

Matthew Malensek

Assistant Professor
Department of Computer Science · University of San Francisco

☎ (415) 720-3780 | ✉ mmalensek@usfca.edu | 🏠 www.cs.usfca.edu/~mmalensek

Research Interests

My research falls broadly within the area of data science, with a focus on performing analytics at scale, accounting for the storage and management of voluminous data streams, and transparently managing federation across cloud and fog domains.

Professional Experience

Assistant Professor

University of San Francisco

San Francisco, California

Aug 2017 — Present

Graduate Research Assistant

Colorado State University

Fort Collins, Colorado

Jan 2011 — Aug 2017

Graduate Systems Administrator

Colorado State University

Fort Collins, Colorado

Jan 2010 — Jan 2011

Education

Colorado State University

Ph.D., Computer Science

Fort Collins, Colorado

Summer 2017

Colorado State University

M.S., Computer Science

Fort Collins, Colorado

Fall 2012

Colorado State University

B.S., Applied Computing Technology

Minor: Business Administration

Fort Collins, Colorado

Fall 2009

Awards & Honors

- | | | |
|------|--|-------------------------|
| 2018 | IEEE TCSC Outstanding Ph.D. Dissertation Award ,
Presented by the IEEE Technical Committee on Scalable Computing | <i>Guangzhou, China</i> |
| 2014 | Best Paper Award , IEEE/ACM Symposium on Big Data Computing
(22% acceptance rate) with Jared Koontz and Sangmi Lee Pallickara | <i>London, UK</i> |
| 2012 | Best Paper Award , IEEE/ACM Conference on Utility and Cloud Computing
(27% acceptance rate) with Sangmi Lee Pallickara and Shrideep Pallickara | <i>Chicago, IL</i> |

Publications

Publications are listed in reverse chronological order. Preprint versions in PDF format are available online at: <https://www.cs.usfca.edu/~mmalensek/publications>

