Kyriakos Psarakis

kPsarakis | kPsarakis | kpsarakis.com | kpsarakis94@gmail.com

Employment

Software Engineer Nov 2025 - present

Ververica GmbH *Remote*

Research Intern Jun 2023 - Sept 2023

Huawei Technologies Research & Development (UK) Ltd *Edinburgh, United Kingdom*

- Implementing query optimization techniques within an MPP database, predominantly regarding join order.
- The new query optimizer is merged in production in Huawei Cloud.

Research Engineer Apr 2020 - Dec 2020

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

- Implemented the scalable schema-matching system used in the Valentine publications.
- Design and development of a benchmark that evaluated the efficiency and consistency (on a large scale) of distributed microservices developed by the student teams of the Web-Scale Data Management course.
- Developed the Data Profiling service of the Topio market platform. Alongside assisting with the development of two Python libraries for geospatial data profiling.

Research Intern Nov 2019 - Jul 2020

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

- Designed, developed, and integrated a scalable data discovery system that worked within the company's data governance pipeline, assisting with providing recommendations about related data assets.

Teaching Assistant Sep 2019 - Apr 2020

Delft University of Technology, Department of Software Technology *Delft, Netherlands*

- Assisting in the labs of the course Big Data Processing in TU Delft's bachelor. The labs had programming assignments in Bash, functional programming with Scala, Apache Spark, and Apache Flink.
- Designed and implemented part of the lab assignments of the MOOC "Taming Big Data Streams: Real-time Data Processing at Scale" which included a complete introduction to Apache Flink.

Education

Ph.D. in Database Systems Jan 2021 - Oct 2025

Delft University of Technology *Delft, Netherlands*

Topic: Working at the intersection of stream processing and serverless computing, aiming to create a novel transactional dataflow system with a programming model that allows users to write complex, distributed stateful serverless functions (SFaaS) free of concurrency or machine failure considerations.

Title: Democratizing Scalable Cloud Applications: Transactional Stateful Functions on Streaming Dataflows

Advisor: Asterios Katsifodimos

M.Sc. in Computer Science Sep 2018 - Dec 2020

Delft University of Technology *Delft, Netherlands*

Thesis: "Holistic Schema Matching at Scale"

Advisor: Asterios Katsifodimos

Diploma (M.Eng.) in Electrical and Computer Engineering Sep 2012 - Dec 2017

Technical University of Crete *Chania, Greece*

Thesis: "Outlier Detection Using Spark Streaming"

Advisor: Antonios Deligiannakis

Technical Skills

- Programming Languages:
  - Expert: Python
  - Proficient: C/C++, Java
  - Familiar: Scala, Go, Rust, Javascript
- Big Data Processing Frameworks: Apache Flink, Apache Storm, Apache Spark, Apache Hadoop.
- Tools: Docker, Kubernetes, GIT.

Software

- **Styx: Transactional Stateful 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: Led the system implementation and publication.
- **Stateflow – Object Oriented Code to Distributed Stateful Dataflows**
  - StateFlow is a framework that compiles object-oriented Python code to distributed stateful dataflows. These dataflows can be executed on different target systems (i.e., PyFlink, Apache Beam, Flink Statefun, AWS Lambda, and Stateflow engine).
  - Contributions: Guided and assisted in the implementation of the compiler. Created the custom Stateflow engine, improving the compiled code performance by an order of magnitude.
- **Valentine: (Schema-) Matching DataFrames Made Easy**
  - A Python package for capturing potential relationships among columns of different tabular datasets, which are given in the form of pandas DataFrames.
  - Contributions: Led the design and development of the package and currently being the primary maintainer.
- **Topio Market Profiler**
  - A data profiling service that scrutinizes the data assets and calculates statistics or informative summaries over them, thus offering extensive and objective indicators for assessing data assets in terms of business value, fitness for purpose, and quality.
  - Contributions: Implementation of the service in Flask and maintained GeoVaex and BigDataVoyant, the two Python packages of the project required for large scale geospatial data profiling.

Awards

- 2023 EDBT Best Demonstration Award for "Topio Marketplace: Search and Discovery of Geospatial Data"
- 2021 ACM DEBS Best Paper Award for "Distributed Transactions on Serverless Stateful Functions"

Teaching

Delft University of technology Sep 2019 - Oct 2025

Faculty of Electrical Engineering, Mathematics and Computer Science *Delft, Netherlands*

- Spring 2020 - 2025: IN4331: Web-scale Data Management, Head Teaching Assistant, Master's level.
- Spring 2020 - 2021: Taming Big Data Streams: Real-time Data Processing at Scale, Head Teaching Assistant, MOOC.
- Fall 2019: CSE2520: Big Data Processing, Teaching Assistant, Bachelor level.

Languages

Greek (native speaker), English (full professional proficiency), German (elementary proficiency).

