Victor Giannakouris

giannakouris.victor@gmail.com • (+1) 9179456484 • https://gsvic.github.io
• Google Scholar. Citations: 145, h-index: 7

INTERESTS

I am interested in the intersection of LLMs and data systems, and database system internals. I have research and industry experience on implementing database system components, ranging from storage (TileDB, MonetDB) to federated query optimizers for in-memory and distributed query engines (Spark SQL, DataFusion, DuckDB, Redshift, Presto). During my PhD, I have built **Dingo** (learned federated query optimizer), **λ-Tune** (exploiting LLMs for database system tuning), **DBG-PT** (LLM-based query plan regression debugging) and **SwellDB** (a database that generates its tables on-the-fly).

EDUCATION

Cornell University, Ithaca, NY, USA

- Ph.D., Computer Science — 2021 – 2026
  - Thesis: Rethinking Data System Architectures for the Generative AI Era
  - Advisor: Immanuel Trummer

National and Kapodistrian University of Athens, Athens, Greece

- M.Sc., Computing Systems — Oct 2018 – Jul 2021
  - Thesis: An Online Sample Index for Approximate Query Processing
  - Advisor: Yannis Ioannidis

Ionian University, Department of Informatics, Corfu, Greece

- B.Sc., Informatics — Sep 2012- May 2017
  - Thesis: Optimization of Distributed SQL Query Execution Over Multiple-Engine Environments
  - Advisor: Dimitrios Tsoumakos

EMPLOYMENT

Cornell University, Ithaca, NY, USA

- Graduate Research Assistant — Aug 2021 – Present
  - TA: CS 6230 - Parallel Architectures, Algorithms, and Applications (Spring 2025)
  - TA: CS 4321 - Database Systems Implementation (Fall 2022, Fall 2023, Fall 2024)
  - TA: CS 4410 - Introduction to Operating Systems (Fall 2021, Spring 2022)

Microsoft Research, Mountain View, CA, US

- Research Intern, Gray Systems Lab (GSL) — May 2025 – Aug 2025
  - Optimization of Queries with LLM Invocations

Microsoft Research, Redmond, WA, US

- Research Intern, Data Systems Group — May 2023 – Aug 2023
  - Query Optimization and Index Tuning

IBM Research, Almaden, CA, US

- Research Intern — May 2022 – Aug 2022
  - Database Internals (Spark SQL, Dremio, DB2)

TileDB, Boston, MA, US

- Senior Software Engineer — Nov 2019 – Jul 2021
  - Integration of the TileDB storage engine with distributed query engines
  - Spark SQL, Presto/Trino, Hail
  - Zero-copy interfaces using Arrow and JNI.

Unravel Data Systems, Palo Alto, CA, US

- Software Engineer — Oct 2018 – Oct 2019
  - Big Data Application Performance Management (APM)

AISERA, Palo Alto, CA, US

- Data Engineer — Apr 2018 – Sep 2018
  - Development of data collection microservices on the company's platform's data ingestion component.
  - Tools: Kubernetes, AWS, Google Cloud Platform, Java, Python

BEAT, Athens

- Big Data Software Engineer — Sep 2017 – Apr 2018
  - Development of streaming and ETL jobs over large-scale data.
  - Tuning and optimization of Spark SQL batch jobs on query and storage level.
  - Tools: Spark, YARN, Hadoop, Hive, Presto, Scala, Python

Computing Systems Laboratory (CSLab), NTUA, Greece

- Researcher — Jul 2015 – Aug 2017

- R&D on Big Data Management, Databases and Multi-Engine Analytics.
- Participated in four publications.
- Development of a federated query optimizer for Spark SQL, achieving execution time improvement for queries over heterogeneous environments up to one order of magnitude.
- Participated in the EU Funded Project ASAP(A Scalable Analytics Platform).

isMOOD Data Technology Services

- Software Engineer — Apr 2014 – Sep 2015
  - Development of web crawlers
  - Tools: Scrapy, Python, Elasticsearch, MongoDB

SERVICE

Reviewer

- ACM Transactions on Database Systems (TODS), 2025
- SoftwareX, 2024

AWARDS

Gerondelis Foundation Scholarship

- Stipend for students studying abroad. — May 2023

Ignite Fellowship for New Ventures

- Awarded the 1 year Cornell's Ignite for New Ventures Fellowship to spin off SwellDB. — May 2025

PUBLICATIONS

CONFERENCES

[1] V. Giannakouris, I. Trummer, "SwellDB: GenAI-Native Query Processing via On-the-Fly Table Generation *In VLDB 2025, PhD Workshop.* Jun 2025.

