DR. OSCAR RICARDO MOLL

Curriculum Vitae

Cambridge, MA — orm@csail.mit.edu — oscar-moll.com — linkedin.com/in/oscarmoll —

EDUCATION

Doctor of Philosophy in Computer Science (PhD)   2015-2023
MIT Department of Electrical Engineering and Computer Science
Thesis Title: Efficiently Searching for Objects Within Large Collections of Images and Video.
Minor in entrepreneurship. Internships at Tesla and Google.
Blockcert: https://credentials.mit.edu/certificate/048c651fc6ae5dce8c9e6242be2a5121

Master of Engineering in Computer Science (MEng)   2010-2012
MIT Department of Electrical Engineering and Computer Science
Thesis Title: Database partitioning strategies for social network data

Bachelor of Science in Mathematics Computer Science (dual BSc)   2011
Massachusetts Institute of Technology (MIT)

EXPERIENCE

Machine Learning Engineer (PhD Intern)
Tesla Autopilot, Palo Alto, CA   **Jun 2017 - Aug 2017**
Researched and developed a data distribution change detection pipeline for Tesla Autopilot fleet telemetry (e.g., looking for sudden braking hotspots in maps which negatively affect driver experience). Aggregated the full Tesla fleet data in S3 and applied statistical tests for data distribution changes. The pipeline was implemented as an Apache Spark Streaming Query via PySpark. Additionally, we implemented an interactive web dashboard to visualize pipeline outputs overlayed on a map that also allowed users to dig into the anomalies found.

Software Engineer (PhD intern)
Google, Mountain View, CA   **Jun 2016 - Aug 2016**
Designed and implemented an efficient basicprogramming expression evaluator directly on top of protocol buffer data. By directly evaluating over serialized protocol buffers, it elides unnecessary overhead and memory allocations. By JIT compiling the expressions via LLVM, it further avoids dispatch, intermediate materialization and bounds checking overheads, allowing faster processing of larger amounts of data.

Software Development Engineer (Level 2)   **Sep 2014 - Jan 2015**
Software Development Engineer (Level 1)   **Sep 2012 - Sep 2014**
Amazon Web Services, Seattle, WA
Designed and implemented data path components for the distributed storage layer that underpins the Amazon Aurora service. Some of this joint design work was awarded multiple US patents.
Designed and implemented an anti-entropy protocol: The mechanism keeps replicas in sync despite transient failures and brings blank-slate replicas up-to-date.
Mentored a summer intern, interviewed engineering candidates, reviewed code

Software Engineer Intern
Twitter, San Francisco, CA   **Jun 2011 - Dec 2011**
Designed and evaluated alternative DB partitioning methods for social graph data (part of my MEng Thesis)

Developed distributed tracing of requests within Apache Cassandra nodes and integrated it with company monitoring tools.

PUBLICATIONS

SELECTED CONFERENCE PAPERS

1. **Oscar Moll**, Manuel Favela, Samuel Madden, Vijay Gadepally, and Michael Cafarella. Seesaw: Interactive ad-hoc search over image databases. *Proc. ACM Manag. Data*, 1(4), dec 2023. `https://dl.acm.org/doi/pdf/10.1145/3626754`
2. **Oscar Moll**, Favyen Bastani, Sam Madden, Mike Stonebraker, Vijay Gadepally, and Tim Kraska. Exsample: Efficient searches on video repositories through adaptive sampling. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 2956–2968, 2022. `https://arxiv.org/abs/2005.09141`

CONFERENCE PAPERS

3. Holger Pirk, **Oscar Moll**, Matei Zaharia, and Sam Madden. Voodoo : A vector algebra for portable database performance on modern hardware. *PVLDB*, 9(14):1707–1718, 2016
4. Yehuda Afek, Alexander Matveev, **Oscar R. Moll**, and Nir Shavit. Amalgamated lock-elision. In Yoram Moses, editor, *Distributed Computing - 29th International Symposium, DISC 2015, Tokyo, Japan, October 7-9, 2015, Proceedings*, volume 9363 of *Lecture Notes in Computer Science*, pages 309–324. Springer, 2015
5. Favyen Bastani, **Oscar Moll**, and Sam Madden. Vaas: video analytics at scale. *Proceedings VLDB Endowment*, 13(12):2877–2880, August 2020

WORKSHOP/DEMO PAPERS

6. **Oscar Moll**, Aaron Zalewski, Sudeep Pillai, Samuel Madden, Michael Stonebraker, and Vijay Gadepally. Exploring big volume sensor data with vroom. *PVLDB (Demo paper)*, 10(12):1973–1976, 2017
7. **Oscar Moll**, Sam Madden, and Vijay Gadepally. Ad-hoc searches on image databases. In *Heterogeneous Data Management, Polystores, and Analytics for Healthcare*, pages 3–9. Springer Nature Switzerland, 2022
8. Holger Pirk, **Oscar Moll**, and Sam Madden. What makes a good physical plan?: Experiencing hardware-conscious query optimization with candomblé. In Fatma Özcan, Georgia Koutrika, and Sam Madden, editors, *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 2149–2152. ACM, 2016

PATENTS

9. Yan Valerie Leshinsky, James Mcclellan Corey, Samuel James McKelvie, Oscar Ricardo Moll Thomae, and Pradeep Jnana Madhavarapu. Efficient garbage collection for a log-structured data storage system, October 2019. US Patent and Trademark Office
10. Samuel James McKelvie, Benjamin Tobler, James Mcclellan Corey, Pradeep Jnana Madhavarapu, Oscar Ricardo Moll Thomae, Christopher Richard Newcombe, Yan Valerie Leshinsky, and Anurag Windlass Gupta. Individual write quorums for a log-structured distributed storage system, March 2019. US Patent and Trademark Office

AWARDS AND HONORS

Guatemaltecos Ilustres Orator - 2021
Seguros Universales, S.A.
Guatemala

Greylock X - Summer 2017
Greylock Partners
San Francisco, CA

Edwin Webster Fellowship - Spring 2015
MIT EECS
Cambridge, MA

TALKS

Efficiently Searching for Objects Within Large Collections of Images and Video
Cambridge, MA - May 2023

Thesis Defense. MIT Dept of EECS

SeeSaw: Interactive Ad-Hoc Image Searches Over Image Databases
Palo Alto, CA - May 2022
Stanford and University of Washington Workshop on Video Analytics

ExSample: Efficient Searches on Video Repositories through Adaptive Sampling
Kuala Lumpur, Malaysia (remote) - May 2022
IEEE International Conference in Data Engineering (ICDE)

The Case for Learned Sampling in Video Datasets
Cambridge, MA - January 2019
North East Database Day (NEDB)

Exploring Big Volume Sensor Data with Vroom
Munich, Germany - September 2017
International Conference on Very Large Data Bases (VLDB)

SERVICE TO PROFESSION

Reviewer
International Conference on Very Large Databases (VLDB) 2024

Reviewer
AI City Challenge 2021 CVPR Workshop

TEACHING EXPERIENCE

Database Systems (Fall 2015)
MIT

Mathematics for Computer Science (Spring 2011)
MIT

REFERENCES

Samuel Madden
MIT College of Computing Distinguished Professor of Computing
MIT CSAIL
madden@csail.mit.edu
`https://db.csail.mit.edu/madden/`

Pierre Brunelle
CEO, Co-Founder
Pixeltable
pbrunelle@pixeltable.com