Journal Articles

- [J14] Matthew Malensek, Walid Budgaga, Ryan Stern, Shrideep Pallickara, and Sangmi Lee Pallickara. "Trident: Distributed Storage, Analysis, and Exploration of Multidimensional Phenomena". *IEEE Transactions on Big Data* (), (To Appear). [[PDF](#) · Impact Factor: 1.203].
- [J13] Naman Shah, Matthew Malensek, Harshil Shah, Shrideep Pallickara, and Sangmi Lee Pallickara. "Scalable Network Analytics for Characterization of Outbreak Influence in Voluminous Epidemiology Datasets". *Concurrency and Computation: Practice and Experience* (), (To Appear). DOI: 10.1002/cpe.4998. [[PDF](#) · Impact Factor: 0.942].
- [J12] Thilina Buddhika, Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Synopsis: A Distributed Sketch over Voluminous Spatiotemporal Observational Streams". *IEEE Transactions on Knowledge and Data Engineering* 29.11 (Nov. 2017), pp. 2552–2566. ISSN: 1041-4347. DOI: 10.1109/TKDE.2017.2734661. [[PDF](#) · Impact Factor: 3.438].
- [J11] Walid Budgaga, Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "A Framework for Scalable Real-Time Anomaly Detection over Voluminous, Geospatial Data Streams". *Concurrency and Computation: Practice and Experience* 29.12 (Mar. 2017). DOI: 10.1002/cpe.4106. [[PDF](#) · Impact Factor: 0.942].
- [J10] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Hermes: Federating Fog and Cloud Nodes to Support Query Evaluations in Continuous Sensing Environments". *IEEE Cloud Computing* 4.2 (Mar. 2017), pp. 54–62. DOI: 10.1109/MCC.2017.26. [[PDF](#) · Impact Factor: 1.86].
- [J9] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Fast, Ad Hoc Query Evaluations over Multidimensional Geospatial Datasets". *IEEE Transactions on Cloud Computing* 5.1 (Jan. 2017), pp. 28–42. ISSN: 2168-7161. DOI: 10.1109/TCC.2015.2398437. [[PDF](#) · Impact Factor: 3.77].
- [J8] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Analytic Queries over Geospatial Time-Series Data Using Distributed Hash Tables". *IEEE Transactions on Knowledge and Data Engineering* 28.6 (June 2016), pp. 1408–1422. ISSN: 1041-4347. DOI: 10.1109/TKDE.2016.2520475. [[PDF](#) · Impact Factor: 3.438].
- [J7] Cameron Tolooee, Matthew Malensek, and Sangmi Lee Pallickara. "A Scalable Framework for Continuous Query Evaluations over Multidimensional, Scientific Datasets". *Concurrency and Computation: Practice and Experience* 28.8 (June 2016), pp. 2546–2563. DOI: 10.1002/cpe.3651. [[PDF](#) · Impact Factor: 0.942].
- [J6] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Autonomous Cloud Federation for High-Throughput Queries over Voluminous Datasets". *IEEE Cloud Computing* 3.3 (May 2016), pp. 40–49. DOI: 10.1109/MCC.2016.65. [[PDF](#) · Impact Factor: 1.86].
- [J5] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Minerva: Proactive Disk Scheduling for QoS in Multitier, Multitenant Cloud Environments". *IEEE Internet Computing* 20.3 (May 2016), pp. 19–27. DOI: 10.1109/MIC.2016.48. [[PDF](#) · Impact Factor: 1.4].
- [J4] Walid Budgaga, Matthew Malensek, Sangmi Pallickara, Neil Harvey, F. Jay Breidt, and Shrideep Pallickara. "Predictive Analytics Using Statistical, Learning, and Ensemble Methods to Support Real-time Exploration of Discrete Event Simulations". *Future Generation Computer Systems* 56.C (Mar. 2016), pp. 360–374. ISSN: 0167-739X. DOI: 10.1016/j.future.2015.06.013. [[PDF](#) · Impact Factor: 2.43].
- [J3] Zhiqian Sui, Matthew Malensek, Neil Harvey, and Shrideep Pallickara. "Autonomous Orchestration of Distributed Discrete Event Simulations in the Presence of Resource Uncertainty". *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 10.3 (Sept. 2015), 18:1–18:20. ISSN: 1556-4665. DOI: 10.1145/2746345. [[PDF](#) · Impact Factor: 2.48].
- [J2] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Evaluating Geospatial Geometry and Proximity Queries Using Distributed Hash Tables". *IEEE Computing in Science Engineering (CiSE)* 16.4 (July 2014), pp. 53–61. ISSN: 1521-9615. DOI: 10.1109/MCSE.2014.48. [[PDF](#) · Impact Factor: 1.361].

- [J1] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Exploiting Geospatial and Chronological Characteristics in Data Streams to Enable Efficient Storage and Retrievals". *Future Generation Computer Systems* 29.4 (June 2013), pp. 1049–1061. ISSN: 0167-739X. DOI: 10.1016/j.future.2012.05.024. [[PDF](#) · Impact Factor: 2.43].

Book Chapters

- [B1] Sangmi Lee Pallickara, Matthew Malensek, and Shrideep Pallickara. "On the Processing of Extreme Scale Datasets in the Geosciences". *Handbook of Data Intensive Computing*. Ed. by Borko Furht and Armando Escalante. New York, NY: Springer New York, 2011, pp. 521–537. ISBN: 978-1-4614-1415-5. DOI: 10.1007/978-1-4614-1415-5_20.

