ecosystem Navigator+ Scalability/Elasticity+

Performance Variability

Grid*, Cloud, Big Data

Benchmarking

Longitudinal Studies

Delegated Matchmaking*

POGGI*, AoS

BTWorld*

Auto-Scalers

Heterogeneous Systems

Socially Aware+

Collaborative Downloads*

Groups in Online Gaming

Toxicity Detection*

Interaction Graphs !

Software Artifacts !

Graphalytics, etc.

Data Artifacts !

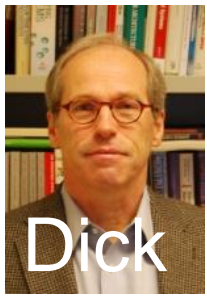
A Distributed Systems Memex*

Fundamental Problems/Research Lines

+ Please ask for a definition

My Contribution So Far ! VIDI-funded

* Award-winning (level of ambition)



Dick



Henk



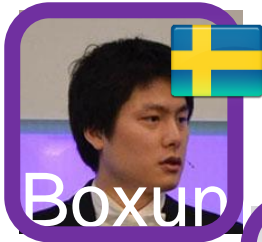
Johan



Ferna

Many Thanks to Many!

(My @large Team at VU and TUD)



Boxun



Yon



Elvan



Nezh



Vincent



Boddar



Otto



Mihai



Tim



Sai



Jie



Orna



Adele



StefanH



Corina



AlexO



Jesse



Wing



Take-Home Message: I Am Massivizing Computer Systems!

- The Golden Age of Computer Systems →
- My Research is about Massivizing Computer Systems →
 - Research approach: Pasteur's Quadrant + Big Science →
 - Fundamental research lines →
- **General Questions** ← we are here now →
- Ask for an Example: Portfolio Scheduling ← suggestion



Education for
Everyone



Business
Services



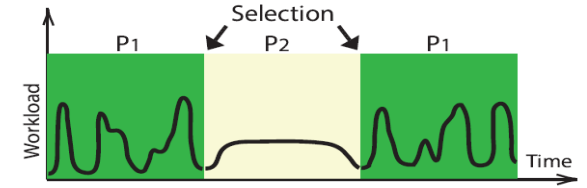
Massivizing
Computer
Systems



Big Science

Consider Reading the Following:

1. Iosup et al. LDBC Graphalytics: A Benchmark for Large-Scale Graph Analysis on Parallel and Distributed Platforms. PVLDB 9(13): 1317-1328 (2016)
2. Guo et al.: Design and Experimental Evaluation of Distributed Heterogeneous Graph-Processing Systems. CCGrid 2016: 203-212
3. van Beek et al.: Self-Expressive Management of Business-Critical Workloads in Virtualized Datacenters. IEEE Computer 48(7): 46-54 (2015)
4. Jia et al.: Socializing by Gaming: Revealing Social Relationships in Multiplayer Online Games. TKDD 10(2): 11 (2015)
5. Ghit et al.: V for Vicissitude: The Challenge of Scaling Complex Big Data Workflows. CCGRID 2014: 927-932
6. Guo et al.: How Well Do Graph-Processing Platforms Perform? An Empirical Performance Evaluation and Analysis. IPDPS 2014: 395-404
7. Javadi et al.: The Failure Trace Archive: Enabling the comparison of failure measurements and models of distributed systems. J. Parallel Distrib. Comput. 73(8): 1208-1223 (2013)
8. Iosup and Epema: Grid Computing Workloads. IEEE Internet Computing 15(2): 19-26 (2011)
9. Iosup et al.: On the Performance Variability of Production Cloud Services. CCGRID 2011: 104-113
10. Iosup et al.: Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing. IEEE Trans. Parallel Distrib. Syst. 22(6): 931-945 (2011)

