

# Vasiliki (Vasia) Kalavri

Assistant Professor, Boston University

vkalavri@bu.edu, (+1) 617 353 8919

## Academic Appointments

**January 2020 - now, Boston University, MA**

**Tenure-Track Assistant Professor**

Department of Computer Science, College of Arts & Sciences

**July 2019 - December 2019, Boston University, MA**

**Adjunct Assistant Professor**

Department of Computer Science, College of Arts & Sciences

**January 2017 - October 2019, ETH Zürich, Switzerland**

**Postdoctoral Fellow**

Systems Group, Department of Computer Science

## Education

**2012 - 2016, KTH Royal Institute of Technology, Sweden, and  
Université Catholique de Louvain, Belgium**

**PhD in Information and Communication Technology**

Erasmus Mundus Joint Doctorate in Distributed Computing (double degree)

Thesis: *Performance Optimization Techniques and Tools for Distributed Graph Processing*

Advisors: Vladimir Vlassov (KTH, Sweden), Seif Haridi (SICS, Sweden), Peter Van Roy (Université Catholique de Louvain, Belgium)

Opponent: Peter Boncz (CWI and VU Amsterdam, Netherlands)

**2010 - 2012, KTH Royal Institute of Technology, Sweden, and  
Universitat Politècnica de Catalunya, Spain**

**MSc in Information and Communication Technology**

European Master in Distributed Computing (double degree)

Thesis: *Integrating Pig and Stratosphere*

Advisor: Per Brand (SICS, Sweden)

**2004 - 2010, National Technical University of Athens, Greece**

**MEng in Electrical and Computer Engineering**

Thesis: *Parallelization of Linear Algebra Algorithms for High-Scale Architectures*

Advisor: Nectarios Koziris (NTUA, Greece)

## Funding

**2021, RedHat Research Incubation Award**

Project title: Serverless Streaming Graph Analytics. *USD 85,000*. (PI).

**2021, RedHat Research Incubation Award**

Project title: Towards high performance and energy efficiency in open-source stream processing. *USD 150,000*. (Primary PI)

**2021, RedHat Research Incubation Award**

Project title: Creating a global open research platform to better understand social sustainability using data from a real-life smart village. *USD 150,000*. (Co-PI).

**2021, Hariri Institute Focused Research Program**

Project title: Continuous Analysis of Mobile Health Data among Medically Vulnerable Populations. *USD 150,000*. (Co-PI)

### **2021, Google DAPA (Data Acquisition, Processing and Analytics) Faculty Award**

Faculty award for research on self-managed streaming systems and workload-aware state management. *USD 50,000*. (PI).

### **2021, Samsung Memory Solutions Lab UR Collaboration Grant**

Project title: SSD-aware Context Serving for Streaming Graph Representation Learning. *USD 85,000*. (PI)

### **2020, Google DAPA (Data Acquisition, Processing and Analytics) Faculty Award**

Faculty award for research on self-managed streaming systems and workload-aware state management. *USD 50,000*. (PI)

## **Fellowships and Awards**

### **2020, USENIX HotStorage Outstanding New Research Direction Award.**

For the paper *In support of workload-aware streaming state management*.

### **2017, IBM Innovation Award**

In recognition of an outstanding PhD thesis that presents an original contribution to informatics or its applications, 5.000 EUR.

### **2017, ETH Zürich Postdoctoral Fellowship**

2-year fellowship for the research project *Automatic scaling of distributed streaming computations using graph analytics on real-time monitoring data*, 228.300 CHF.

### **2012, Erasmus Mundus Doctoral Fellowship**

3-year PhD fellowship awarded by the European Commission, 129.900 EUR.

### **2011, Movilidad de Estudiantes en Masteres Oficiales**

Grant for MSc Students awarded by the Ministry of Education in Spain, 1.380 EUR.

### **Travel Grants**

2018, Usenix NSDI diversity grant

2014, ACM SIGMOD travel grant

2013, Google Students Conference and travel grant

2013, Knut och Alice Wallenbergs travel grant

## **Service**

### **Boston University**

2021, CS PhD Advising Policies Committee.

2021, Dissertation Committee (Sohan Sinha).

2020-now, CS PhD admissions committee.

