

## **Fact Sheet Achievements Project P20 of COMMIT/NL**

### *Infrastructure Virtualization for e-Science*

Between 2001 and 2017, Project P20 of COMMIT/NL, the Dutch National research program in Computer Science, was executed, with as main participating universities/researchers Delft University of Technology (Dick Epema and Alexandru Iosup), the University of Amsterdam (Cees de Laat and Paola Grosso), and the Vrije University Amsterdam (Henri Bal and Thilo Kielmann).

In line with its original mission, the P20 consortium made major scientific contributions on programming and managing modern computer infrastructures. This work was indispensable for many data-intensive applications that need to scale to real-world data sizes. This essential technological foundation for big data processing also received wide international recognition and attracted important new grants and valorization activities.

This fact sheet records the main achievements of the project.

#### **Awards**

- Dutch Prize for ICT Research 2016
- Teacher of the Year of the Netherlands 2015 and Best Lecturer of TU Delft 2013-2014
- Achievement Award EuroPar 2014
- IEEE TCSC Scale Challenge 2014
- Best Paper Award DEBS 2014
- Best Paper Award MASCOTS 2013
- Best Paper Award Cloud Computing 2013
- Best Impact Award SC/MTAGS 2012
- Best Paper Award SC/MTAGS 2012
- SPECTacular Award from SPEC 2012, 2013, and 2017

#### **Grants**

- NWO/STW VIDI
- NWO/M DAS-5
- NWO Cybersecurity
- NWO Veni
- NWO/STW Veni
- Six H2020 projects on environmental data, intercloud, virtual lab, rt-cloud bench and exascale reconfigurable technologies
- Industry grants and gifts for promising and directly transferrable research

## **Recognition:**

- Member of De Jonge Akademie of the KNAW
- Berkeley LBNL Policy board member
- Berkeley LBNL Computer Science division research assessment committee member
- Membership Academia Europaea
- Keynote NCSA, Urbana Champaign 2015
- Keynote HPCS 2015
- Keynote ANSP BMA6 2014
- Keynote EuroPar 2014
- Keynote CCGrid 2012
- Keynote CineGrid 2012

**PhD theses completed:** 11

## **Valorization activities**

- Addition of KLM to the project
- Spin Off company Guido van 't Noordende
- Close cooperation with KPMG on big data for business and society
- Joint work with Bitbrains, one of the largest Dutch cloud computing providers
- COMMIssioner valorization grant from COMMIT (2014-15)
- Two NWO KIEM projects
- Chair of the Cloud Working Group of SPEC
- Ongoing collaboration and published work with Intel Research Labs, OR; Oracle Research Labs, CA; Google Inc., CA; IBM TJ Watson, NY, etc.

## **Dissemination activities**

- NOS journal opening item (10 Sept 2015) about electronic patient files
- Advised the Dutch parliament about electronic patient files
- Much publicity on sports analytics for ice skating research
- Coverage in several national newspapers, TV, and radio

## **Top publications**

- HPDC 2017: resilience in in-memory data-processing systems
- ICPE 2017: comprehensive experimental study of cloud auto-scalers (nominated for BPA)
- PVLDB 2016: industry-supported process for benchmarking graph-processing platforms
- IEEE Computer 2016: strategy of and achievements with the DAS systems
- IEEE Computer 2015: datacenter management jointly with industry (Bitbrains)
- ICDCS 2015: Scalable, Instant VM Startup for IaaS Clouds
- HPDC 2014: storing VM images in clouds
- SIGMETRICS 2014: big data processing with MapReduce
- ICDCS 2014: distributed middleware for MapReduce and stream processing
- IPDPS 2014: multi-dimensional experimental study of distributed graph-processing platforms

- IEEE TPDS 2014: Parallel Workload Modeling with Realistic Characteristics
- ISWC 2013: distributed reasoning on dynamic semantic web data
- SC 2013: VM deployment
- SC 2013: portfolio scheduling
- IEEE Big Data 2013: first Dutch article in a top big data conference
- IEEE Internet Computing 2013: Distributed Computing on an Ensemble of Browsers
- IEEE Internet Computing 2012: Enabling Web Services to Consume and Produce Large Datasets
- IEEE TPDS 2012: Cost-driven Scheduling of Grid Workflows Using Partial Critical Paths
- IEEE Internet Computing 2011: Grid Computing Workloads
- ISWC 2011: an OWL reasoner that scales to 1 billion triples
- IEEE TPDS 2011: performance analysis of cloud services, highest cited article in the highest impact journal in the field, for the period 2009-2014 (source: Google Scholar)
- HPDC 2011: Incremental Placement of Interactive Perception Applications

### **Community service**

- Area Chair SC 2016
- General Chair ACM/SPEC ICPE 2016
- Organizer PIRE 2013-2014: Open Science Data Cloud symposium
- Program chair RDA-P4 2014: Research Data Alliance plenary
- Program Chair HPCS 2015
- General Chair CCGRID 2013
- PC Co-Chair HPDC 2013, and 2017
- General Chair HPDC 2012 and 2015
- Team Chair SuperComputing / SCinet

### **Computing infrastructure and artifacts**

- DAS-5, main research infrastructure for big data processing and cloud computing in the Netherlands (NWO-M)
- Ibis/JavaGAT (deployed by NLeSC)
- Graphalytics, open-source benchmarking software for big data systems (used by the LDBC graph-processing community, aim to become industry standard through SPEC)
- The Grid Workloads Archive, open-access datasets (275 citations)
- The Failure Trace Archive, open-access datasets (over 200 citations)
- The Game Trace Archive, open-access datasets (de facto standard in the community)
- First European GENI node for virtualized networked environments.

### **Bibliometrics in *past 5 years*** (source: public personal pages on Google Scholar, august 2017)

- Henri Bal, h-index 27 and over 3,000 citations
- Dick Epema, h-index 32 and over 5,250 citations
- Paola Grosso, h-index 16 and over 1,800 citations
- Alexandru Iosup, h-index 30 and over 4,100 citations
- Thilo Kielmann, h-index 22 and over 1,800 citations
- Cees de Laat, h-index 24 and over 2,300 citations