

Daniel Rammer

Ph.D. Student
Colorado State University

Email: daniel.rammer@protonmail.com
Phone (Cell): (920)-740-2445
GitHub: github.com/hamersaw
Website: blackpine.io

Research Interests

My research interests broadly envelop systems approaches for distributed storage and analytics at scale; with a focus on harnessing spatiotemporal data characteristics, accounting for memory contention and data distributions, and orchestrating deep-learning workloads.

Education

Colorado State University Ph.D. in Computer Science	Fort Collins, CO Summer 2021 (Expected)
<ul style="list-style-type: none">– Advisor: Dr. Shrideep Pallickara / Co-Advisor: Dr. Sangmi Lee Pallickara– Dissertation: Harnessing Spatiotemporal Data Characteristics to Facilitate Large-Scale Analytics over Voluminous, High-Dimensional Observational Datasets	

Colorado State University M.S. in Computer Science, GPA: 3.87/4.00	Fort Collins, CO Spring 2018
<ul style="list-style-type: none">– Thesis: Monitoring and Characterizing Application Service Availability	

University of Wisconsin - Oshkosh B.S. in Computer Science, Major GPA: 3.6/4.00	Oshkosh, WI Spring 2013
---	----------------------------

Experience

Colorado State University Graduate Research Assistant - Distributed Systems	Fort Collins, CO August 2017 - Present
---	---

Colorado State University Graduate Research Assistant - Network Security	Fort Collins, CO May 2015 - August 2017
--	--

VersiFit Technologies Solutions Developer - ETL for Educational Data Warehouse	Appleton, WI August 2012 - May 2015
--	--

Teaching

Graduate Teaching Assistant at Colorado State University <i>Distributed Systems (CS555)</i>	Fall 2019
---	-----------

Publications

Journal Articles

- [J1] **Daniel Rammer**, Thilina Buddhika, Matthew Malensek, Shrideep Pallickara, and Sangmi Lee Pallickara. Enabling Fast Exploratory Analyses Over Voluminous Spatiotemporal Data Using Analytical Engines. IEEE Transactions on Big Data. 2019. [5.16 Impact Factor, [PDF](#)]

Conference Proceedings

- [C6] Paahuni Khandelwal, **Daniel Rammer**, Shrideep Pallickara and Sangmi Pallickara. Mind the Gap: Generating Imputations for Satellite Data Collections at Myriad Spatiotemporal Scopes. (To Appear) IEEE International Symposium on Cluster Computing and the Grid (IEEE CCGrid). Melbourne, Australia. 2021. [26% Acceptance Rate, [PDF](#)]
- [C5] **Daniel Rammer**, Kevin Bruhwiler, Paahuni Khandelwal, Sam Armstrong, Shrideep Pallickara and Sangmi Pallickara. Small is Beautiful: Distributed Orchestration of Spatial Deep Learning Workloads. Proceedings of the 13th IEEE/ACM Conference on Utility and Cloud Computing. Leicester, UK. 2020. [31% Acceptance Rate, [PDF](#)]
- [C4] **Daniel Rammer**, Sangmi Lee Pallickara, and Shrideep Pallickara. Towards Timely, Resource-Efficient Analyses Through Spatially-Aware Constructs within Spark. Proceedings of the 13th IEEE/ACM Conference on Utility and Cloud Computing. Leicester, UK. 2020. [31% Acceptance Rate, [PDF](#)]
- [C3] Kevin Bruhwiler, Paahuni Khandelwal, **Daniel Rammer**, Samuel Armstrong, Sangmi Lee Pallickara, and Shrideep Pallickara. Lightweight, Embeddings Based Storage and Model Construction Over Satellite Data Collections. Proceedings of the IEEE International Conference on Big Data (IEEE BigData). Atlanta, USA. 2020. [15.5% Acceptance Rate, [PDF](#)]
- [C2] **Daniel Rammer**, Sangmi Lee Pallickara, and Shrideep Pallickara. ATLAS: A Distributed File System for Spatiotemporal Data. Proceedings of the 12th IEEE/ACM Conference on Utility and Cloud Computing. Auckland, NZ. 2019. [29.2% Acceptance Rate, [PDF](#)]
- [C1] **Daniel Rammer**, Walid Budgaga, Thilina Buddhika, Shrideep Pallickara, and Sangmi Lee Pallickara. Alleviating I/O Inefficiencies to Enable Effective Model Training Over Voluminous, High-Dimensional Datasets. Proceedings of the IEEE International Conference on Big Data (IEEE BigData). Seattle, USA. 2018. [18.9% Acceptance Rate, [PDF](#)]