2020-now, Faculty Advisor for the ACM-Women student chapter.

### **Academic Conferences**

APSys 2022 (PC Member)

ACM SIGMOD 2022, 2023 (PC Member)

ACM SIGMOD Student Research Competition 2022 (co-Chair)

ACM SIGMOD 2022 Demos, Reproducibility (PC Member)

VLDB 2022 (Review board member)

USENIX ATC 2021 (Heavy PC Member)

ACM Middleware 2021 (PC Member)

ACM Middleware Doctoral Symposium 2020, 2021

IEEE ICDE 2021 (PC Member), 2022 (Associate Chair)

EuroSys 2021 (AMA Chair)

ACM DEBS 2020, 2021, 2022 (PC member)

ICDE 2020 (Demo PC)

EDBT 2020 (Demo PC)  
CCGrid 2019 (Applications and Data Science track co-Chair)  
OPODIS 2018 (PC member)  
ICAC 2018 (Sub-reviewer)

### Academic Workshops

Co-Chair of GRADES-NDA 2021, 2022 (co-located with SIGMOD 2021, 2022)  
USENIX HotStorage 2020  
GRADES-NDA 2020 (co-located with SIGMOD 2020)  
GRADES-NDA 2019 (co-located with SIGMOD 2019)  
DBTest 2018 (co-located with SIGMOD 2018)  
GRADES-NDA 2018 (co-located with SIGMOD 2018)  
GABB 2018 (co-located with IPDPS 2018)  
GABB 2017 (co-located with IPDPS 2017)  
DEEM 2017 (co-located with SIGMOD 2017)

### Academic Journals

Journal of Systems Research (Streaming Systems Area Chair)  
IEEE Transactions on Computers  
VLDB Journal  
IEEE Transactions on Knowledge and Data Engineering  
IEEE Transactions on Parallel and Distributed Systems

### Industrial Conferences

Flink Forward Europe 2019  
Flink Forward San Francisco 2017  
Berlin Buzzwords 2017  
Flink Forward Berlin 2016  
Berlin Buzzwords 2016

## Teaching Activities

### University Courses

#### **Streaming and Event-Based Systems**

(CS 551), Boston University, Spring 2022.

#### **Computer Systems (undergraduate)**

(CS 210), Boston University, Fall 2020, Fall 2021.

#### **Data Stream Processing and Analytics**

(CS 591 K1), Boston University, Spring 2020, Spring 2021.

#### **Data Stream Processing and Analytics**

(263-3826-00), ETH Zürich, Spring 2019.

#### **Implementation of Distributed Systems** (Semester project supervisor)

(ID2219), KTH Royal Institute of Technology, 2011 & 2015.

#### **Network programming with Java** (Guest lecturer)

(ID2212), KTH Royal Institute of Technology, 2015.

#### **Cloud Computing** (Guest lecturer)

Universitat Politècnica de Catalunya, Barcelona, 2015.

#### **Languages and Algorithms for Distributed Applications** (Head TA)

(LSINF2345), UCLouvain, 2014.

#### **Applied Programming** (Head TA)

(ID1218), KTH Royal Institute of Technology, 2011.

#### **Distributed Systems, Basic Course** (Head TA)

(ID220) KTH Royal Institute of Technology, 2011.

## Industry Courses

### **Distributed Systems and Algorithms with Erlang (TA)**

Seminar Course for Ericsson Engineers.

Kista (Sweden), Shanghai, Beijing (China), San Jose (USA), Budapest (Hungary), 2012.

### **Functional Programming with Haskell (TA)**

Seminar Course for Ericsson Engineers.

Shanghai, China, 2012.

## Summer Schools and Conference Tutorials

**Beyond Analytics: the Evolution of Stream Processing Systems**, ACM SIGMOD 2020.

3rd International **Summer School on Data Science**, Croatia, 2018.

**Big Data Analytics Summer School**, Stockholm, 2017 & 2018.

Tutorial at the 31st **British International Conference on Databases**, London, 2017.

Apache Flink tutorials at **BOSS workshop (VLDB)**, 2016 & 2017.

2nd Int'l ScaDS **Summer School on Big Data**, Germany, 2016.

**EIT Summer School on Cloud and Big Data**, Sweden, 2016.

Technical training at **Flink Forward**, Germany, 2015.

