

Vasiliki (Vasia) Kalavri

Assistant Professor

Boston University College of Arts & Sciences
Department of Computer Science
665 Commonwealth Avenue, Boston, MA 02215
vkalavri@bu.edu
<https://cs-people.bu.edu/vkalavri/>

Academic Appointments

January 2020 - now

Tenure-Track Assistant Professor | Boston University
Department of Computer Science, College of Arts & Sciences

July 2019 - December 2019

Adjunct Assistant Professor | Boston University
Department of Computer Science, College of Arts & Sciences

January 2017 - October 2019

Postdoctoral Fellow | ETH Zürich
Systems Group, Department of Computer Science

Education

2012 - 2016

PhD in Information and Communication Technology (double degree) |
KTH Royal Institute of Technology and Université Catholique de Louvain
Thesis: *Performance Optimization Techniques and Tools for Distributed Graph Processing*
Advisors: Vladimir Vlassov (KTH), Seif Haridi (SICS), Peter Van Roy (UCLouvain)
Opponent: Peter Boncz (CWI and VU Amsterdam)

2010 - 2012

MSc in Distributed Computing (double degree) |
KTH Royal Institute of Technology and Universitat Politècnica de Catalunya
Thesis: *Integrating Pig and Stratosphere*
Advisor: Per Brand (SICS)

2004 - 2010

MEng in Electrical and Computer Engineering | National Technical University of Athens
Thesis: *Parallelization of Linear Algebra Algorithms for High-Scale Architectures*
Advisors: Georgios Goumas and Nectarios Koziris (NTUA)

Awards and Honors

2026

- ACM SIGMOD Distinguished Reviewer Award

2025

- NSF Faculty Early Career Development Program (CAREER) Award
- SOSP Distinguished Artifact Award
for the paper *ORQ: Complex Analytics on Private Data with Strong Security Guarantees*
- VLDB Distinguished Reviewer Award

2024

- Digital Futures Summer Early Career Scholar

- **DaMoN Best Paper Award**
for the paper *In Situ Neighborhood Sampling for Large-Scale GNN Training*

2023

- **SIGMOD Systems Award**
for *Apache Flink*, which greatly expanded the use of stream data processing
- **VLDB Distinguished Reviewer Award**
- **SIGMOD Best Reviewer Award**
- **Confluent Community Catalyst**
The Catalysts are elite contributors who multiply and empower the developer community and share the goodness of event streaming platforms with others.

2022

- **DEEM Best Paper Award**
for the paper *Evaluating model serving strategies over streaming data*

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
- **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

Research Funding

2026 - 2029: **National Science Foundation | Award Number: 2613424**

CICI: TCR: Automated Orchestration of Scientific Data Workflows in Open Clouds with Strong Security Guarantees.

Co-PI | Primary PI: Mayank Varia, co-PI: John Liagouris; USD 1,199,997

2026 - 2027: **i-Scale The Center for Systems Innovation at Scale**

Streaming GNN Inference for Real-time Anomaly Detection on Microservice Graphs.

Single PI; USD 75,000

2025 - 2029: **National Science Foundation | Award Number: 2440334**

CAREER: Adaptive resource management and reconfiguration mechanisms for streaming dataflow systems.

Single PI; USD 676,807

2023 - 2025: **National Science Foundation | Award Number: 2237193**

CNS: CORE: Small: Scaling Graph Machine Learning Workloads on Modern Storage.

Single PI; USD 469,787

2023: **RedHat Research Incubation Award**

Towards high performance and energy efficiency in open-source stream processing.

Primary PI | Co-PI: Jonathan Appavoo; USD 174,876

2022 - 2025: **National Science Foundation | Award Number: 2209194**

SaTC: CORE: Medium: Secure outsourced analytics in untrusted clouds.

Co-PI | Primary PI: John Liagouris, co-PI: Mayank Varia; USD 749,952

2022: RedHat Research Incubation Award

Serverless Streaming Graph Analytics.

