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
- Big Science
- Online Gaming
- Grid Computing
- Datacenters

*Average daily online gamers worldwide*
This Is the Golden Age of Computer Systems

Do you recognize this App?

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*, as operator/customer?
What is efficient? What is the future of 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
… Yet We Are in a Crisis

Need to Help Real Users Choose Their Tools

Need Smarter Schedulers

Need Dependable Systems

Need to Address The “Data Deluge”

Need to Be Much More Efficient, But Also to Educate Our Customers

... 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 These Challenges...

... and Is Relevant, Impactful, and I Believe Inspiring for Many Young Scientists
My Story From Now On…
Massivizing Computer Systems
(Problem Found)

In Pasteur’s Quadrant+
- Fundamental research
- Inspired by real use
- Experimental in nature

+ Please ask for an example
Fundamental Problems/Research Lines

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*

My Contribution So Far! VIDI-funded

+ Please ask for a definition
* Award-winning (level of ambition)
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
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)
The Golden Age of Computer Systems

My Research is about Massivizing Computer Systems

- Research approach: Pasteur’s Quadrant
- Fundamental research lines

An Example: Portfolio Scheduling

General Questions ← we are here now
Many Thanks to Many!
(My @large Team at TU Delft)
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)