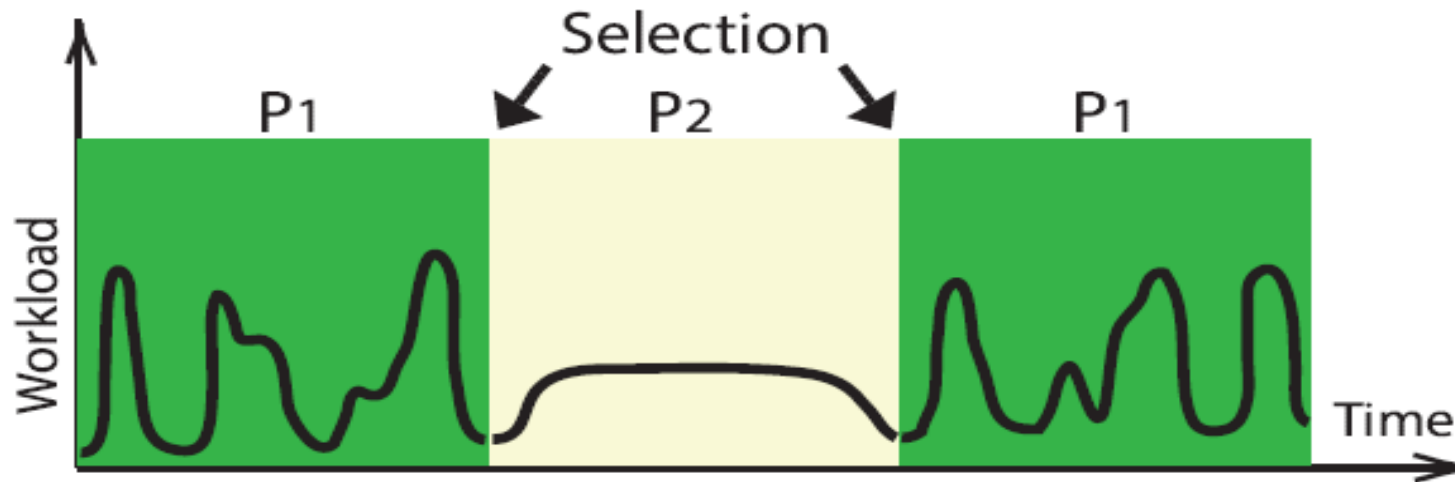


An Example: Portfolio Scheduling for Datacenters (what's in a name)



Portfolio Scheduling, In A Nutshell

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



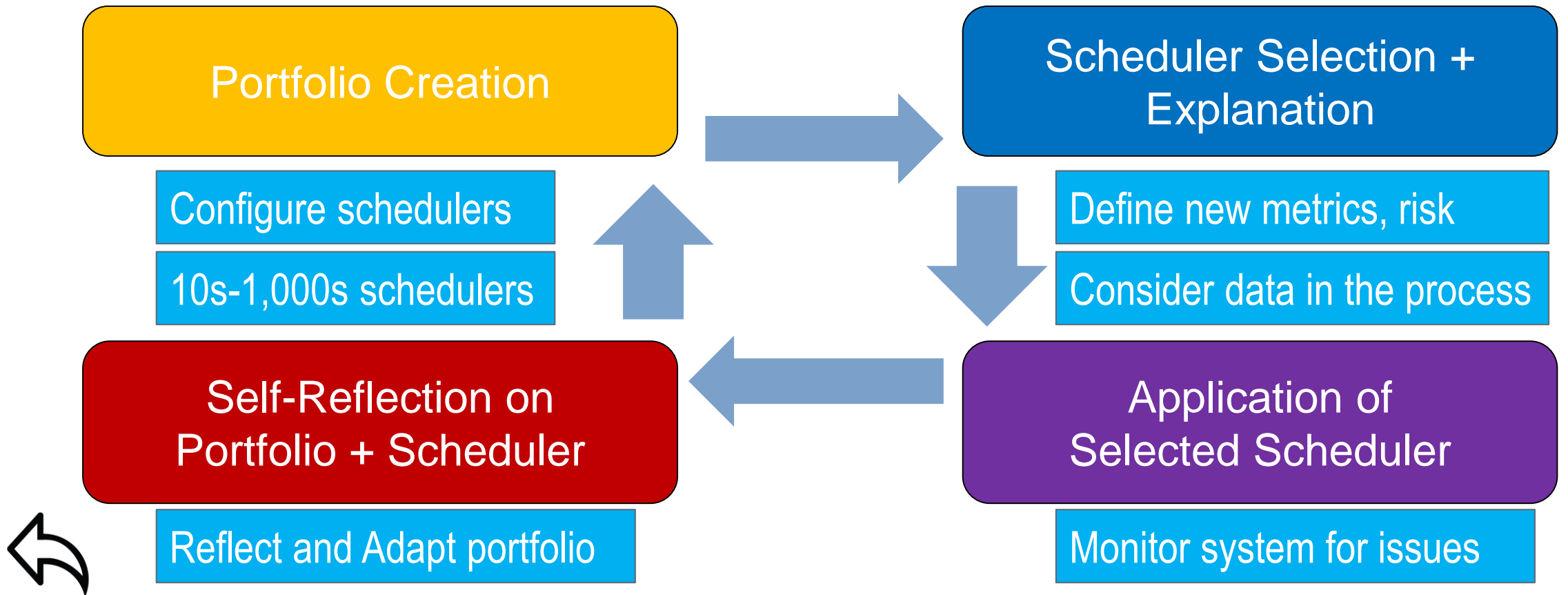
1. Create a set of schedulers
 - Resource provisioning and allocation policies for datacenters
2. Select active scheduler online, apply for the next period, analyze results
(Repeat)





