

# IOANNIS MYTILINIS

✉ ioannis.mytilinis@oracle.com, gmytilinis@gmail.com

📍 Lausanne, Switzerland ☎ (+41) 0786212679 or (+30) 6984750374

## EDUCATION

---

### Doctor of Philosophy (PhD) in Computer Science

Dec 2012 - Nov 2019

National Technical University of Athens (NTUA), Heroon Polytechniou 9, 15780 Zografou, Greece

Title of Thesis: "Wavelet-based Algorithms for Approximate Processing in the Big Data Era"

Defence Date: 20/11/2019

Advisor: Prof. Nectarios Koziris

### Diploma (MEng) in Electrical and Computer Engineering

Sept 2006 - Feb 2012

National Technical University of Athens (NTUA), Heroon Polytechniou 9, 15780 Zografou, Greece

Title of Thesis: "Distributed Storage and Indexing of Social Network Information"

Advisor: Prof. Nectarios Koziris

## AWARDS

---

### Next-Generation Data Infrastructure Research Award by Facebook

2021

Proposal on "Runtime-optimized Analytics with Compilation Hints"

Collaboration with Prof. Anastasia Ailamaki

**Best Paper Runner up Award** in the 31st International Conference on Scientific and Statistical Database Management (SSDBM) 2019

**NTUA Thomaidion Award** for outstanding publications in scientific journal or conference proceedings 2018

**NTUA Thomaidion Award** for outstanding publications in scientific journal or conference proceedings 2016

## PUBLICATIONS

---

### 1. Theses

[T01] Mytilinis, I. (2019). Wavelet-based Algorithms for Approximate Processing in the Big Data Era.

[T02] Mytilinis, I. (2012). Distributed Storage and Indexing of Social Network Information.

### 2. Articles in Refereed International Journals

[J01] Mytilinis, I., Tsoumakos, D., Koziris, N. (2021). Workload-aware wavelet synopses for sliding window aggregates. Distributed and Parallel Databases (**DAPD**), 39(2), 445-482. (**Invited, Special Issue**)

[J02] Bakogiannis, T., Mytilinis, I., Doka, K., Goumas, G. (2020). Leveraging blockchain technology to break the cloud computing market monopoly. **Computers**, 9(1), 9.

[J03] Mytilinis, I., Tsoumakos, D., Koziris, N. (2018). Scaling the construction of wavelet synopses for maximum error metrics. IEEE Transactions on Knowledge and Data Engineering (**TKDE**), 31(9), 1794-1808.

[J04] Doka, K., Mytilinis, I., Papailiou, N., Giannakouris, V., Tsoumakos, D., Koziris, N. (2015). Multi-engine Analytics with IReS. In Real-Time Business Intelligence and Analytics (pp. 133-154). Springer, Cham.

### 3. Articles in Refereed International Conferences

[C01] Zapridou, E., Mytilinis, I., Ailamaki, A. Dalton: Learned Partitioning for Distributed Data Streams. Expected to appear in the proceedings of **VLDB 2023**.

[C02] Michas, G., Chrysogelos, P., Mytilinis, I., Ailamaki, A. (2021, June). Hardware-Conscious Sliding Window Aggregation on GPUs. In Proceedings of the 17th International Workshop on Data Management on New Hardware (**DaMoN 2021**) (pp. 1-5).

[C03] Konsta, A., Mytilinis, I., Doka, K., Niarchos, S., Koziris, N. (2021, April). Clouseau: Blockchain-based Data Integrity for HDFS Clusters. In 2021 IEEE 37th International Conference on Data Engineering (**ICDE**) (pp. 2725-2728). IEEE.

[C04] Sioulas, P., Sanca, V., Mytilinis, I., Ailamaki, A. (2021). Accelerating complex analytics using speculation. In **CIDR**.

[C05] Bian, H., Chandra, B., Mytilinis, I., Ailamaki, A. (2021). Storage Management in Smart Data Lake. In **EDBT/ICDT Workshops**.