[2] V. Giannakouris, I. Trummer, "Rethinking Pluggable Federated Query Optimization: From Laptops to Data Warehouses *In VLDB 2025, Composable Data Systems (CDMS) Workshop.* Jun 2025.

[3] V. Giannakouris, I. Trummer, "SwellDB: Dynamic Query-Driven Table Generation with Large Language Models", *In proceedings of the 2025 ACM SIGMOD/PODS, International Conference on Management of Data*, Berlin, Germany, May 2025.

[4] V. Giannakouris, I. Trummer, "λ-Tune: Harnessing Large Language Models for Automated Database System Tuning", *In proceedings of the 2025 ACM SIGMOD/PODS, International Conference on Management of Data*, Berlin, Germany, Apr 2025.

[5] V. Giannakouris, I. Trummer, "DBG-PT: A Large Language Model Assisted Query Performance Regression Debugger", *In proceedings of the VLDB 2024 Endowment*, Guangzhou, China, Aug 2024.

[6] V. Giannakouris, I. Trummer, "Demonstrating λ-Tune: Exploiting Large Language Models for Workload-Adaptive Database System Tuning", *In proceedings of the 2024 ACM SIGMOD/PODS, International Conference on Management of Data*, Santiago, Chile, Jun 2024.

[7] V. Giannakouris, I. Trummer, "Building learned federated query optimizers", *In proceedings of the VLDB 2022 Endowment (PhD Workshop)*, Sydney, Australia, Sep 2022.

[8] V. Giannakouris, A. Fernandez, A. Simitsis,and S. Babu, "Cost-Effective, Workload-Adaptive Migration of Big Data Applications to the Cloud," *In proceedings of the 2019 ACM SIGMOD/PODS, International Conference on Management of Data*, Amsterdam, The Netherlands, Jun 2019.

[9] V. Giannakouris, N. Papailiou, D. Tsoumakos,and N. Koziris, "MuSQLE: Distributed SQL Query Execution Over Multiple-Engine Environments," *In proceedings of the 2016 IEEE International Conference on Big Data (BigData 2016)*, Washington, DC, USA, Dec 2016.

[10] K. Doka, N. Papailiou, V. Giannakouris, D. Tsoumakos and N. Koziris, "Mix 'n Match Multi-Engine Analytics,," *In proceedings of the 2016 IEEE International Conference on Big Data (BigData 2016)*, Washington, DC, USA, Dec 2016.

[11] K. Doka, M. Filatov, V. Giannakouris, V. Kantere, N. Koziris, C. Mantas, N. Papailiou, V. Papaioannou and D. Tsoumakos, "Optimizing, Planning and Executing Analytics Workflows over Multiple Engines.," *In the 1st International Workshop on Multi-Engine Data AnaLytics (MEDAL 2016), in conjuction with the EDBT Conference*, Bordeaux, France, Dec 2016.

[12] N .Papailiou, K. Doka, V. Giannakouris, V. Papaioannou, D. Tsoumakos and N. Koziris, "Robust and Adaptive Multi-Engine Analytics using IReS (demo paper)," *In the 10th International Workshop on Enabling Real-Time Business Intelligence (BIRTE 2016), in conjunction with the VLDB Conference*, New Delhi, India, Mar 2016.

[13] V. Giannakouris, A. Plerou, and S. Sioutas, "CSMR: A Scalable Algorithm for Text Clustering with Cosine Similarity and MapReduce," *In the 10th International Conference on Artificial Intelligence Applications and Innovations*, Rhodes, Greece, Sep 2014.

JOURNAL PUBLICATIONS

[1] K. Doka, I. Mytilinis, N .Papailiou, V. Giannakouris, D. Tsoumakos and N. Koziris, "Multi-Engine Analytics with IReS," *IIn: Castellanos M., Chrysanthis P., Pelechrinis K. (eds) Real-Time Business Intelligence and Analytics. BIRTE 2015, BIRTE 2016, BIRTE 2017. Lecture Notes in Business Information Processing, vol 337. Springer, Cham*, New Delhi, India, Mar 2016.

[2] X. Evangelopoulos, V. Giannakouris, L. Iliadis, C. Makris, Y. Plegas, A. Plerou, S. Sioutas, "Evaluating Information Retrieval using Document Popularity: A case study using MapReduce," *Engineering Applications of Artificial Intelligence*, vol. 89, no. 2, pp. 4123–4133, Feb 2008.

LANGUAGES

- **English** (fluent), **Greek** (native)