Single PI; USD 85,000

2022: RedHat Research Incubation Award

Towards high performance and energy efficiency in open-source stream processing.

Primary PI | Co-PI: Jonathan Appavoo; USD 150,000

2022: Google DAPA Faculty Award

Self-managed streaming systems and workload-aware state management.

Single PI; USD 50,000

2021: RedHat Research Incubation Award

Serverless Streaming Graph Analytics.

Single PI; USD 85,000

2021: RedHat Research Incubation Award

Towards high performance and energy efficiency in open-source stream processing.

Primary PI | Co-PI: Jonathan Appavoo; USD 150,000

2021: RedHat Research Incubation Award

Creating a global open research platform to better understand social sustainability using data from a real-life smart village.

Collaborator | PI: Christos G. Cassandras; USD 150,000

2021: Hariri Institute Focused Research Program

Continuous Analysis of Mobile Health Data among Medically Vulnerable Populations.

Co-PI | Co-PIs: Lisa Quantilliani and Nicole Spartano; USD 150,000

2021: Google DAPA Faculty Award

Self-managed streaming systems and workload-aware state management.

Single PI; USD 50,000

2021: Samsung Memory Solutions Lab UR Collaboration Grant

SSD-aware Context Serving for Streaming Graph Representation Learning.

Single PI; USD 85,000

2020: Google DAPA Faculty Award

Self-managed streaming systems and workload-aware state management.

Single PI; USD 50,000

Teaching Activities

Boston University

CS 210 Computer Systems (undergraduate)

- **Fall 2023** | Enrollment: 281 | Co-Instructors: James Cadden, Preethi Narayanan
- **Fall 2021** | Enrollment: 132 | Sole instructor
- **Fall 2020** | Enrollment: 245 | Co-Instructor: Jonathan Appavoo

CS 351 Distributed Systems (undergraduate)

- **Fall 2025** | Enrollment: 185 | Co-Instructor: Anna Arpaci-Dusseau

CS 551 Streaming and Event-Based Systems (graduate)

- **Spring 2026** | Enrollment: 35 | Sole instructor
- **Spring 2025** | Enrollment: 20 | Sole instructor
- **Spring 2024** | Enrollment: 35 | Sole instructor
- **Spring 2023** | Enrollment: 37 | Sole instructor

- **Spring 2022** | Enrollment: 38 | Sole instructor

CS 591 K1 Data Stream Processing and Analytics

- **Spring 2021** | Enrollment: 18 | Sole instructor

ETH Zürich

263-3826-00 Data Stream Processing and Analytics

- **Spring 2021** | Sole instructor

KTH Royal Institute of Technology

ID2219 Implementation of Distributed Systems

- **2015** | Semester project supervisor

ID2212 Network programming with Java

- **2015** | Guest lecturer

ID1218 Applied Programming

- **2011** | Head TA

ID220 Distributed Systems, Basic Course

- **2011** | Head TA

Université Catholique de Louvain

LSINF2345 Languages and Algorithms for Distributed Applications

- **2014** | Head TA

Industry Courses for Ericsson

Distributed Systems and Algorithms with Erlang

- **2012** | TA | Locations: Sweden, China, USA, Hungary

Functional Programming with Haskell

- **2012** | TA | Locations: China

Summer Schools and Tutorials

- **Secure MPC from a Systems Perspective: A Hands-on Tutorial with ORQ**, ACM EuroSys 2026 Tutorial
- **VLDB Summer School**, Romania 2023
- **Beyond Analytics: the Evolution of Stream Processing Systems**, ACM SIGMOD 2020 Tutorial
- 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 at Boston University

- 2020-now: Muhammad Faisal
- 2021-now: Yuanli Wang
- 2021-now: Lei Huang | Co-advised with Abraham Matta
- 2023-now: Naima Abrar Shami
- 2024-now: Yuhang Song
- 2025-now: Sakshi Sharma | Co-advised with John Liagouris