Refereed Conference Proceedings

- [C11] Naman Shah, Harshil Shah, Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Network analysis for identifying and characterizing disease outbreak influence from voluminous epidemiology data". *Proceedings of the 2016 IEEE International Conference on Big Data*. Washington, D.C., USA, Dec. 2016, pp. 1222–1231. DOI: 10.1109/BigData.2016.7840726. [[PDF](#) · 18.68% Acceptance Rate].
- [C10] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Alleviation of Disk I/O Contention in Virtualized Settings for Data-Intensive Computing". *Proceedings of the 2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC)*. Limassol, Cyprus, Dec. 2015, pp. 1–10. DOI: 10.1109/BDC.2015.32. [[PDF](#) · 16% Acceptance Rate].
- [C9] Jared Koontz, Matthew Malensek, and Sangmi Lee Pallickara. "GeoLens: Enabling Interactive Visual Analytics over Large-Scale, Multidimensional Geospatial Datasets". *Proceedings of the 2014 IEEE/ACM International Symposium on Big Data Computing (BDC)*. London, UK, Dec. 2014, pp. 35–44. DOI: 10.1109/BDC.2014.12. [[PDF](#) · 22% Acceptance Rate · **Best Paper Award**].
- [C8] Matthew Malensek, Walid Budgaga, Sangmi Pallickara, Neil Harvey, F. Jay Breidt, and Shrideep Pallickara. "Using Distributed Analytics to Enable Real-Time Exploration of Discrete Event Simulations". *Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*. London, UK, Dec. 2014, pp. 49–58. DOI: 10.1109/UCC.2014.13. [[PDF](#) · 19% Acceptance Rate].
- [C7] Cameron Tolooee, Matthew Malensek, and Sangmi Lee Pallickara. "A Framework for Managing Continuous Query Evaluations over Voluminous, Multidimensional Datasets". *Proceedings of the 2014 IEEE International Cloud and Autonomic Computing Conference (ICAC)*. London, UK, Sept. 2014, pp. 73–82. DOI: 10.1109/ICAC.2014.25. [[PDF](#)].
- [C6] Matthew Malensek, Sangmi Pallickara, and Shrideep Pallickara. "Polygon-Based Query Evaluation over Geospatial Data Using Distributed Hash Tables". *Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing (UCC)*. Dresden, Germany, Dec. 2013, pp. 219–226. DOI: 10.1109/UCC.2013.46. [[PDF](#) · 24% Acceptance Rate].
- [C5] Matthew Malensek, Sangmi Pallickara, and Shrideep Pallickara. "Autonomously Improving Query Evaluations over Multidimensional Data in Distributed Hash Tables". *Proceedings of the 2013 ACM Cloud and Autonomic Computing Conference (CAC)*. Miami, Florida, USA, Sept. 2013, 15:1–15:10. DOI: 10.1145/2494621.2494638. [[PDF](#) · 35% Acceptance Rate].
- [C4] Matthew Malensek, Zhiqian Sui, Neil Harvey, and Shrideep Pallickara. "Autonomous, Failure-resilient Orchestration of Distributed Discrete Event Simulations". *Proceedings of the 2013 ACM Cloud and Autonomic Computing Conference (CAC)*. Miami, Florida, USA, Sept. 2013, 3:1–3:10. DOI: 10.1145/2494621.2494625. [[PDF](#) · 35% Acceptance Rate].
- [C3] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Expressive Query Support for Multidimensional Data in Distributed Hash Tables". *Proceedings of the 2012 IEEE/ACM 5th International Conference on Utility and Cloud Computing (UCC)*. Chicago, Illinois, USA, Nov. 2012, pp. 31–38. DOI: 10.1109/UCC.2012.41. [[PDF](#) · 27% Acceptance Rate · **Best Paper Award**].

- [C2] Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara. "Galileo: A Framework for Distributed Storage of High-Throughput Data Streams". *Proceedings of the 2011 IEEE/ACM 4th International Conference on Utility and Cloud Computing (UCC)*. Melbourne, Australia, Dec. 2011, pp. 17–24. DOI: 10.1109/UCC.2011.13. [[PDF](#) · 26.7% Acceptance Rate].
- [C1] Sangmi Lee Pallickara, Matthew Malensek, and Shrideep Pallickara. "Enabling access to timeseries, geospatial data for on-demand visualization". *IEEE Symposium on Large Data Analysis and Visualization, (LDAV)*. Providence, Rhode Island, USA, Oct. 2011, pp. 141–142. DOI: 10.1109/LDAV.2011.6092339. [[PDF](#)].

