Georgios Siachamis

GitHub | Google Scholar | LinkedIn | Personal website | george@gsiachamis.dev

Distributed systems and backend-oriented engineer with a PhD in database systems and postdoctoral experience in data-intensive software, stream processing, graph data management and data integration. Strong in Java and Python, with hands-on experience building, evaluating, and deploying scalable and fault-tolerant systems. Interested in backend, platform, and distributed systems engineering roles.

Technical Skills

- Programming Languages: Java (Expert), Python (Expert), Shell (intermediate), C (intermediate), Rust (basic), Go (basic), JavaScript (basic), C++ (basic).
- Distributed/Data Systems: Apache Flink, Apache Spark, Apache Hadoop.
- Infrastructure & Tools: Linux, Docker, Kubernetes, Git
- Databases: PostgreSQL, MySQL, MongoDB, Elasticsearch
- Systems concepts: fault tolerance, autoscaling, stream processing, distributed transactions
- Semantic/Data Integration: RDF, OBDA, schema matching, data integration

Employment

Postdoctoral Researcher / Research Engineer Oct 2024 - Present

Inria, CEDAR *Saclay, France*

- Designed and implemented software components for ontology-based data access, graph data lakes, and keyword-search query evaluation.
- Built software components for the Data Exchange Platform (DXP) project, including access-control mechanisms for RDF/OBDA systems and ontology/schema-mapping pipelines for airline-sector data integration.

Ph.D. Candidate Nov 2019 - Apr 2024

Delft University of Technology, Web Information Systems *Delft, Netherlands*

- Designed and evaluated adaptive mechanisms for distributed stream processing systems, focusing on autoscaling, checkpointing, and skew-aware similarity joins.
- Built experimental frameworks and workloads in Java/Python for benchmarking stream processing engines under dynamic conditions.
- Conducted large-scale empirical evaluations of fault-tolerance and autoscaling strategies.
- Co-authored publications in ICDE, DEBS, SIGMOD, and other related venues.

Academic Consultant Jan 2020 - Dec 2023

ING Group, Global Analytics and Tech Infra *Amsterdam, Netherlands*

- Member of the AI for FinTech Research collaboration between ING and TU Delft.
- Contributed to industry-academia collaboration on data systems topics in financial technology and infra management.
- Explored approaches for infrastructure asset management and real-time internal processing tools.
- Facilitated knowledge transfer through seminars and technical exchanges across academic and industrial stakeholders.

Research Intern Apr 2019 - Jun 2019

Université de Cergy-Paris *Paris, France*

- Extended the publicly available code of OrpheusDB to include all the data versioning strategies discussed in the paper, and designed and performed an extended experimental evaluation.

Software Engineer Jan 2015 - Feb 2016

FOCUS ON DIGITAL SERVICES LTD. *Athens, Greece*

- Part of a team that designed and implemented the digital content of an English learning book publisher.
- Worked with Flash, HTML5 and video/audio processing tools.

Education

Ph.D. in Database Systems Nov 2019 - Nov 2024

Delft University of Technology *Delft, Netherlands*

Title: Adaptivity for Streaming Dataflow Engines

Summary: Worked on the adaptivity capabilities of modern streaming dataflow engines, focusing on load balancing, fault tolerance, and autoscaling, where I conducted two extensive experimental evaluations of checkpointing and autoscaling, and proposed a novel solution for adaptive streaming similarity joins.
Advisor: Dr. Asterios Katsifodimos

Diploma (M.Eng.) in Electrical and Computer Engineering Apr 2014 - Oct 2019

Technical University of Athens *Athens, Greece*

Thesis: "Exploration of query similarity using machine learning" (in Greek)

Advisor: Dr. Verena Kantere

Languages

Greek (native speaker), English (professional working proficiency), French (elementary proficiency).

Awards

- 2025 International Conference on Cooperative Information Systems (CoopIS) **Best Paper Award** for "RDF Query Answering in the Presence of Access Restrictions".