PhD students at KTH Royal Institute of Technology

- 2018-2021 (Graduated): Zainab Abbas | Co-advised with Vladimir Vlassov and Paris Carbone | Next: Solutions Architect @ Databricks

Undergraduate researchers at Boston University

- 2025-now: Arnav Kulkarni
- 2025-now: Alessandra Lanz
- 2025-now: Adam Godel
- 2023-2025: Qixuan (Talia) Chen | Next: PhD student at UT Austin
- 2023-2025: Zikun Wang | Next: PhD student at Northeastern University
- 2023-2024: Devon Lewis | Next: PhD student at Rutgers University
- 2022-2024: Yuchen Lu | Next: PhD student at Brown University
- 2022-2024: Po Hao Chen | Next: PhD student at Brown University
- 2023-2024: Shuhan Gu | Next: MS in Computer Science at Cornell University
- 2024: Jake Gustin | Next: MS in Computer Science at Georgia Tech
- 2024: Katherine Rimey | Next: MS in Computer Science at Yale University
- 2022-2023: Jerry Zhang | Next: Computer Vision Engineer at Qualcomm
- 2021-2022: Baiqing Lyu | Next: Software Engineer at TikTok
- 2022: Ke Li | Next: MS in Software Engineering at CMU
- 2020-2021: Zhifan Xue | Next: MS in Computer Science at University of Michigan

Master's student researchers at Boston University

- 2025-now: Chenxu Yu
- 2025: Pranav Goyanka | Next: Software Engineer at Goldman Sachs
- 2023-2024: Atul Lal | Next: Software Engineer at Confluent
- 2023-2024: Yan Tong | Next: PhD student at UC Santa Cruz
- 2023: Shubham Kaushik | Next: PhD student at Brandeis University
- 2022-2023: Yizheng Xie | Next: PhD student at Brown University
- 2022-2023: Jingyu Su | Next: Software Engineer at Qualcomm
- 2022: Shaolin Xie | Next: PhD student at University of Southern California

Bachelor and Master Thesis Supervision

Accelerating Large-Scale GNN Training with a SmartSSD

MSc Thesis, Yuhang Song, 2024, Boston University

Offloading the Sampling Stage of GNN Training to Smart Storage

MSc Thesis, Emmanouil Kritharakis, 2023, Boston University

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 Poulovassilis (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

Service

Reviewer for Conferences, Journals, and Workshops

ACM EuroSys 2026, 2027

ACM HotStorage 2026

SOSP 2026 Student Research Competition

GRADES-NDA 2026

SeQureDB 2026

USENIX NSDI 2026

Northeast Systems Day 2026

USENIX ATC 2021, 2023, 2024, 2025

ACM SIGMOD 2022, 2023, 2026

VLDB 2022, 2023, 2024, 2025

ACM DEBS 2020, 2021, 2022, 2023, 2024, 2026

IEEE Transactions on Computer Systems (TOCS) 2025

VLDB Journal 2020, 2021, 2022, 2023, 2025

APSys 2022, 2024

IEEE Transactions on Parallel and Distributed Systems 2019, 2021, 2022, 2023, 2024

IEEE Transactions on Knowledge and Data Engineering 2020, 2024

DEEM 2017, 2024 (co-located with SIGMOD)

CIDR 2023

SOSP 2023 Posters

SOSP 2023 Student Research Competition

ACM SIGMOD 2022 Demos

ACM SIGMOD 2022 Reproducibility

ACM SoCC 2022

ACM Middleware 2021

ACM Middleware Doctoral Symposium 2020, 2021

IEEE ICDE 2021

USENIX HotStorage 2020

IEEE ICDE Demos 2020

EDBT Demos 2020

OPODIS 2018

GRADES-NDA 2018, 2019, 2020 (co-located with SIGMOD)

DBTest 2018 (co-located with SIGMOD)

GABB 2017, 2018 (co-located with IPDPS)

