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)

Big Science

Business Services

Grid Computing

Online Gaming

Datacenters

Daily Life

Source: comScore MMO, Worldwide, April 2011, Age 18+
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…

Creators → Digital Services → Workload

Scheduler

Datacenter

Performance, Dependability, Efficiency
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

… 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

+ Please ask for an example
**Fundamental Research in Massivizing Comp. Sys.**

<table>
<thead>
<tr>
<th>Scheduling</th>
<th>Dependability</th>
<th>New World+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bags-Of-Tasks Workflows ! Portfolio !</td>
<td>Failure Analysis* Space-/Time-Correlation Availability-On-Demand</td>
<td>Workload Modeling Business-Critical Online Gaming</td>
</tr>
<tr>
<td>Performance Variability Grid*, Cloud, Big Data Benchmarking Longitudinal Studies</td>
<td>Heterogeneous Systems</td>
<td></td>
</tr>
</tbody>
</table>

**Software Artifacts !**
Graphalytics, etc.

**Data Artifacts !**
A Distributed Systems Memex* + Please ask for a definition

* Award-winning (level of ambition)

**Fundamental Problems/Research Lines**
My Contribution So Far ! VIDI-funded
Many Thanks to Many!
(My @large Team at VU and TUD)

Dick
Henk
Johan
Ferna
Boxun
Yon
Boodar
Otto
Orna
Adele
StefanH
Corina
AlexO
Jie
StefanH
Wing
Jesse
Elvan
Mihai
Tim
Sai
Vincent
Nezih
Adele
Ph.D. student
Post-doc
Assistant Prof.
AssociateProf.
Scientist
Teacher
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
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)
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 Creation
- Configure schedulers
- 10s-1,000s schedulers

Self-Reflection on Portfolio + Scheduler
- Reflect and Adapt portfolio

Scheduler Selection + Explanation
- Define new metrics, risk
- Consider data in the process

Application of Selected Scheduler
- Monitor system for issues
Experimental Research Methodology
My Main Scientific Instrument: DAS-5

300+ scientists as users
We won IEEE Scale Challenge 2014
Portfolio Scheduling in Practice: Massive Datacenters

Policy selected, fraction of decisions

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)

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 Associate Prof.  7 Scientist  2 Teacher
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

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, ...

Note: Psy has >3.5 billion views (Oct 2016).