



HAL
open science

E2Clab: Reproducible Analysis of Complex Workflows on the Edge-to-Cloud Continuum

Daniel Rosendo, Alexandru Costan, Gabriel Antoniu, Patrick Valduriez

► **To cite this version:**

Daniel Rosendo, Alexandru Costan, Gabriel Antoniu, Patrick Valduriez. E2Clab: Reproducible Analysis of Complex Workflows on the Edge-to-Cloud Continuum. IPDPS 2021 - 35th IEEE International Parallel and Distributed Processing Symposium, May 2021, Virtual, France. . hal-03269852v2

HAL Id: hal-03269852

<https://hal.science/hal-03269852v2>

Submitted on 7 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

E2Clab: Reproducible Analysis of Complex Workflows on the Edge-to-Cloud Continuum

Daniel Rosendo

University of Rennes
Inria, CNRS, IRISA - Rennes, France
daniel.rosendo@inria.fr

Advisors:

Gabriel Antoniu

University of Rennes
Inria, CNRS, IRISA
Rennes, France

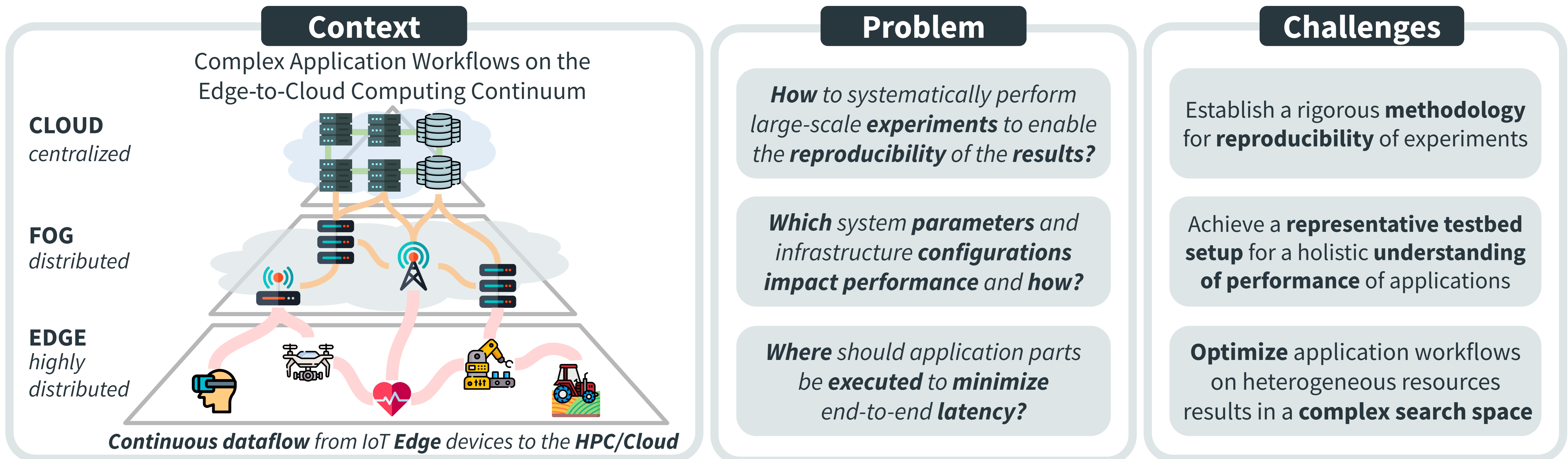
Alexandru Costan

University of Rennes
Inria, CNRS, IRISA
Rennes, France

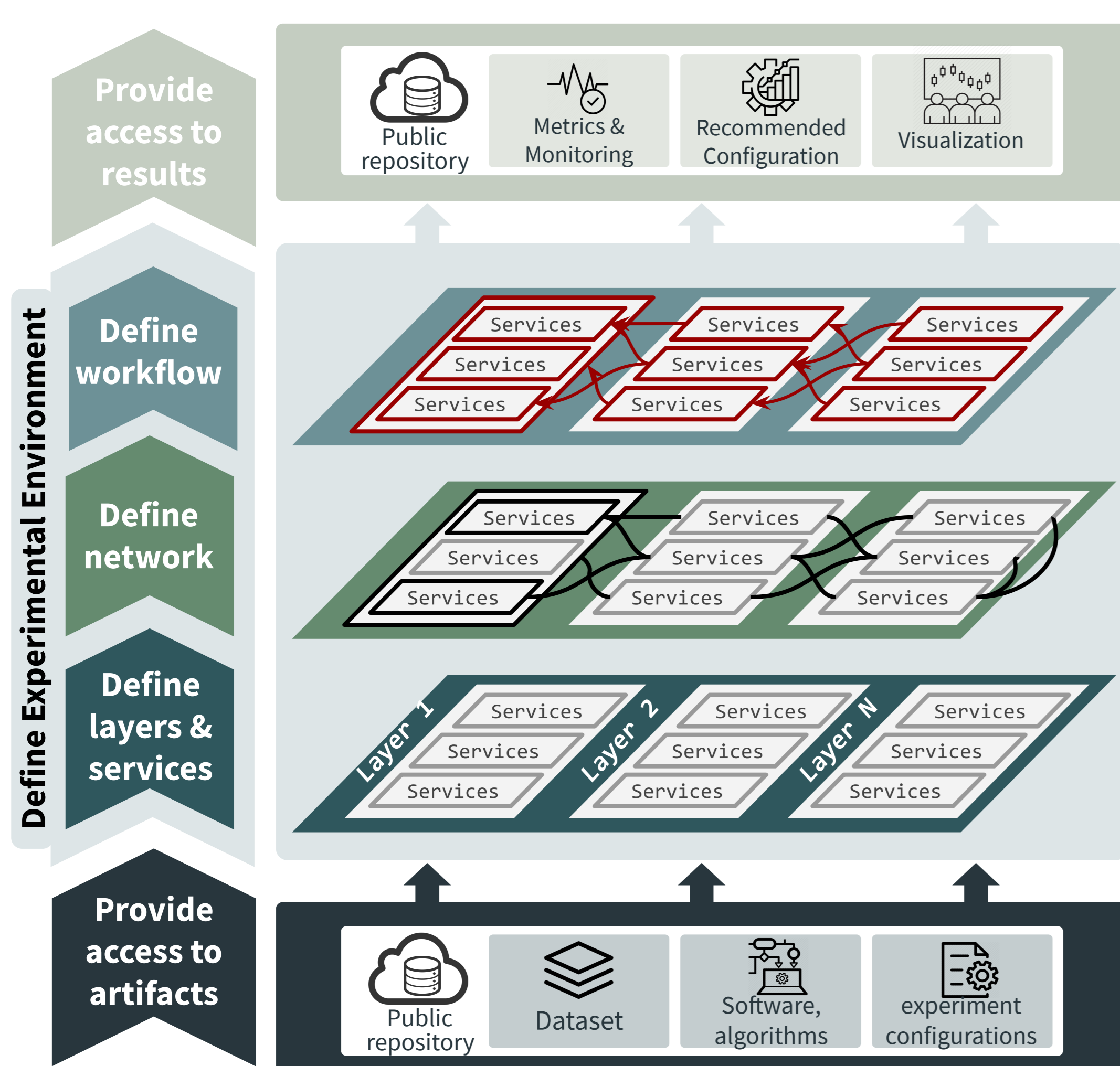
Patrick Valduriez

University of Montpellier
Inria, CNRS, LIRMM
Montpellier, France

Introduction

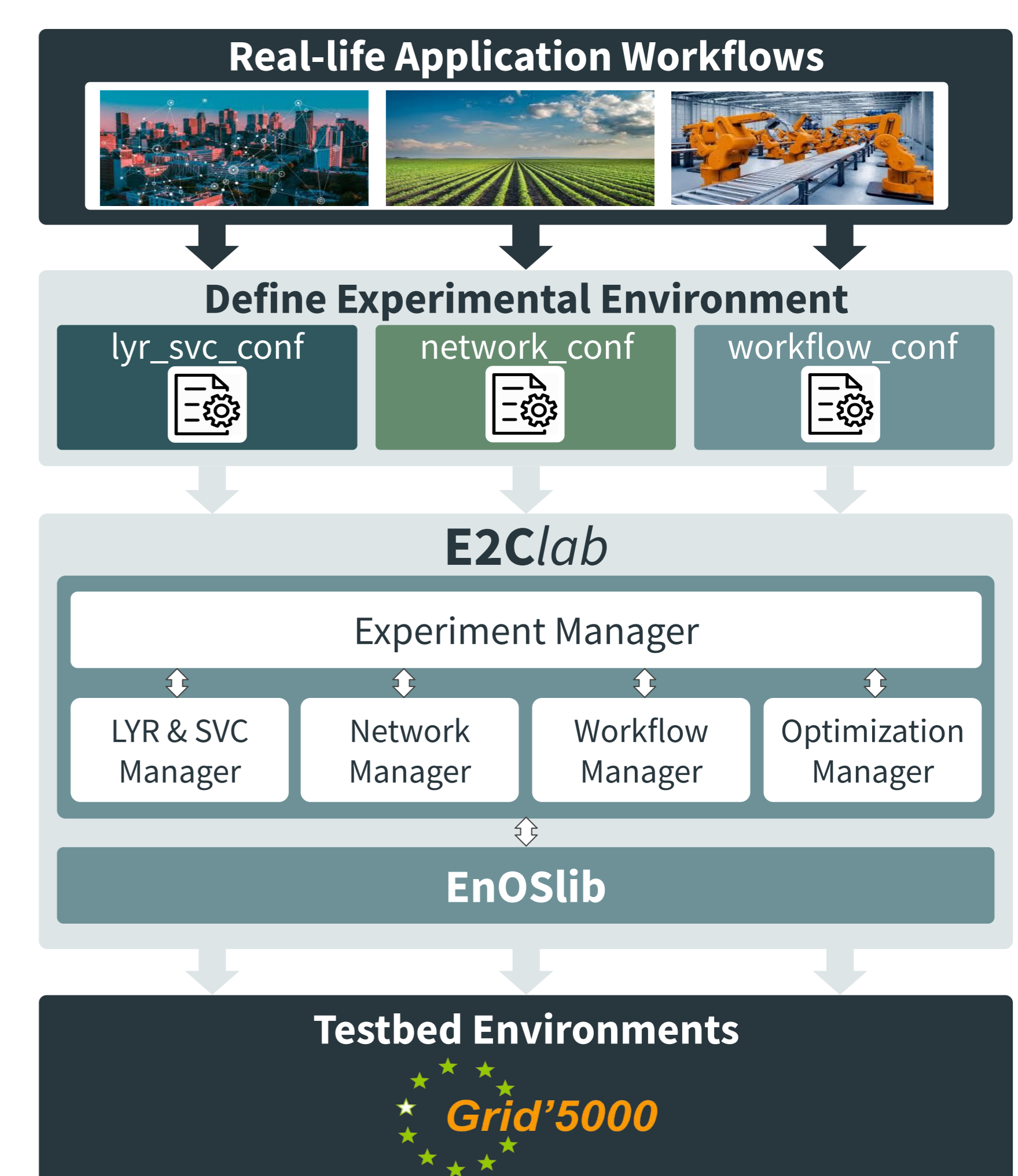


Our Contribution

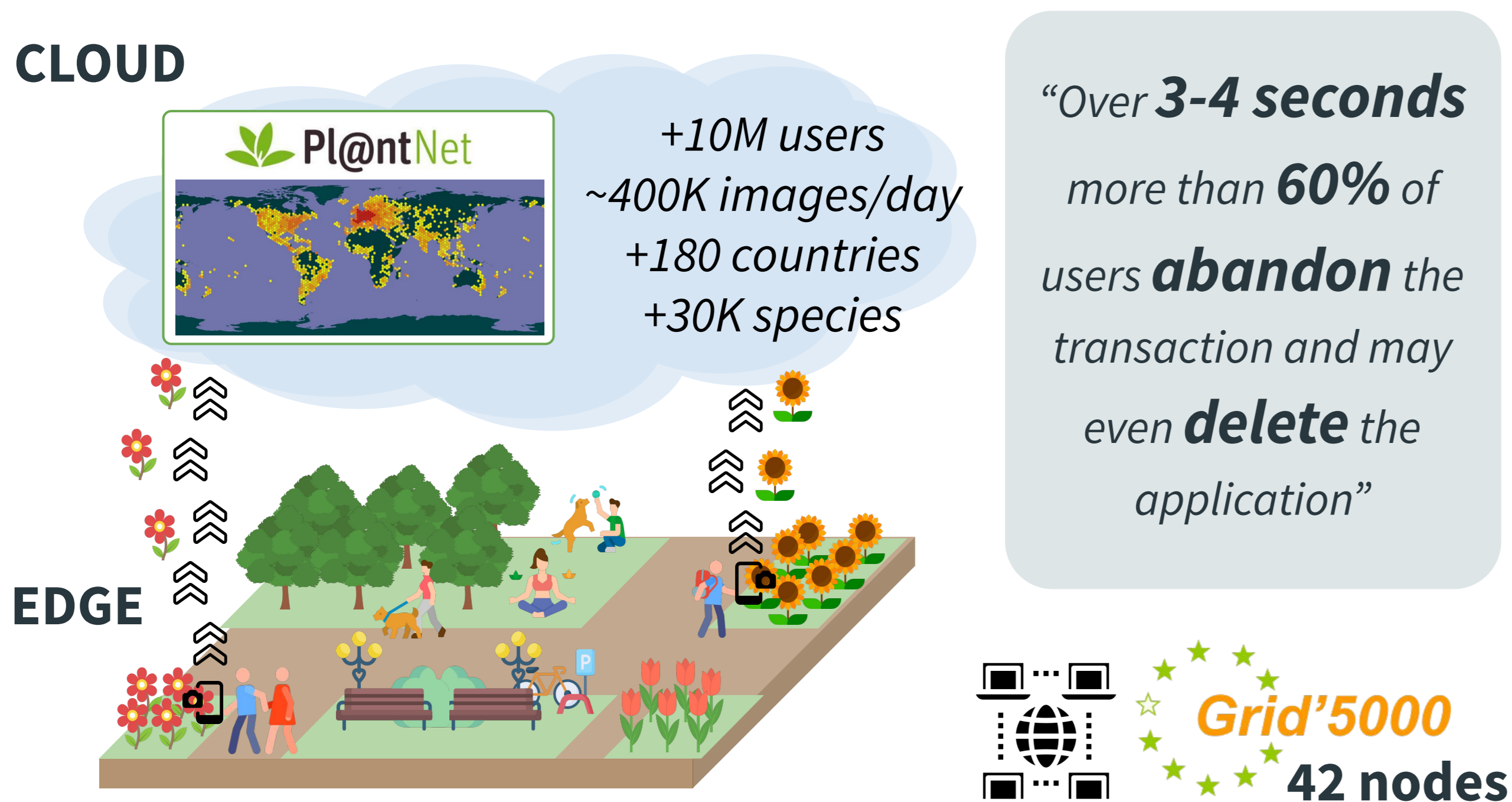


Methodology

- ★ **Reproducible Experiments**
Repeatability, Replicability & Reproducibility
- ★ **Mapping**
Application parts & physical testbed
- ★ **Variation & Scaling**
Experiment variation and transparent scaling
- ★ **Network Emulation**
Edge-to-Cloud communication constraints
- ★ **Application Optimization**
Parameter search of application and infrastructure configurations