Ph.D. Dissertation and Master's Thesis

- [T2] Matthew Malensek. "Low-latency, Query-Driven Analytics over Voluminous Multidimensional, Spatiotemporal Datasets". PhD Dissertation. Colorado State University, July 2017. Advisors: Shrideep Pallickara and Sangmi Lee Pallickara. [[IEEE TCSC Outstanding Ph.D. Dissertation Award](#)].
- [T1] Matthew Malensek. "On the Evaluation of Exact-match and Range Queries over Multidimensional Data in Distributed Hash Tables". MS Thesis. Colorado State University, Aug. 2012. Advisor: Shrideep Pallickara.

Research Projects

My research has produced software in the areas of big data, cloud computing, and distributed systems. These projects are released as open source software under permissive licenses to encourage community involvement and collaboration. A complete overview of my projects is available at: <https://www.cs.usfca.edu/~mmalensek/lab>

Galileo

Observational datasets collected from radars, satellites, and other types of remote sensing devices are often composed of spatiotemporal records and several other *features* of interest. Galileo is a scalable distributed storage framework for managing these multidimensional datasets, with an emphasis on expressive queries for retrieval and analysis. The system supports petascale datasets and is designed to handle billions of small files comprising trillions of observations. To enable efficient distributed queries, Galileo features a hierarchical, multidimensional indexing scheme that generates a compact, in-memory *metamodel* of the dataset and its attributes. This allows low-latency approximate queries, geometry-constrained or proximity queries, and several types of analytic queries including outlier detection, hypothesis testing, and kernel density estimations of the underlying dataset.

<http://galileo.cs.colostate.edu>

Minerva

While cloud environments typically provide resource isolation between collocated virtual machines, certain usage patterns by *noisy neighbors* can have a negative impact on other VMs by monopolizing resources. I/O devices such as disks or network interfaces are particularly difficult to divide fairly among virtual machines, especially when the host hardware is overprovisioned. Minerva monitors the disk and network usage of virtual machines and allows enforcement of scheduling policies that promote *good neighbors*: clients that buffer disk writes, opt for sequential access patterns, or carry out I/O operations in short bursts. Combined with I/O forecasting to predict future resource usage, Minerva provides a substantial increase in overall throughput on busy VM host hardware.

<https://www.cs.usfca.edu/~mmalensek/projects/minerva>

Forager

Distributed execution engines often manage long-running, CPU-intensive computations. However, large quantities of short-term, stateless computations require a different set of considerations. Forager is part of the *Granules* cloud runtime, with a focus on scheduling and managing small tasks. The system employs *cycle scavenging* to locate resources such as idle workstations or servers, executes the computations, and collects the results. This functionality has been leveraged to orchestrate large-scale discrete event simulations across thousands of CPU cores. Forager supports fault tolerance, federation to the cloud, and state checkpointing and migration via an extended version of the *BSDiff* algorithm.

<http://granules.cs.colostate.edu>

Teaching

CS 326 · Operating Systems	Fall 2018 · Enrollment: 34
CS 677 · Big Data	Fall 2018 · Enrollment: 31
CS 220 · Introduction to Parallel Computing	Spring 2018 · Enrollment: 33
CS 326 · Operating Systems	Spring 2018 · Enrollment: 34
CS 220 · Introduction to Parallel Computing	Fall 2017 · Enrollment: 33
CS 686 · Special Topics in Big Data	Fall 2017 · Enrollment: 31

Professional & Academic Service

IEEE/ACM International Conference on Utility and Cloud Computing *Zurich, Switzerland*
Tutorials Co-chair *2018*

IEEE/ACM International Conference on Utility and Cloud Computing *Shanghai, China*
Tutorials Co-chair *2016*

IEEE/ACM International Conference on Utility and Cloud Computing *Dresden, Germany*
Session Chair *2013*

Upsilon Pi Epsilon, Colorado Beta Chapter *Fort Collins, Colorado*
Member, Officer *2012 – 2017*

- President (2016 – 2017)
- Vice President (2014 – 2015)
- Financial Officer (2013)

Journal Reviewer

Future Generation Computer Systems, IEEE Cloud Computing Magazine, IEEE Transactions on Cloud Computing, IEEE Transactions on Systems, Man, and Cybernetics: Systems, International Journal of Geographical Information Science