Matthew Saltz

saltzm@gmail.com · +1 225 205 7903
https://saltzm.github.io

SUMMARY

Software engineer experienced in programming distributed systems, databases, and custom implementations of distributed machine learning algorithms.

WORK EXPERIENCE

EllieDB, New York, NY, *Founder & CTO (https://elliedb.com)* November 2023-Present

- Implemented an LSM tree with a fully async API, no internal threads, and no memory allocation after startup
- Implemented a generic key-value store benchmarking tool that works against an async API and accounts for coordinated omission. Used to benchmark RocksDB, WiredTiger, and the EllieDB prototype
- Implementing an embedded key-value store with a novel log-based design, a fully async API, no internal threads, and no memory allocation after startup, with a focus on high throughput/low latency point queries and predictable performance. Currently supports point queries and transactions with snapshot isolation

MongoDB, New York, NY, *Multiple Roles* March 2018-February 2023

- *Senior Software Engineer, Storage Execution* (August 2022-February 2023)
  - Integrated support for reading the MongoDB collection catalog at points in time in the past
- *Team Lead, Service Architecture* (February 2021-August 2022)
  - Managed a team of four engineers, with three direct reports
  - Led a project to allow running MongoDB behind an L4 load balancer
  - Investigated the viability of the use of C++ coroutines for use in the MongoDB server
- *Senior Software Engineer, Distributed Systems* (March 2019-February 2021)
  - Led a project to design and implement a unified API for task cancellation in the server
  - Led a project to ensure that data is always properly cleaned up after a data migration
- *Software Engineer 3, Distributed Systems* (March 2018-March 2019)
  - Designed/implemented a fault-tolerant transaction coordinator for distributed transactions
  - Led a project to move logic for splitting ranges of data from routers to servers

Vertica, Cambridge, MA, *Software Engineer, Machine Learning* August 2015-March 2018

- Helped lead the initial development of Vertica's in-database machine learning capabilities
- Responsible for the design and development in C++ of distributed machine learning algorithms in Vertica, including kmeans and data normalization
- Designed and implemented in C++ a resource-managed infrastructure for caching tables that led to a speedup of 1-2 orders of magnitude for our machine learning algorithms

PUBLICATIONS

- *A New Kernelized Associative Memory and Some of its Applications*, **M Saltz**, LA Belanche, *ECAI 2016*
- *Distributed Community Detection with the WCC Metric*, **M Saltz**, A Prat-Pérez, D Dominguez-Sal, *Proceedings of the 24th International Conference on World Wide Web, 1095-1100*
- *Dualiso: An Algorithm for Subgraph Pattern Matching on Very Large Labeled Graphs*, **M Saltz**, A Jain, A Kothari, A Fard, JA Miller, L Ramaswamy, *2014 IEEE International Congress on Big Data, 498-505*
- *A Distributed Vertex-Centric Approach for Pattern Matching in Massive Graphs*, Arash Fard, M Usman Nisar, Lakshmish Ramaswamy, John A Miller, **Matthew Saltz**, *2013 IEEE International Conference on Big Data*
- *Distributed Algorithms for Graph Pattern Matching*, Arash Fard, Usman Nisar, Lakshmish Ramaswamy, John A Miller, Matthew Saltz, *University of Georgia*
- *A Scalable Vertex-Centric Approach for Distributed Graph Pattern Matching*, Arash Fard, MU Nisar, Lakshmish Ramaswamy, John A Miller, **Matthew Saltz**, *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*
- *A Fast Algorithm for Subgraph Pattern Matching on Large Labeled Graphs*, **MW Saltz**, J Miller, *University of Georgia*

RESEARCH EXPERIENCE

Master's Thesis Research, Universitat Politècnica de Catalunya February 2015-June 2015

· Kernelized the energy function of the bidirectional associative memory (BAM) and demonstrated that treating inference as an optimization problem on the kernelized energy function improves recall with more patterns and higher amounts of noise

Data Management (DAMA) Group, Universitat Politècnica de Catalunya July-August 2014

· Designed and developed in Java a distributed vertex-centric community detection algorithm based on the WCC metric using Apache Giraph

Master's Thesis Research, University of Georgia May 2013-August 2013

· Created and implemented in Scala a novel algorithm for subgraph isomorphism that outperformed two well-known algorithms by up to several orders of magnitude on synthetic graphs of up to 10M nodes and 250M edges and on two real life datasets

Large Scale Distributed Information Systems Lab, University of Georgia August 2012-August 2013

· Implemented in Java and Scala centralized sequential and parallel versions of strong simulation, an algorithm that uses various topological criteria to find matches of a query graph in a data graph

EDUCATION

Erasmus Mundus Master of Data Mining and Knowledge Management Sept 2013-June 2015

· **Université Lumière Lyon 2**, Lyon, France, *Data Mining & Complex System Modeling in Social Science* 2013-2014

· **Universitat Politècnica de Catalunya**, Barcelona, Spain, *Statistical Modeling & Data Mining* 2014-2015

Awarded the Category A scholarship of Erasmus Mundus, an organization sponsored by the E.U. to encourage international cooperation in a variety of cross-university master's programs. Includes full tuition, living stipend, personal travel budget

University of Georgia, B.S., M.S. in Computer Science, *GPA: 4.0/4.0 (B.S.), 3.96/4.0 (M.S.)* 2009-2013

Oxford University, Course: Modernist Literature, *GPA: 4.0/4.0* May-June 2010

Honors & Awards: Foundation Fellowship (1 of 11, University of Georgia's highest academic scholarship), First Honor Graduate (1 of 14 students graduating with a 4.0), Outstanding Undergraduate Student Award for Computer Science (only recipient)