## Advising

### PhD students

- Zainab Abbas, KTH (Graduated, 2021. Co-advised with Vladimir Vlassov and Paris Carbone)
- Muhammad Faisal, BU
- Emmanouil Kritharakis, BU
- Yuanli Wang, BU

### Bachelor and Master students

- Baiqing Lyu, BU
- Jerry Zhang, BU

### Bachelor and Master Thesis Supervision

#### **FASTER State Management for Strymon.**

MSc Thesis, Matthew Brookes, 2019, ETH Zürich.

#### **Evaluation of Datasets and Methods for Online Automatic Detection of Web Trackers.**

BSc Thesis, Jean Kaufmann, 2018, ETH Zürich.

#### **Network Verification with Streaming Computation.**

MSc Thesis, Basile Maret, 2018, ETH Zürich.

#### **Automatic Detection of Web Trackers in Real-Time.**

MSc Thesis, Lu Chen, 2017, ETH Zürich.

#### **An Algebra and an Implementation of Cypher on Top of Apache Flink.**

MSc Thesis, Mengqi Yang, 2016, Technical University of Eindhoven.

Co-supervision with George Fletcher, Alex Poulouvasilis (Birkbeck), and Alex Averbuch (Neo Technology).

#### **Streaming Graph Partitioning for Apache Flink.**

MSc Thesis, Zainab Abbas, 2016, KTH Royal Institute of Technology.

Co-supervision with Paris Carbone.

#### **Optimization of Execution Plans for Fixpoint Iterative Algorithms in Large-Scale Graph Processing.**

MSc Thesis, Riccardo Diomedi, 2016, KTH Royal Institute of Technology.

#### **Scaling up Computations on Highly Skewed Graphs.**

MSc Thesis, Andra Lungu, 2015, TU Berlin.

Co-supervision with Asterios Katsifodimos.

### **Streaming Graph Analytics Framework Design.**

MSc Thesis, Daniel Bali, 2015, KTH. Royal Institute of Technology.

Co-supervision with Paris Carbone.

### **Result Reuse in Big-Data Processing Frameworks.**

MSc Thesis, Hui Shang, 2013, KTH Royal Institute of Technology.

### **Techniques and Applications of Early Approximate Results for Big-Data Analytics.**

MSc Thesis, Vaidas Brundza, 2013, KTH Royal Institute of Technology.

## **Events Organization**

### **Workshops**

February 2022: *Uncovering synergies between clinical mHealth use cases, streaming data processing, and machine learning approaches.* Boston University.

November 2021: *Secure analytics on mobile health data streams.* Boston University.

July 2021: *Development of Use Cases for Continuous Mobile Data in Health Promotion & Disease Detection.* Boston University.

### **Meetups**

2017 - present: Apache Flink Meetup, Zurich

2017 - present: Papers we Love Meetup, Zurich

2015 - 2016: Apache Flink Meetup, Stockholm

### **Developer Conferences**

2016 - present: HPC, Big Data, and Data Science, FOSDEM, Brussels

2017: RustFest Zurich

## **Industry Experience**

### **Feb 2016 - April 2016, Intern Software Engineer**

Data Artisans, Berlin, Germany

Design and implementation of Apache Flink's SQL and Stream SQL APIs

### **Sept 2014 - March 2015, Intern Software Engineer**

Telefonica Research, Barcelona, Spain

Large-Scale Graph Processing Optimizations and Automatic Detection of Web Trackers.

Advisors: Dionysios Logothetis (Facebook), Jeremy Blackburn (University of Alabama at Birmingham)

### **Feb 2012 - April 2012, Research Engineer**

Ericsson & KTH Royal Institute of Technology, Several locations

Organization and provision of intensive training seminars to Ericsson engineers on Distributed Systems and Functional Programming.

### **July 2009 - Aug 2010, Web Developer and Applications Designer**

e-on Integration S.A., Athens, Greece

Development, testing, and production deployment of CRM and ERP products for the TT Hellenic Postbank and the Volkswagen Bank Greece.