Organizing Roles

USENIX OSDI 2025 & ATC 2025 Mentorship co-Chair

ACM DEBS 2025 Workshops and Tutorials co-Chair

Northeast Database Day Program co-Chair

ACM SIGMOD Student Research Competition 2022 co-Chair

IEEE ICDE 2022 Associate Chair

Journal of Systems Research Streaming Systems Area Chair 2022-2024
GRADES-NDA 2021, 2022 co-Chair (co-located with SIGMOD)
ACM EuroSys 2021 AMA Chair
CCGrid 2019 Applications and Data Science track co-Chair

Proposal Reviewer

2026: Served in NSF Future CoRe panel
2026: Served as a Scientific Evaluator for the Swiss National Science Foundation
2025: Served in NSF Medium Panel

Reviewer for Industrial Conferences

Confluent Current 2022, 2024, 2025
Flink Forward Europe 2019, 2024
Flink Forward San Francisco 2017
Berlin Buzzwords 2016, 2017
Flink Forward Berlin 2016

Doctoral Dissertation Committees

2025:

- Ankit Chaudhary, TU Berlin
- Kajetan Maliszewski, TU Berlin
- Parul Sohal, Boston University
- Anam Farrukh, Boston University
- Aneesh Raman, Boston University
- Tarikul Islam Papon, Boston University

2024:

- George Siachamis, Delft University of Technology
- Zhongshun Zhang, Boston University

2023:

- Khaled Ammar, University of Waterloo
- Tommy Unger, Boston University
- Ali Raza, Boston University

2022:

- Sohan Sinha, Boston University

2021:

- Zainab Abbas, KTH Royal Institute of Technology
- Miguel Coimbra, Instituto Superior Tecnico

Master Thesis Committees

2025: Hyunjin Jung, Boston University
2024: Sanskriti Sharma, Boston University
2023: Emmanouil Kritharakis, Boston University
2022: Vivek Unnikrishnan, Boston University

Other Boston University Service

2022-2023: CS Tenure-Track Hiring Committee
2022-2024: CS Graduate Awards Committee Chair
2021: CS PhD Advising Policies Committee
2020-2022: CS PhD Admissions Committee
2020-2023: Faculty Advisor for the ACM-Women student chapter

Events Organization

Workshops at Boston University

- July 2025: *1st Workshop on Systems for Secure Multi-Party Computation (SysteMPC'25)*
- February 2022: *Uncovering synergies between clinical mHealth use cases, streaming data processing, and machine learning approaches.*
- November 2021: *Secure analytics on mobile health data streams.*
- July 2021: *Development of Use Cases for Continuous Mobile Data in Health Promotion & Disease Detection.*

Meetups

2017 - 2019: Apache Flink Meetup, Zurich

2017 - 2019: Papers we Love Meetup, Zurich

2015 - 2016: Apache Flink Meetup, Stockholm

Developer Conferences

2016 - 2019: 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, Jeremy Blackburn.

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

- September 2024, Visiting Professor, KTH, Sweden
- 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

Talks

2026, Open Source Data Streaming Meetup Bay Area @Snowflake (Menlo Park, CA)

Talk title: *Deep Dive into Flink's Disaggregated State Management*

2026, Columbia University & UC Santa Barbara Colloquium

Talk title: *Rethinking Streaming Dataflow Systems: A Case for Disaggregated, Adaptive, and Resource-Aware Engines*

2025, NSF CSR PI Meeting (Boca Raton, FL)

Talk title: *Scaling Graph Machine Learning Workloads on Modern Storage*

2025, Olympia Research Summit (Athens, Greece)

Talk title: *Towards adaptive and resource-efficient data stream processing systems*

2025, BIFOLD Colloquium (TU Berlin, Germany)

Talk title: *Secure collaborative analytics with MPC made practical*

2024, Brown University (Providence, RI)

Guest lecture for CSCI2380 “Distributed Systems”