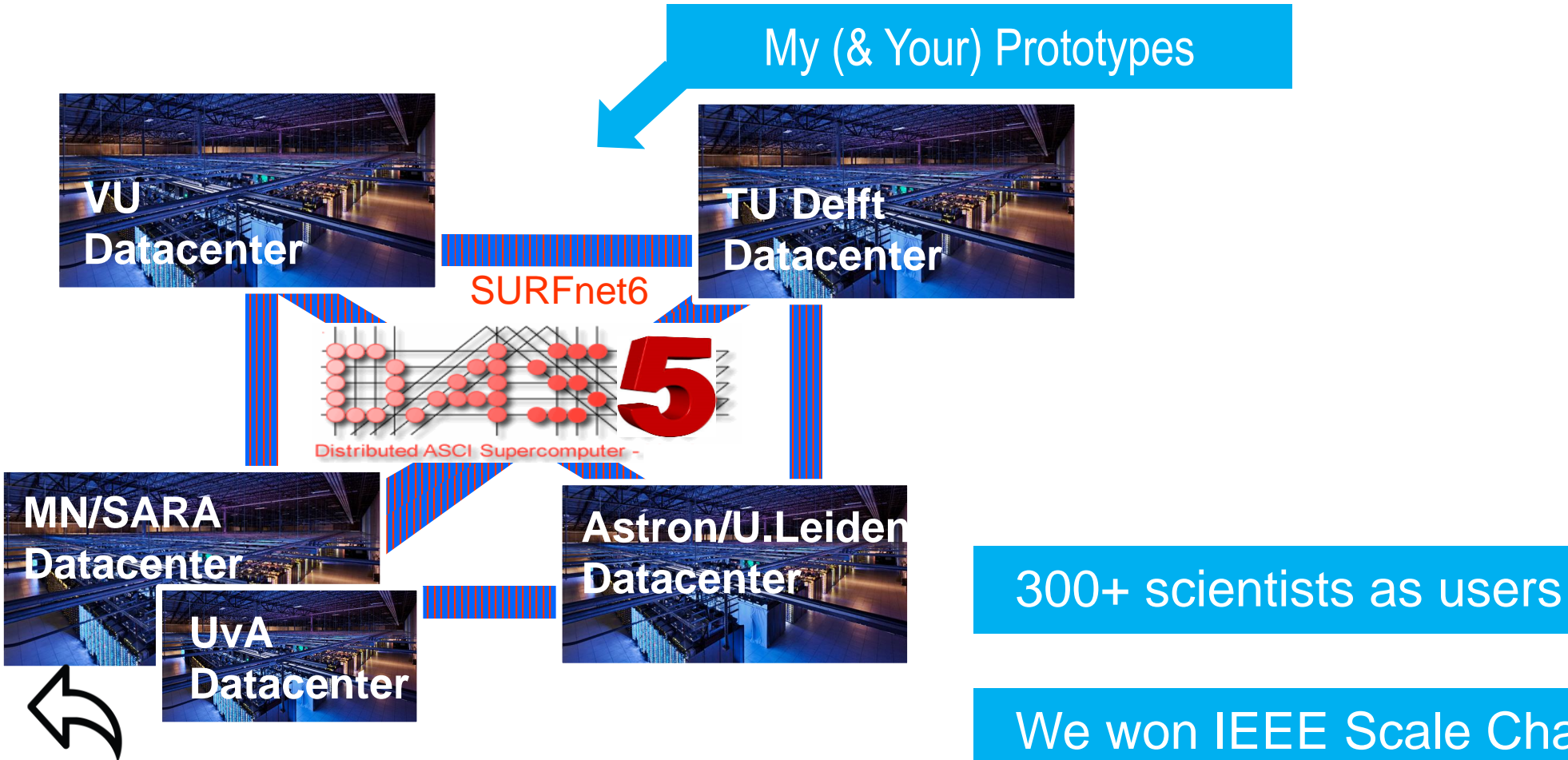
Portfolio Scheduling for Computer Systems