## **Research Visits and Meetings**

**December 2019, Dagstuhl Seminar 19491: "Big Graph Processing Systems", Germany**

**June 2017, NII Shonan Meeting, Shonan Village Center (SVC), Japan**

**March - May 2013, TU Berlin, Germany**

## Selected Invited Talks

### 2020, Tufts CS Colloquium

Towards self-managed and re-configurable stream processing systems

### 2020, North East Database Day 2020, Cambridge, MA

From data stream management to distributed dataflows and beyond (Keynote)

### 2018, 3rd Int'l Summer School on Data Science, Split, Croatia

Platforms for big data analytics and stream processing

### 2018, Distributed Computing & Analytics Workshop, Kista, Sweden

Automatic scaling decisions for distributed streaming dataflows

### 2017, QCon San Francisco, California, USA

Predictive Datacenter Analytics with Strymon

### 2017, 31st British Int'l Conference on Databases, London, UK

Programming Models and Tools for Distributed Graph Processing

### 2017, O'Reilly Velocity London, UK

Online performance analysis of distributed dataflow systems

### 2016, 2nd Int'l ScaDS Summer School on Big Data, Leipzig, Germany

Batch and Stream Graph Processing with Apache Flink

### 2016, Summer School on Cloud and Big Data, Stockholm, Sweden

Introduction to Distributed Graph Processing with Apache Flink

### 2016, Berlin Buzzwords, Berlin, Germany

Graphs As Streams: Rethinking Graph Processing in the Streaming Era

### 2016, dotScale: The European Conference on Scalability, Paris, France

Demystifying Distributed Graph Processing

### 2015, Bay Area Apache Flink Meetup, San Jose, California, USA

Graph Processing with Gelly

## List of Publications

### Conferences and Journals

**A New Benchmark Harness for Systematic and Robust Evaluation of Streaming State Stores.** Esmail Asyabi, Yuanli Wang, John Liagouris, *Vasiliki Kalavri*, Azer Bestavros. Proceedings of the 17th European Conference on Computer Systems, (EuroSys'22, to appear).

**Learning on Streaming Graphs with Experience Replay.** Massimo Perini, Giorgia Ramponi, Paris Carbone, *Vasiliki Kalavri*. Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing (SAC '22).

**SoK: Function-As-A-Service: From An Application Developer's Perspective.** Ali Raza, Ibrahim Matta, Nabeel Akhtar, *Vasiliki Kalavri*, Vatche Isahagian. Journal of Systems Research (JSys) 1(2) 01 Sep 2021.

**The future is big graphs: a community view on graph processing systems.** Sherif Sakr, Angela Bonifati, Hannes Voigt, Alexandru Iosup, et al. Communications of the ACM 64(9):62-71 Sep 2021.

**Practice of streaming and dynamic graphs: Concepts, models, systems, and parallelism.** Besta Maciej, Marc Fischer, *Vasiliki Kalavri*, Michael Kapralov, and Torsten Hoefler. IEEE

Transactions on Parallel and Distributed Systems (TPDS) 1-1, 2021.

**Beyond Analytics: the Evolution of Stream Processing Systems.** Paris Carbone, Marios Fragkoulis, *Vasiliki Kalavri*, Asterios Katsifodimos. Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (Tutorial).

**Megaphone: latency-conscious state migration for distributed streaming dataflows.** Moritz Hoffmann, Andrea Lattuada, Frank McSherry, *Vasiliki Kalavri*, John Liagouris, Timothy Roscoe. Proceedings of the VLDB Endowment, Vol. 12, No. 9 (PVLDB), 2019.

**Streaming Graph Partitioning: An Experimental Study.** Zainab Abbas, *Vasiliki Kalavri*, Paris Carbone, Vladimir Vlassov. Proceedings of the VLDB Endowment, Vol. 11, No. 11 (PVLDB), 2018.

**High-Level Programming Abstractions for Distributed Graph Processing.** *Vasiliki Kalavri*, Vladimir Vlassov, Seif Haridi. IEEE Transactions in Knowledge and Data Engineering (TKDE), 30 (2), 305-324, 2018.

**Three steps is all you need: fast accurate, automatic scaling decisions for distributed streaming dataflows.** *Vasiliki Kalavri*, John Liagouris, Moritz Hoffmann, Desislava Dimitrova, Matthew Forshaw, Timothy Roscoe. 13th USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2018.