2024, HPI & TU Darmstadt Research Seminar (virtual)

Talk title: *Secure and expressive time series analytics in untrusted clouds*

2024, Digital Futures Seminar (KTH, Sweden)

Talk title: *Cryptographically secure analytics in the Cloud*

2024, ITU Advanced Data Systems seminar (Copenhagen, Denmark)

Talk title: *Cryptographically secure analytics in the Cloud*

2023, CASTOR Software Days (KTH, Sweden)

Talk title: *Scaling Graph Machine Learning Workloads on Modern Storage*

2023, National Technical University of Athens (Athens, Greece)

Talk title: *SECRECY: Secure collaborative analytics in untrusted clouds*

2023, University of Massachusetts Dartmouth (Dartmouth, MA)

Talk title: *Secure collaborative analytics in untrusted clouds*

2022, University of Waterloo Data Systems Group Seminar (Canada)

Talk title: *Efficient collaborative analytics with no information leakage*

2022, The Dutch Seminar on Data Systems Design (virtual)

Talk title: *Efficient collaborative analytics with no information leakage*

2022, ACM EuroSys (Rennes, France)

Talk title: *A New Benchmark Harness for Systematic and Robust Evaluation of Streaming State Stores*

2022, Samsung Memory Solutions Lab Tech Talk (virtual)

Talk title: *Next-generation state management for streaming workloads*

2021, Google DAPA Seminar (virtual)

Talk title: *Next-generation state management for streaming workloads*

2021, 14th LDBC TUC meeting (virtual)

Talk title: *Learning to partition unbounded graph streams*

2020, Tufts CS Colloquium (virtual)

Talk title: *Towards self-managed and re-configurable stream processing systems*

2020, Database Seminar, Columbia University (virtual)

Talk title: *Recent advances and open problems in large-scale stream processing*

2020, University of Texas at Austin, Distributed Systems Guest Lecture (virtual)

Talk title: *Recent advances and open problems in large-scale stream processing*

2020, Keynote at the North East Database Day (Cambridge, MA)

Talk title: *From data stream management to distributed dataflows and beyond*

2018, 3rd Int’l Summer School on Data Science (Split, Croatia)

Talk title: *Platforms for big data analytics and stream processing*

2018, Distributed Computing & Analytics Workshop (Kista, Sweden)

Talk title: *Automatic scaling decisions for distributed streaming dataflows*

2017, QCon San Francisco (California)

Talk title: *Predictive Datacenter Analytics with Strymon*

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

Talk title: *Programming Models and Tools for Distributed Graph Processing*

2017, O'Reilly Velocity (London, UK)

Talk title: *Online performance analysis of distributed dataflow systems*

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

Talk title: *Batch and Stream Graph Processing with Apache Flink*

2016, Summer School on Cloud and Big Data (Stockholm, Sweden)

Talk title: *Introduction to Distributed Graph Processing with Apache Flink*

2016, Berlin Buzzwords (Berlin, Germany)

Talk title: *Graphs As Streams: Rethinking Graph Processing in the Streaming Era*

2016, dotScale: The European Conference on Scalability (Paris, France)

Talk title: *Demystifying Distributed Graph Processing*

2015, Bay Area Apache Flink Meetup (San Jose, CA)

Talk title: *Graph Processing with Gelly*

List of Publications

Conferences and Journals

CJ1: Koala: Non-disruptive Reconfiguration for Long-running Stateful Dataflows.

Lei Huang, Sakshi Sharma, Chenxu Yu, Zikun Wang, Yuanli Wang, Zichen Zhu, John Liagouris, *Vasiliki Kalavri*. In 24th USENIX Symposium on Networked Systems Design and Implementation (NSDI'27). To appear.

CJ2: CryptDough: A Unified Analytics Engine for Secure Multiparty Computation.