Master Student Theses Supervision

- Smruti Kshirsagar ('25) "Global-State Querying in Stream Processing using Snapshots (H-Styx)"
- Mitali Patil ('25) "Live Global State Queries in Stream Processing"
- Theodoros Veneti ('23) "Online state migration in modern stream processing engines"
- Wouter van Lil ('23) "Consistency in Stateful FaaS Platforms"
- Marcus Schutte ('23) "Minimizing aborts in an epoch based transaction protocol for deterministic databases"
- Nikos Gavalas ('23) "Incremental Snapshotting in Transactional Dataflow SFaaS Systems"
- Martijn Comans ('23) "Benchmarking Stateful Serverless Functions"
- Konstantinos Chronas ('23) "Generating Labeled Datasets for Schema Matching"
- Alex Walker ('22) "Distributed Dataflow Transactions"
- Wouter Zorgdrager ('21) "Anyone Can Cloud: Democratizing Cloud Application Programming"
- Martijn de Heus ('21) "Distributed Transactions on SFaaS using Coordinator Functions"

Service

Reviewer

- **ACM WWW** (2025)
- **Springer Cluster Computing Journal** (2025)

External Reviewer

- **ACM SIGMOD** (2021, 2022, 2023)
- **VLDB** (2021, 2022, 2023)
- **IEEE ICDE** (2021, 2022, 2023, 2024)
- **CIKM** (2024)

Publications

International Journals

- **Transactions across serverless functions leveraging stateful dataflows**
  M. de Heus, **K. Psarakis**, M. Fragkoulis, A. Katsifodimos.
  In the Elsevier Information Systems Journal. **(CORE A*)**

International Conferences

- **Event Horizon: Asymmetric Dependencies for Fast Geo-Distributed Operations**
  J. Arns, H. Ng, **K. Psarakis**, A. Katsifodimos, P. Carbone.
  Conference on Innovative Data Systems Research (CIDR) 2026 **(CORE A)**
- **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) **(CORE A*)**
- **Styx: Transactional Stateful Functions on Streaming Dataflows**
  K. Psarakis, G. Christodoulou, G. Siachamis, M. Fragkoulis, A. Katsifodimos.
  ACM Special Interest Group on Management of Data (SIGMOD) 2025 **(CORE A*)**
- **Transactional Cloud Applications: Status Quo, Challenges, and Opportunities**
  R. Laigner, G. Christodoulou, **K. Psarakis**, A. Katsifodimos, Y. Zhou.
  ACM Special Interest Group on Management of Data (SIGMOD) 2025 (Tutorial) **(CORE A*)**

- **Transactional Cloud Applications Go with the (Data)Flow**
  K. Psarakis, G. Christodoulou, M. Fargkoulis, A. Katsifodimos.
  Conference on Innovative Data Systems Research (CIDR) 2025 **(CORE A)**
- **Stateful Entities: Object-oriented Cloud Applications as Distributed Dataflows**
  K. Psarakis, W. Zorgdrager, M. Fragkoulis, G. Salvaneschi, A. Katsifodimos.
  International Conference on Extending Database Technology (EDBT) 2024 (Vision Paper) **(CORE A)**
- **CheckMate: Evaluating Checkpointing Protocols for Streaming Dataflows**
  G. Siachamis, **K. Psarakis**, M. Fragkoulis, A. van Deursen, P. Carbone, A. Katsifodimos.
  IEEE International Conference on Data Engineering (ICDE) 2024 **(CORE A*)**
- **Evaluating Stream Processing Autoscalers**
  G. Siachamis, G. Christodoulou, **K. Psarakis**, M. Fragkoulis, A. van Deursen, A. Katsifodimos.
  ACM Conference on Distributed and Event-Based Systems (DEBS) 2024
- **Adaptive Distributed Streaming Similarity Joins**
  G. Siachamis, **K. Psarakis**, M. Fragkoulis, O. Papapetrou, A. van Deursen, A. Katsifodimos.
  ACM Conference on Distributed and Event-Based Systems (DEBS) 2023
- **Topio: an Open-Source Web Platform for Trading Geospatial Data**
  A. Ionescu, K. Patroumpas, **K. Psarakis**, G. Chatzigeorgakidis, D. Collarana, K. Barenscher,
  D. Skoutas, A. Katsifodimos, S. Athanasiou.
  International Conference on Web Engineering (ICWE) 2023
- **Topio Marketplace: Search and Discovery of Geospatial Data**
  A. Ionescu, A. Alexandridou, L. Ikonomou, **K. Psarakis**, K. Patroumpas, G. Chatzigeorgakidis,
  D. Skoutas, S. Athanasiou, R. Hai, A. Katsifodimos.
  International Conference on Extending Database Technology (EDBT) 2023 (Demo Track) **(CORE A)**
- **Stateful Entities: Object-oriented Cloud Applications as Distributed Dataflows**
  K. Psarakis, W. Zorgdrager, M. Fargkoulis, G. Salvaneschi, A. Katsifodimos.
  Conference on Innovative Data Systems Research (CIDR) 2023 (Abstract) **(CORE A)**
- **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 (Demo Track) **(CORE A*)**
- **Distributed Transactions on Serverless Stateful Functions**
  M. de Heus, **K. Psarakis**, M. Fragkoulis, A. Katsifodimos.
  ACM Conference on Distributed and Event-Based Systems (DEBS) 2021
- **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 **(CORE A*)**

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
- **SiMa: Federating Data Silos using GNNs**
  C. Koutras, R. Hai, **K. Psarakis**, M. Fragkoulis, A. Katsifodimos.
  Table Representation Learning workshop (NeurIPS) 2022