- [C06] Kotselidis, C., Diamantopoulos, S., Akrivopoulos, O., Rosenfeld, V., Doka, K., Mohammed, H., ... Morgan, W. (2020, March). Efficient compilation and execution of JVM-based data processing frameworks on heterogeneous co-processors. In 2020 Design, Automation Test in Europe Conference Exhibition (**DATE**) (pp. 175-179). IEEE.
- [C07] Bakogiannis, T., Mytilinis, I., Doka, K., Goumas, G. (2019, October). Building Ad-Hoc Clouds with CloudAgora. In 2019 38th Symposium on Reliable Distributed Systems (**SRDS**) (pp. 360-3602). IEEE.
- [C08] Mytilinis, I., Tsoumakos, D., Koziris, N. (2019, July). Maintaining wavelet synopses for sliding-window aggregates. In Proceedings of the 31st International Conference on Scientific and Statistical Database Management (**SSDBM**) (pp. 73-84). (**Best Paper Runner up**)
- [C09] Doka, K., Bakogiannis, T., Mytilinis, I., Goumas, G. (2019, June). Cloudagora: Democratizing the cloud. In International Conference on Blockchain (pp. 142-156). Springer, Cham.
- [C10] Mytilinis, I., Bitsakos, C., Doka, K., Konstantinou, I., Koziris, N. (2018, December). The vision of a heterogeneous scheduler. In 2018 IEEE International Conference on Cloud Computing Technology and Science (**CloudCom**) (pp. 302-307). IEEE.
- [C11] Doka, K., Mytilinis, I., Giannakopoulos, I., Konstantinou, I., Tsitsigkos, D., Terrovitis, M., Koziris, N. (2017, October). Exploiting social networking and mobile data for crisis detection and management. In International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries (pp. 28-40). Springer, Cham.
- [C12] Mytilinis, I., Tsoumakos, D., Koziris, N. (2016, June). Distributed wavelet thresholding for maximum error metrics. In Proceedings of the 2016 ACM **SIGMOD** International Conference on Management of Data (pp. 663-677).
- [C13] Mytilinis, I., Giannakopoulos, I., Konstantinou, I., Doka, K., Tsitsigkos, D., Terrovitis, M., ... Koziris, N. (2015, May). Modissense: A distributed spatio-temporal and textual processing platform for social networking services. In Proceedings of the 2015 ACM **SIGMOD** International Conference on Management of Data (pp. 895-900).
- [C14] Mytilinis, I., Tsoumakos, D., Kantere, V., Nanos, A., Koziris, N. (2015, March). I/O performance modeling for big data applications over cloud infrastructures. In 2015 IEEE International Conference on Cloud Engineering (**IC2E**) (pp. 201-206). IEEE.
- [C15] Mytilinis, I., Giannakopoulos, I., Konstantinou, I., Doka, K., Koziris, N. (2014, October). MoDisSENSE: A distributed platform for social networking services over mobile devices. In 2014 IEEE International Conference on Big Data (**Big Data**) (pp. 49-51). IEEE.
- [C16] Konstantinou, I., Tsoumakos, D., Mytilinis, I., Koziris, N. (2013, June). DBalancer: distributed load balancing for NoSQL data-stores. In Proceedings of the 2013 ACM **SIGMOD** International Conference on Management of Data (pp. 1037-1040).

#### 4. Book Chapters

- [B01] Fumero, J., Kotselidis, C., Zakkak, F., Papadimitriou, M., Akrivopoulos, O., Tselios, C., ... Bitsakos, C. (2019). Programming and Architecture Models. In Heterogeneous Computing Architectures (pp. 53-87). CRC Press.