**SnailTrail: Generalizing Critical Paths for Online Analysis of Distributed Dataflows.** Moritz Hoffmann, Andrea Lattuada, John Liagouris, *Vasiliki Kalavri*, Desislava Dimitrova, Sebastian Wicki, Zaheer Chothia, Timothy Roscoe. 15th USENIX Symposium on Networked Systems Design and Implementation (NSDI), 2018.

**The shortest path is not always a straight line: Leveraging semi-metricity in graph analysis.** *Vasiliki Kalavri*, Tiago Simas, Dionysios Logothetis. Proceedings of the VLDB Endowment, Vol. 9, No. 9 (PVLDB), 2016.

**Like a Pack of Wolves: Community Structure of Web Trackers.** *Vasiliki Kalavri*, Jeremy Blackburn, Matteo Varvello, Konstantina Papagiannaki. 17th International Conference on Passive and Active Measurement. Vol. 9631. Springer (PAM), 2016.

**MapReduce: Limitations, Optimizations and Open Issues.** *Vasiliki Kalavri*, Vladimir Vlassov. 11th IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA), 2013.

**PonIC: Using Stratosphere to Speed Up Pig Analytics.** *Vasiliki Kalavri*, Vladimir Vlassov, Per Brand. 19th European Conference on Parallel Processing, (Euro-Par) 2013.

**m2r2: A Framework for Results Materialization and Reuse in High-Level Dataflow Systems for Big Data.** *Vasiliki Kalavri*, Hui Shang, Vladimir Vlassov. 16th IEEE International Conference on Computational Science and Engineering (CSE), 2013.

**Block Sampling: Efficient Accurate Online Aggregation in MapReduce.** *Vasiliki Kalavri*, Vaidas Brundza, Vladimir Vlassov. 5th IEEE International Conference on Cloud Computing Technology and Science (CloudCom) 2013.

## Books

**Stream Processing with Apache Flink *Fundamentals, Implementation, and Operation of Streaming Applications***. Fabian Hueske, *Vasiliki Kalavri*. O'Reilly Media 2019.

## Workshops

**In support of workload-aware streaming state management**. *Vasiliki Kalavri*, John Liagouris . 12th USENIX Workshop on Hot Topics in Storage and and File Systems, 2020.

**FASTER State Management for Timely Dataflow**. Matthew Brookes, *Vasiliki Kalavri*, John Liagouris. Proceedings of Real-Time Business Intelligence and Analytics,(BIRTE), 2019.

**Asymmetry in large-scale graph analysis, explained**. *Vasiliki Kalavri*, Stephan Ewen, Kostas Tzoumas, Vladimir Vlassov, Volker Markl, Seif Haridi. Proceedings of Workshop on GRAPH Data management Experiences and Systems, 1-7, (GRADES), 2014.

## Theses

**Performance Optimization Techniques and Tools for Distributed Graph Processing**. Doctoral Thesis in Information and Communication Technology, KTH Royal Institute of Technology, Sweden, and Université catholique de Louvain, Belgium, 2016.

**Performance Optimization Techniques and Tools for Data-Intensive Computation Platforms**. Licentiate Thesis in Information and Communication Technology, KTH Royal Institute of Technology, Sweden, 2014.

**Integrating Pig and Stratosphere**. MSc Thesis in Distributed Computing, KTH Royal Institute of Technology, Sweden, 2012.

**Parallelization of Linear Algebra Algorithms for High-Scale Architectures**. Diploma Thesis in Electrical and Computer Engineering, National Technical University of Athens, Greece, 2010.

## Pre-prints

**Secrecy: Secure collaborative analytics on secret-shared data**. John Liagouris, *Vasiliki Kalavri*, Muhammad Faisal, Mayank Varia, arXiv:2102.01048, 2021.

**A Survey on the Evolution of Stream Processing Systems**. Marios Fragkoulis, Paris Carbone, *Vasiliki Kalavri*, Asterios Katsifodimos, arXiv:2008.00842, 2020.

**DeltaPath: dataflow-based high-performance incremental routing**. Desislava Dimitrova, John Liagouris, Sebastian Wicki, Moritz Hoffmann, *Vasiliki Kalavri*, Timothy Roscoe. arXiv:1808.06893, 2018.