Muhammad Faisal, Alessandra Lanz, Sam Buxbaum, Adam Godel, *Vasiliki Kalavri*, Mayank Varia, John Liagouris. In Proceedings of the ACM SIGOPS 32nd Symposium on Operating Systems Principles (SOSP'26). To appear.

CJ3: ConRAD: Conformal Risk-Aware Neural Databases.

Sonia Horchidan, Fabian Zeiher, Xiangyu Shi, *Vasiliki Kalavri*, Henrik Boström, Ioannis Kontoyiannis, Paris Carbone. Proceedings of the VLDB Endowment, Vol. 19 (VLDB'26). To appear. <https://arxiv.org/abs/2605.03806>

CJ4: ORQ: Scaling Complex Multiparty Computations to Large Private Datasets.

Eli Baum, Sam Buxbaum, Nitin Mathai, Muhammad Faisal, *Vasiliki Kalavri*, Mayank Varia, John Liagouris. ACM Transactions on Computer Systems (TOCS), 2026. Just Accepted. <https://doi.org/10.1145/3820248>

CJ5: Slowpoke: End-to-end Throughput Optimization Modeling for Microservice Applications.

Yizheng Xie, Di Jin, Oguzhan Colkesen, *Vasiliki Kalavri*, John Liagouris, and Nikos Vasilakis. In 23rd USENIX Symposium on Networked Systems Design and Implementation (NSDI'26). <https://www.usenix.org/conference/nsdi26/presentation/xie>

CJ6: Disaggregated State Management in Apache Flink 2.0.

Mei, Yuan, Rui Xia, Zhaoqian Lan, Kaitian Hu, Lei Huang, Paris Carbone, Yanfei Lei, *Vasiliki Kalavri*, Han Yin, and Feng Wang. Proceedings of the VLDB Endowment 18, no. 12 (VLDB'25), pp. 4846–4859. <https://www.vldb.org/pvldb/vol18/p4846-mei.pdf>

- CJ7: ORQ: Complex Analytics on Private Data with Strong Security Guarantees.** Eli Baum, Sam Buxbaum, Nitin Mathai, Muhammad Faisal, *Vasiliki Kalavri*, Mayank Varia, and John Liagouris. In Proceedings of the ACM SIGOPS 31st Symposium on Operating Systems Principles, (SOSP’25), pp. 802–833. <https://dl.acm.org/doi/epdf/10.1145/3731569.3764833>
- CJ8: CAPSys: Contention-aware task placement for data stream processing.** Yuanli Wang, Lei Huang, Zikun Wang, *Vasiliki Kalavri*, Ibrahim Matta. In the 20th European Conference on Computer Systems (EuroSys ’25), pp. 654–670. <https://dl.acm.org/doi/epdf/10.1145/3689031.3696085>
- CJ9: Crayfish: Navigating the Labyrinth of Machine Learning Inference in Stream Processing Systems.** Sonia Horchidan, Po Hao Chen, Emmanouil Kritharakis, Paris Carbone, *Vasiliki Kalavri*. In the 27th International Conference on Extending Database Technology (EDBT’24), pp. 676–689. <https://openproceedings.org/2024/conf/edbt/paper-156.pdf>
- CJ10: A Survey on the Evolution of Stream Processing Systems.** Marios Fragkoulis, Paris Carbone, *Vasiliki Kalavri*, Asterios Katsifodimos. The VLDB Journal, 2023, pp.1–35. <https://doi.org/10.1007/s00778-023-00819-8>
- CJ11: TVA: A multi-party computation system for secure and expressive time series analytics.** Muhammad Faisal, Jerry Zhang, John Liagouris, *Vasiliki Kalavri*, Mayank Varia. 32nd USENIX Security Symposium (USENIX Security’23), pp. 5395–5412. <https://www.usenix.org/system/files/usenixsecurity23-faisal.pdf>
- CJ12: SECRECY: Secure collaborative analytics in untrusted clouds.** John Liagouris, *Vasiliki Kalavri*, Muhammad Faisal, Mayank Varia. 20th USENIX Symposium on Networked Systems Design and Implementation (NSDI’23), pp. 1031–1056. <https://www.usenix.org/system/files/nsdi23-liagouris.pdf>
- CJ13: 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), pp. 559–574. <https://dl.acm.org/doi/epdf/10.1145/3492321.3519592>
- CJ14: 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), pp. 470–478. <https://doi.org/10.1145/3477314.3507113>
- CJ15: 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. <https://doi.org/10.5070/SR31154815>
- CJ16: The future is big graphs: a community view on graph processing systems.** Sherif Sakr, Angela Bonifati, Hannes Voigt, et al. Communications of the ACM 2021. <https://dl.acm.org/doi/epdf/10.1145/3434642>
- CJ17: Practice of streaming and dynamic graphs: Concepts, models, systems, and parallelism.** Besta Maciej, Marc Fischer, *Vasiliki Kalavri*, Michael Kapralov, and Torsten Hoeftler. IEEE Transactions on Parallel and Distributed Systems (TPDS)