Portfolio Scheduling

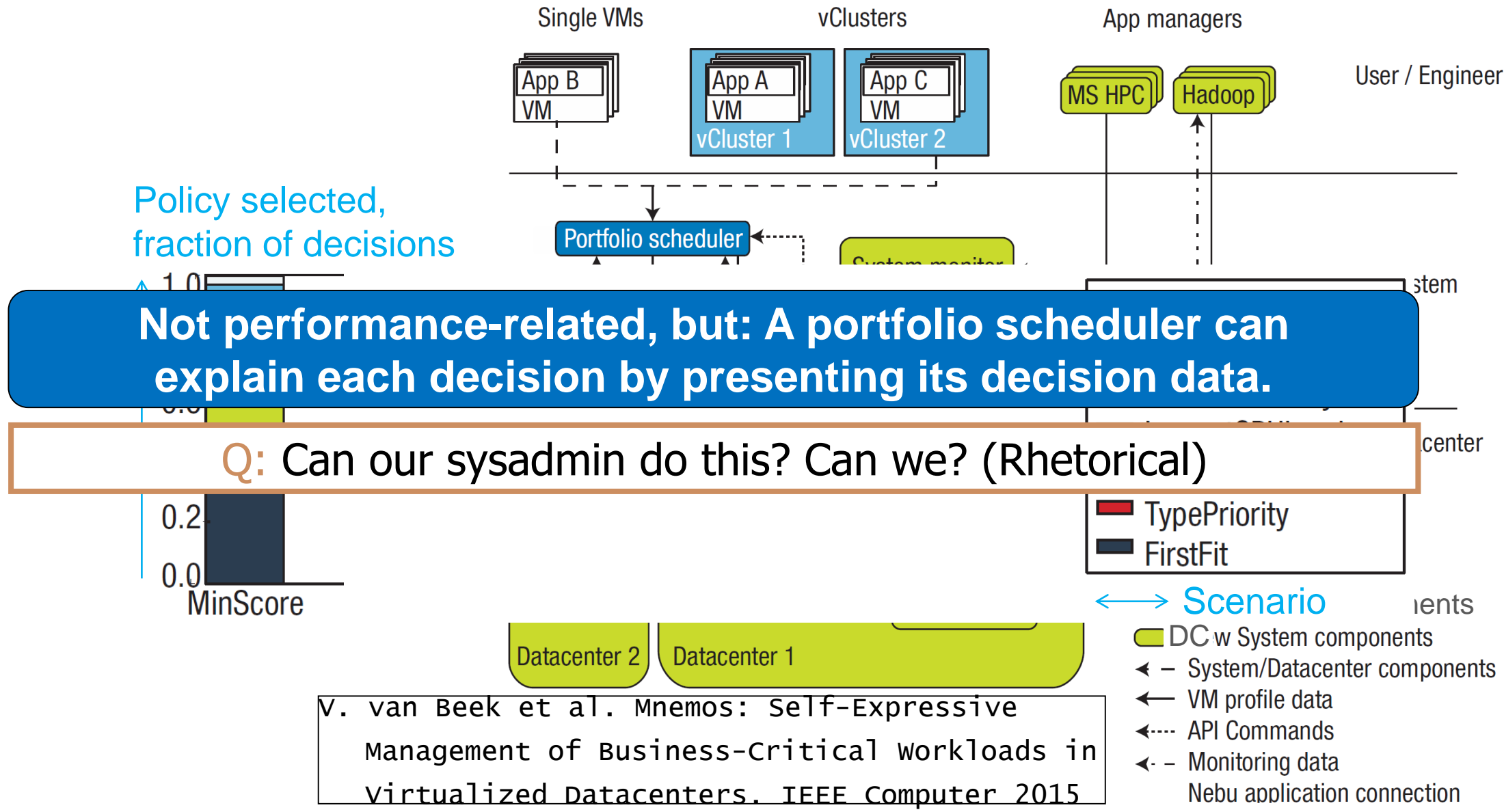


Experimental Research Methodology

My Main Scientific Instrument: DAS-5



Portfolio Scheduling in Practice: Massive Datacenters



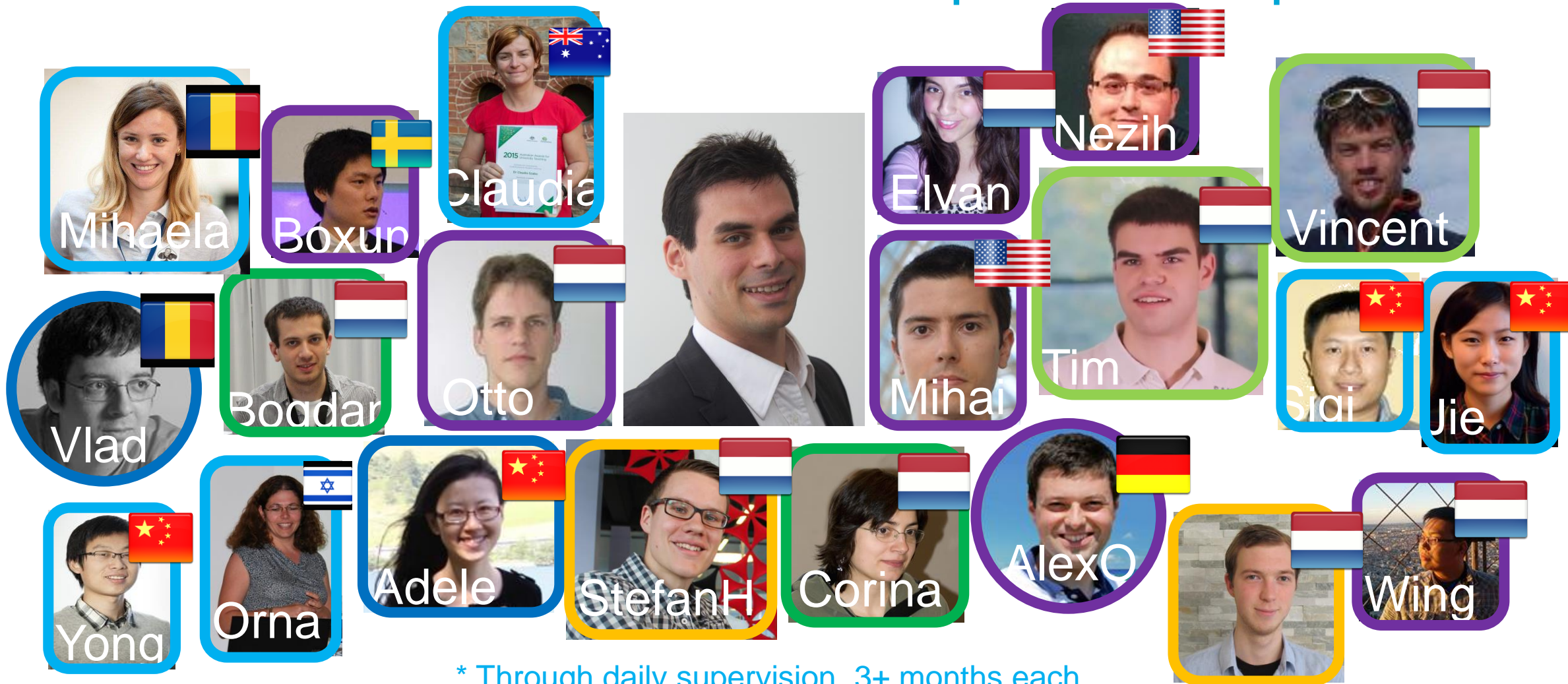
End of Example:
Portfolio Scheduling for Datacenters
(what's in a name)





Supporting Emerging Scholars: For A New Generation of Top-Quality, Socially Responsible Professionals

The New Generation I Helped Develop*



* Through daily supervision, 3+ months each

2

Ph.D. student

2

Post-doc

6

Assistant Prof.

2

AssociateProf.

7

Scientist

2

Teacher

Honors Track at TU Delft

Inventing things: VLDB'16

Teaching others: IEEE/ACM SC'15 Tutorial



Competition: IEEE Scale Challenge Winners 2014

New knowledge: IEEE Big Data 2013





Societal Impact: Ethical Innovation By and For Many

Valorization: Innovation By All For All

Comp.sci. for comp.sci.+
Students (M.Sc.,Ph.D.) +
Public lectures and info +
Public Software/Data

+

Collaborators

+

Application Domains



Big Science



Personalized &
Online Education



Business Services



Online Gaming



Ethical Issues To Warn About: Jevons' Effect Is Here

Over 500 YouTube videos have at least 100,000,000 viewers each.

If you want to help kill the planet:

https://www.youtube.com/playlist?list=PLirAqAtl_h2r5g8xGajEwdXd3x1s

**Need To Be Much More
Efficient, But Also To
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 (Oct 2016).