## TEACHING EXPERIENCE

---

### 1. Teaching

#### School of Computer and Communication Sciences, EPFL

##### Graduate Classes:

- |                             |      |
|-----------------------------|------|
| 1. Database Systems (CS422) | 2021 |
| 2. Database Systems (CS422) | 2020 |

#### School of Electrical and Computer Engineering, National Technical University of Athens

##### Teaching Assistant:

- |                                |           |
|--------------------------------|-----------|
| 1. Introduction to Programming | 2012      |
| 2. Operating Systems           | 2012-2017 |
| 3. Advanced Database Systems   | 2015-2019 |
| 4. Distributed Systems         | 2018-2019 |

### 2. Advising of Graduate and Undergraduate Students

#### - Co-advising PhD Students

- |   |           |
|---|-----------|
| 1. Eleni Zapridou, "Scalable Distributed Stream Processing", EPFL | 2020-2021 |
|---|-----------|

#### - MSc Students

1. Christina Mantonanaki, “Analyzing and Improving Database Statistics for In-Memory Analytical Query Processing”, EPFL 2021
2. Frédéric Gessler, “Cloud data-warehouses for SQL analytics, revolution or evolution?”, EPFL 2021
3. Richard Roubaty, “Query processing over heterogeneous data sources using the Olympe composition platform”, EPFL 2021
4. Belmpas Theofilos, “LSM Trees for HTAP Workloads”, EPFL 2021
5. Yifei Li, “Increasing Data Freshness in HTAP Engines Through Incremental Processing”, EPFL 2021

#### - Summer Interns

1. Evangelos Danias, “Optimizing control-flow in analytics using speculative execution”, EPFL 2021
2. Georgios Michas, “Scaling up window aggregation in JIT compiled engines”, EPFL 2020

#### - Co-advising Undergraduate Students

1. Alyzia Konsta, “A Verifiable Distributed Storage with the Use of Blockchain”, NTUA 2020
2. Georgios Kiritsas, “Compression Techniques for Large Scale Data”, NTUA 2019
3. Eirini Michelakaki, “Experimental Evaluation of Approximate Database Engines”, NTUA 2018

## PROFESSIONAL AND RESEARCH EXPERIENCE

---

### 1. Professional Positions

- Senior Member of Technical Staff at Oracle, Zurich Jan 2022 - present
- Postdoctoral Researcher at DIAS Lab, EPFL (Supervised by Prof. Anastasia Ailamaki) Jan 2020 - Dec 2021
- Researcher at CsLab, NTUA Dec 2012 - Nov 2019
- Software Engineer at Guppy Nov 2018 - Sept 2019
- Research Intern, IBM, T.J. Watson Research Center, NY Summer 2017
- Research Intern, IBM, T.J. Watson Research Center, NY Summer 2016

### 2. Funded Research Projects

[P01] **Batch Query Optimization - Collaboration with Huawei** 2020-2021

*Role in the project: Team leader/Researcher*

This project proposes a system that achieves high throughput and resource efficiency in multi-tenant infrastructures by sharing data and operators among different concurrent queries. The conducted work lies in three axes: (a) design and implementation of sharing-aware analytical operators for a well-known open-source analytical engine, (b) algorithms on multi-query optimization, and (c) locality-aware data placement and partitioning.

[P02] **Near Data Processing - Collaboration with Huawei** 2020-2021

*Role in the project: Team leader/Researcher*

Traditional compute-centric data processing systems suffer from excessive data movement costs. The recent trend of decoupling compute and storage in favor of scalability and resilience further exacerbates data transfer overheads. This project enables a data-centric architecture that adaptively offloads computation to storage nodes and elastically adjusts computational resources in order to eliminate funnel bottlenecks.

[P03] **SmartDataLake: Sustainable DataLakes for Extreme-Scale Analytics, H2020-825041** 2020-2021

*Role in the project: Team leader/Researcher/Developer*

The project empowers data scientists to extract insights from Big Data Lakes through interactive data visualizations. To achieve interactivity and reduced costs, SmartDataLake provides an adaptive, scalable and elastic data management system that offers: (a) data virtualization for abstracting and optimizing access over heterogeneous data, (b) data synopses that enable interactive approximate query processing, and (c) automated data placement in different storage tiers of diverse characteristics.

[P04] **E2Data: European Extreme Performing Big Data Stacks, H2020-780245** 2018-2019

*Role in the project: Researcher/Developer*

This project provides a new Big Data software paradigm for achieving maximum resource utilization for heterogeneous cloud deployments without requiring developers to change their code. The proposed solution takes a cross-layer

approach by allowing vertical communication between the four key layers of Big Data deployments: the application, the Big Data platform, the cloud provider, and the execution run time.