Projects

STIP

Effective analyses of satellite imagery faces several intrinsic challenges stemming from the volume and variety of data. STIP (SpatioTemporal Image Partitioner) spatially partitions input images and coordinates storage and analytics within a DHT, facilitating distributed evaluations with data locality. Image partitioning is performed with geocodes, which support varying levels of precision and spatial reference systems. STIP reconciles differences in image format, pixel resolution, band types, and spatial reference systems among others. Additionally, it rectifies occlusions, which may be attributed to atmospheric phenomena or instrumentation failure, with preclude accurate analysis. [[Repository](#)]

Atlas

Observational dataset formats seldom align with spatiotemporal access patterns, introducing challenges in identification and retrieval of specific spatiotemporal extents. AtlasFS addresses these inefficiencies with two approaches. Blocked data is reordered so observations are spatiotemporally contiguous, facilitating targeted, sequential disk access. Additionally, informed data replication reconciles the competing pulls of dispersion and locality to mitigate analytical data movements. AtlasFS provides an HDFS-compliant interface with extensions for URL-embedded spatiotemporal filtering predicates, providing efficient, interoperable data space reduction operations. [[Repository](#)]

Effective spatiotemporal analytics must account for intrinsic data characteristics such as voluminous data and unique access patterns. AtlasSpark, an Apache Spark spatiotemporal extension, supports a rich collection of data wrangling functionality over native points, lines, and polygons including inspection, calculation, and transformation operations. AtlasSpark ensures performant analytics by injecting domain-specific rules into

Spark's logical and physical plans. It mitigates data movements by scheduling analytics with data locality, aggregates spatiotemporal filtering, and prunes unnecessary abstract syntax tree branches. [[Repository](#)]

Anamnesis

Memory access is several orders of magnitude faster than disk access. However, modern datasets quickly exceed the memory capacity of commodity clusters. Data sketches reduce dataset sizes by maintaining summary statistics and inter-feature relationships. Anamnesis is the first in-memory, sketch-aligned, HDFS-compliant distributed file system. It leverages data sketches to alleviate memory contention and reduce network I/O. Upon request, Anamnesis may generate full resolution data in myriad formats and presents an HDFS-compliant interface, achieving unprecedented compatibility with popular analytics engines. [[Repository](#)]

Proddle

Near real-time detection of Internet outages is a difficult problem, where individual events may be short-lived, localized, or service specific. Proddle is a global service monitoring framework used to detect such outages. It supports script-driven monitor definitions facilitating diverse functionality including incrementally refined monitor precision. Monitors are evaluated periodically on globally diverse vantages providing wide-spread outage coverage. Additionally, Proddle manages the orchestration of monitor data collection and archiving, providing well-defined API's for community retrieval of raw data. Accompanying is a web UI with live monitoring information, including event-driven alerts with the perceived outage cause. [[Repository](#)]

Awards

CS Graduate Fellowship presented by Computer Science Department <i>Colorado State University</i>	Spring 2021
Academic All-American presented by NCAA Division III Track & Field <i>University of Wisconsin - Oshkosh</i>	March 2010

Technical Skills

Programming Languages: Java, Rust, Python, C++, Go, Bash

Development Tools: Git, Gradle, Maven, Vim

Libraries: gRPC, Protobuf, FlatBuffers

Analytics Tools: Hadoop MapReduce, Apache Spark, PyTorch, Tensorflow, Keras

Storage Frameworks: HDFS, Apache Cassandra, MongoDB

Service Management Tools: Envoy Proxy, Docker