Validation with a Large-scale Real-life Application: Pl@ntNet

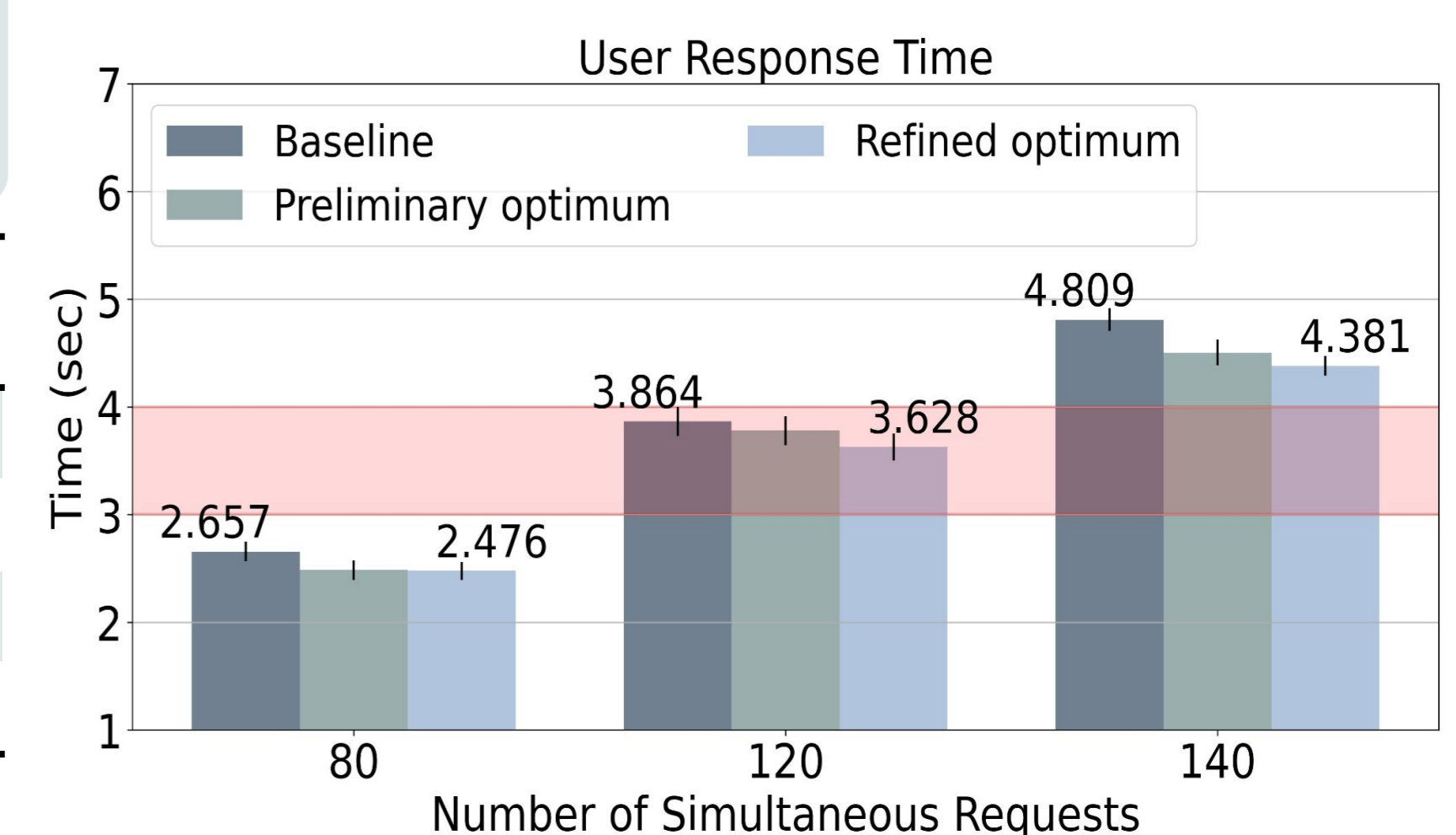


What is the software configuration that minimizes the user response time?

Find (http, download, simsearch, extract), in order to Minimize UserResponseTime
Subject to $20 \leq (\text{http, download, simsearch}) \leq 60$, Pool Size. $3 \leq (\text{extract}) \leq 9$, Pool Size.

Thread pool	baseline	preliminary optimum	refined optimum
HTTP	40	54	54
Download	40	54	54
Extract	7	7	6
Simsearch	40	53	53
User resp. time (sec)	2.65	2.48	2.47

How does the number of simultaneous users accessing the application impact on the user response time?



Takeaways

- Our methodology has proved useful for **understanding** and **improving** the **performance** of a real life application used at very large-scale.
- The configuration found by E2Clab can **process 35% more** requests and presents a **smaller user response time** compared to the baseline.

Perspectives

- E2Clab enables the Computing Continuum vision by allowing **reproducible** experimental **research** on **large-scale testbeds**.
- E2Clab supports the complete experimental cycle across the Computing Continuum: **deployment, analysis, optimization**.