Kyriakos Psarakis

kPsarakis | kPsarakis | kpsarakis.com | k.psarakis@tudelft.nl

Education

Ph.D. Candidate in Database Systems Jan 2021 - present

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.
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
GPA: 8.1/10

Diploma (M.Eng.) in Electrical and Computer Engineering Sep 2012 - Feb 2018

Technical University of Crete *Chania, Greece*

Thesis: "Outlier Detection Using Spark Streaming"

Advisor: Antonios Deligiannakis
GPA: 8.11/10

Employment

Researh Intern Jun 2023 - Sept 2023

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

- Implementing query optimization techniques within an MPP database.

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.

Researh Intern Nov 2019 - Jul 2020

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

- Design, development, and integration of a scalable schema matching system as an MVP 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.

Teaching

Delft University of technology Sep 2019 - present

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

- Spring 2020 - present: IN4331: Web-scale Data Management, Head Teaching Assistant, Master 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.

Student Supervision

Master Theses Supervision

- 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"

Awards

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

Languages

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

Software

- **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.

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, special issue on the DEBS 2020 and 2021 Conferences.

International Conferences

- **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)
- **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)
- **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)
- **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)
- **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

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

Technical Skills

- Programming Languages: Python (advanced), Java/Scala (intermediate), C/C++ (basic), JavaScript (basic).
- Big Data Processing Frameworks: Apache Flink, Apache Storm, Apache Spark, Apache Hadoop.
- Tools: Docker, Kubernetes, GIT.