**[P05] Data Sovereignty through the use of Blockchain (E $\Delta$ BM34: Support for young researchers) 2018-2019**

*Role in the project: Researcher/Developer*

The project proposes the combination of Cloud Computing and Blockchain Technology into a platform that enables the on-demand creation of a fully distributed, scalable and secure virtual infrastructure for data storage and processing. One of the main targets of the project is to face the limitations of current blockchain implementations in terms of required storage, latency and throughput by adopting different architectural choices that will render blockchain usable in the context of Cloud Computing.

**[P06] ASAP: An Adaptive, highly Scalable Analytics Platform, FP7-619706 2014-2017**

*Role in the project: Researcher/Developer*

This project proposes a unified, open-source execution framework for scalable data analytics. Data analytics tools have become essential for harnessing the power of our era's data deluge. Current technologies are restrictive, as their efficacy is usually bound to a single data and compute model, often depending on proprietary systems. The main idea behind ASAP is that no single execution model is suitable for all types of tasks and no single data model (and store) is suitable for all types of data.

**[P07] CLARIN-EL: Common Language Resources and Technology Infrastructure 2016-2017**

*Role in the project: Researcher/Developer*

CLARIN-EL is the Greek counterpart of the CLARIN project, a pan-european effort for the collection and distribution of language resources (text/speech corpora, lexica, etc.) and processing tools (syntactic analyzers, parsers, taggers, etc.) through a web-based Research Infrastructure.

**[P08] CELAR: Automatic, multi-grained elasticity-provisioning for the Cloud, FP7-317790 2012-2015**

*Role in the project: Researcher/Developer*

The vision of the CELAR project is to provide automatic, multi-grained resource allocation for cloud applications. This enables the commitment of just the right amount of resources based on application demand, performance and requirements, results in optimal use of infrastructure resources and significant reductions in administrative costs.

**[P09] MoDisSENSE: A Distributed Platform for the Development of Social Networking Services over Mobile Devices, National Action "09SYN-72-881" 2012-2014**

*Role in the project: Researcher/Developer*

The project enriches social networking services by exploiting the continuous data flow from the daily use of mobile devices. By combining and analyzing the heterogeneous collected data (text, spatiotemporal data), it offers innovative services that rely on state-of-the art distributed data processing techniques.

## VARIOUS RESPONSIBILITIES IN INTERNATIONAL JOURNALS AND CONFERENCES

---

### Referee in International Journals

1. IEEE Transactions on Knowledge Discovery from Data (TKDD)
2. Information Systems

### PC Member in International Conferences

1. International Workshop on Data Analytics and Machine Learning Made Simple (Co-located with EDBT 2021)
2. Conference on Information and Knowledge Management (CIKM 2020)
3. IEEE International Conference on Machine Learning and Data Science (ICMLDS 2018)

### External Referee in International Conferences

1. IEEE International Conference on Big Data (2019)
2. Conference on Information and Knowledge Management (CIKM 2019)
3. IEEE/ACM International Symposium in Cluster, Cloud, and Grid Computing (CCGrid 2019)
4. Conference on Information and Knowledge Management (CIKM 2018)
5. IEEE International Conference on Big Data (2018)
6. IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2018)
7. International Conference on Smart Cities and Green ICT Systems (SMARTGREENS 2018)
8. IEEE International Conference on Data Engineering (ICDE 2017)

9. Conference on Information and Knowledge Management (CIKM 2017)
10. IEEE/ACM International Symposium in Cluster, Cloud, and Grid Computing (CCGrid 2017)
11. IEEE International Conference on Big Data (2017)
12. International Conference on Computer Communications and Networks (ICCCN 2017)
13. International Conference on Machine Learning and Data Science (ICMLDS 2017)
14. AlgoCloud (2017)
15. IEEE NCA (2017)
16. Algorithms and Systems for MapReduce and Beyond (BeyondMR 2017)
17. IEEE International Conference on Big Data (2016)
18. International Conference on Service Oriented Computed (RMSOC workshop 2015)
19. IEEE/ACM International Symposium in Cluster, Cloud, and Grid Computing (CCGrid 2015)