1-1, 2021, pp. 1860–1876. <https://doi.org/10.1109/TPDS.2021.3131677>

- CJ18: 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, pp. 1002–1015. <http://www.vldb.org/pvldb/vol12/p1002-hoffmann.pdf>
- CJ19: 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, pp. 1590–1603. <http://www.vldb.org/pvldb/vol11/p1590-abbas.pdf>
- CJ20: High-Level Programming Abstractions for Distributed Graph Processing.** *Vasiliki Kalavri*, Vladimir Vlassov, Seif Haridi. IEEE Transactions in Knowledge and Data Engineering (TKDE), 30 (2), pp. 305–324, 2018, pp. 305-324. <https://doi.org/10.1109/TKDE.2017.2762294>
- CJ21: 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’18), pp. 783–798. <https://www.usenix.org/system/files/osdi18-kalavri.pdf>
- CJ22: 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’18), pp. 95–110. <https://www.usenix.org/system/files/conference/nsdi18/nsdi18-hoffmann.pdf>
- CJ23: 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, pp. 672–683. <http://www.vldb.org/pvldb/vol9/p672-kalavri.pdf>
- CJ24: 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 (PAM’16), pp. 42–54. https://link.springer.com/chapter/10.1007/978-3-319-30505-9_4
- CJ25: MapReduce: Limitations, Optimizations and Open Issues.** *Vasiliki Kalavri*, Vladimir Vlassov. 11th IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA), 2013, pp. 250–257. <https://doi.org/10.1109/TrustCom.2013.126>
- CJ26: PonIC: Using Stratosphere to Speed Up Pig Analytics.** *Vasiliki Kalavri*, Vladimir Vlassov, Per Brand. 19th European Conference on Parallel Processing, (Euro-Par’13), Springer Vol. 8097, pp. 279–290. https://link.springer.com/chapter/10.1007/978-3-642-40047-6_30
- CJ27: 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, pp. 894–901. <https://doi.org/10.1109/CSE.2013.134>

CJ28: 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, pp. 250–257. <https://doi.org/10.1109/CloudCom.2013.40>

Preprints

P1: Not All Neighbors Matter: Understanding the Impact of Graph Sparsification on GNN Pipelines. Song, Yuhang, Naima Abrar Shami, Romaric Duvignau, and *Vasiliki Kalavri*. arXiv preprint arXiv:2603.06952 (2026). <https://arxiv.org/abs/2603.06952>

Books

B1: Stream Processing with Apache Flink: Fundamentals, Implementation, and Operation of Streaming Applications. Fabian Hueske, *Vasiliki Kalavri*. O'Reilly Media 2019.