Selected Engineering Projects

- **Valentine: (Schema-) Matching DataFrames Made Easy**
  - A Python package for capturing potential relationships among columns of different tabular datasets, which are given as pandas DataFrames.
  - Contributions: Implemented and evaluated matching algorithms included in the package.
- **Extracting Ontologies out of JSON Schema specifications**
  - A set of Python scripts that turn a provided JSON Schema specification into an RDFS ontology, provide statistics on the ontology, and create synthetic datasets corresponding to the airline sector. The scripts use sentence embeddings (SBERT) and clustering to group properties, and LLMs to provide meaningful naming. As this has been developed for the DXP project, the provided specification and the ontology target the airline sector.
  - Contributions: Design and implementation.
- **Styx: Transactional Serverless Functions on Streaming Dataflows**
  - Styx is a distributed system that serves stateful functions. It provides exactly-once processing semantics, seamless fault-tolerance, serializable transactional guarantees, and an intuitive Python API. In our publication, we showed that Styx outperforms the SotA by at least an order of magnitude.
  - Contributions: Implementation of the workloads for all our competitors: Boki (Go), Beldi (Go), T-Statefun (Java/Python). Deployment and execution of all competitors and workloads.
- **Self-hosted Homelab**
  - Built and operate a self-hosted Linux server platform using Docker Compose, monitoring, backup automation, reverse proxying, and internal DNS/service management.
  - Contributions: Deploy and maintain storage, media, and productivity services with an emphasis on reliability, observability, and maintainability.

Teaching

École Polytechnique, Télécom Paris & Institut Polytechnique de Paris November 2025 - May 2026
Saclay, France

- ECE-5DA04-TP : Big Graph Databases, Instructor (Lectures & Labs), Master level.
- CSC-52640-EP: Database Management Systems, Teaching Assistant, Master level.
- CSC-52083-EP: Systems for Big Data, Teaching Assistant, Master level.

Delft University of Technology April 2021 - Apr 2024
Faculty of Electrical Engineering, Mathematics and Computer Science *Delft, Netherlands*

- IN4331: Web-scale Data Management, Teaching Assistant, Master level.
- DSAIT4210: Research in Web Data and Information Management, Teaching Assistant, Master level.

Publications

International Conferences

- **RDF Query Answering in the Presence of Access Restrictions**
  M. Buron, H. Kathuria, I. Manolescu, **G. Siachamis**.
  International Conference on Cooperative Information Systems (CoopIS) 2025.
- **Styx in Action: Transactional Cloud Applications Made Easy**
  K. Psarakis, O.Mraz, G. Christodoulou, **G. Siachamis**, M. Fragkoulis, A. Katsifodimos.
  In the proceedings of the VLDB Endowment (PVLDB) 2025 (demo track)
- **Styx: Transactional Serverless Functions on Streaming Dataflows** [Code]
  K. Psarakis, G. Christodoulou, **G. Siachamis**, M. Fragkoulis, A. Katsifodimos.
  ACM International Conference on Management of Data (SIGMOD) 2026
- **Evaluating Stream Processing Autoscalers** [Code]
  G. Siachamis, G.Christodoulou, K. Psarakis, M. Fragkoulis, A. van Deursen, A. Katsifodimos.
  ACM Conference on Distributed and Event-Based Systems (DEBS) 2024
- **CheckMate: Evaluating Checkpointing Protocols for Streaming Dataflows** [Code]
  G. Siachamis, K. Psarakis, M. Fragkoulis, A. van Deursen, P. Carbone, A. Katsifodimos.
  IEEE International Conference on Data Engineering (ICDE) 2024
- **Adaptive Distributed Streaming Similarity Joins** [Code]
  G. Siachamis, K. Psarakis, M. Fragkoulis, O. Papapetrou, A. van Deursen, A. Katsifodimos.
  ACM Conference on Distributed and Event-Based Systems (DEBS) 2023
- **Valentine in Action: Matching Tabular Data at Scale**
  C. Koutras, K. Psarakis, **G. Siachamis**, A. Ionescu, M. Fragkoulis, A. Bonifati, A. Katsifodimos.
  In the Proceedings of the VLDB Endowment (PVLDB) 2021 (Vol.14, No. 12.) (demo track)
- **Valentine: Evaluating matching techniques for dataset discovery**
  C. Koutras, **G. Siachamis**, A. Ionescu, K. Psarakis, J. Brons, M. Fragkoulis, C. Lofi, A. Bonifati, A. Katsifodimos.
  IEEE International Conference on Data Engineering (ICDE) 2021

International Workshops

- **Towards Evaluating Stream Processing Autoscalers**
  G. Siachamis, J. Kanis, W. Koper, K. Psarakis, M. Fragkoulis, A. van Deursen, A. Katsifodimos.
  International Workshop on Self-managing Database Systems (SMDB) 2023