Workshops, Demos, and Tutorials

W1: RingSampler: GNN sampling on large-scale graphs with io_uring. Qixuan Chen, Yuhang Song, Melissa Martinez, and *Vasiliki Kalavri*. In Proceedings of the 17th ACM Workshop on Hot Topics in Storage and File Systems (HotStorage'25), pp. 52–60. <https://dl.acm.org/doi/pdf/10.1145/3736548.3737829>

W2: Bridging GNN Inference and Dataflow Stream Processing: Challenges and Opportunities. Naima Abrar Shami and *Vasiliki Kalavri*. In Proceedings of the 8th Joint Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA). 2025, pp. 7:1–7:10. <https://dl.acm.org/doi/pdf/10.1145/3735546.3735856>

W3: QueryShield: Cryptographically Secure Analytics in the Cloud. Ethan Seow, Yan Tong, Eli Baum, Sam Buxbaum, Muhammad Faisal, John Liagouris, *Vasiliki Kalavri*, Mayank Varia. 2025 International Conference on Management of Data (ACM SIGMOD'25 Demo Track), pp. 436–439. <https://dl.acm.org/doi/10.1145/3626246.3654749>

W4: In situ neighborhood sampling for large-scale GNN training. Yuhang Song, Po Hao Chen, Yuchen Lu, Naima Abrar, *Vasiliki Kalavri*. Proceedings of the 20th International Workshop on Data Management on New Hardware, DaMoN'24, pp. 11:1–11:5. <https://dl.acm.org/doi/pdf/10.1145/3662010.3663443>

W5: Bounding the State of Streaming Graph Partitioning. Michał Zwolak, Zainab Abbas, Sonia Horchidan, Paris Carbone, *Vasiliki Kalavri*. Proceedings of the Fifth International Workshop on Exploiting Artificial Intelligence Techniques for Data Management, aiDM'22, co-located with SIGMOD'22, pp. 3:1–3:12. <https://dl.acm.org/doi/pdf/10.1145/3533702.3534920>

W6: Evaluating Model Serving Strategies over Streaming Data. Sonia Horchidan, Emmanouil Kritharakis, *Vasiliki Kalavri*, Paris Carbone. Proceedings of the Sixth Workshop on Data Management for End-To-End Machine Learning (DEEM'22), pp. 4:1–4:5. <https://dl.acm.org/doi/10.1145/3533028.3533308>

W7: The Non-Expert Tax: Quantifying the cost of auto-scaling in Cloud-based data

stream analytics. Yuanli Wang, Baiqing, Lyu, *Vasiliki Kalavri*. Proceedings of The International Workshop on Big Data in Emergent Distributed Environments (BiDEDE'22, co-located with SIGMOD'22), pp. 7:1–7:7. <https://dl.acm.org/doi/10.1145/3530050.3532925>

W8: 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), pp. 2651–2658. <https://dl.acm.org/doi/10.1145/3318464.3383131>

W9: In support of workload-aware streaming state management. *Vasiliki Kalavri*, John Liagouris. 12th USENIX Workshop on Hot Topics in Storage and File Systems, (HotStorage'20). https://www.usenix.org/system/files/hotstorage20_paper_kalavri.pdf

W10: FASTER State Management for Timely Dataflow. Matthew Brookes, *Vasiliki Kalavri*, John Liagouris. Proceedings of Real-Time Business Intelligence and Analytics, (BIRTE), 2019. <https://dl.acm.org/doi/pdf/10.1145/3350489.3350493>

W11: 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, pp. 4:1–4:7. <https://dl.acm.org/doi/pdf/10.1145/2621934.2621940>

Theses

T1: 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.

T2: Performance Optimization Techniques and Tools for Data-Intensive Computation Platforms. Licentiate Thesis in Information and Communication Technology, KTH Royal Institute of Technology, Sweden, 2014.

T3: Integrating Pig and Stratosphere. MSc Thesis in Distributed Computing, KTH Royal Institute of Technology, Sweden, 2012.

T4: Parallelization of Linear Algebra Algorithms for High-Scale Architectures. Diploma Thesis in Electrical and Computer Engineering, National Technical University of Athens, Greece, 2010.