

JEFF LEFEVRE

Data Management in the Cloud

PROFESSIONAL SUMMARY

Software Engineer and Adjunct Professor of Computer Science and Engineering with over 20 years of experience in research, databases, and storage. My focus is on cloud data management, database physical design, and computational storage. I have released open source projects and commercial products, in addition to peer-reviewed papers and patents at various companies and institutions.

EMPLOYMENT HISTORY

ADJUNCT PROFESSOR OF COMPUTER SCIENCE AND ENGINEERING 2017 - Present
University of California, Santa Cruz *Santa Cruz, CA*

Fellow, Center for Research in Open Source Software. Faculty Mentor, Google Summer of Code (GSoC) [2019](#), [2020](#). Co-advised several PhD and MS students (ongoing). Led the Skyhook Data Management project, which embeds Apache Arrow directly within the object storage layer of Ceph, by extending read/write methods at the byte access level with database processing functions. The Skyhook project was successfully merged into Apache Arrow mainline in 2021 ([announcement](#)).

PROJECT SCIENTIST AND GUEST LECTURER 2016 - 2019
University of California, Santa Cruz *Santa Cruz, CA*

Taught database systems and operating systems courses including graduate research seminars. Led open-source project to extend Ceph object storage for database processing. Mentored undergraduate and graduate students.

SOFTWARE ENGINEER, VERTICA DATABASE 2014 - 2016
Hewlett Packard Enterprise *Sunnyvale, CA*

Integrated Vertica columnar database with external computation engines Apache Spark and Distributed R, which included fast parallel data import and export. Released beta version of product, available since Vertica 8.0. Produced 2 SIGMOD papers and 3 patents.

RESEARCH INTERN 2011 - 2013
NEC Labs *Cupertino, CA*

Conducted original research in the Data Management group. Worked with Hive, Impala, IBM DB2, Hadoop, HDFS, Parquet, and other data storage and processing engines. Produced 3 SIGMOD papers and 3 patents.

SOFTWARE ENGINEER INTERN 2007 - 2010
Google *Mountain View, CA*

Contributed to performance trending and analysis for Google storage, including enhancements to disk drive SMART data collection and analysis.

SOFTWARE ENGINEER INTERN 2005 - 2006
Teradata *San Diego, CA*

Created initial storage performance grading software used for data placement by Teradata Virtual Storage (TVS).

SYSTEM ANALYST, WORLDWIDE PROTEIN DATA BANK 2004 - 2005
San Diego Supercomputer Center *San Diego, CA*

Automated systems reliability testing for primary PDB servers and mirrors.

CO-FOUNDER 1998 - 2003
CitrusToGo.com *Tampa, FL*

Co-founded and managed www.CitrusToGo.com for online sales of Florida gift fruit.

EDUCATION

PHD, COMPUTER SCIENCE 2014
University of California-Santa Cruz

PHD Thesis: Physical Design Tuning Methods for Emerging System Architectures.

MS, COMPUTER SCIENCE 2008
University of California-San Diego

MS Thesis: Improving disk array performance and reliability.

BS Thesis: Encoding methods for DNA -based languages.

PUBLICATIONS

- ♦ H. Casaletto, A. Montana, J. LeFevre, P. Alvaro, "*A Transformation Embedded LSM Tree*", (under review) Jan 2025.
- ♦ A. Montana, Y. Xue, J. LeFevre, C. Maltzahn, J. Stuart, P. Kufeldt, P. Alvaro, "*A Moveable Beast: Partitioning Data and Compute for Computational Storage*", (pre-print) 2023.
- ♦ J. Chakraborty, I. Jimenez, S. A. Rodriguez, A. Uta, J. LeFevre, C. Maltzahn, "*Skybook: Towards an Arrow-Native Storage System*", CCGrid 2022 (pdf).
- ♦ S.A. Rodriguez, J. Chakraborty, A. Chu, I. Jimenez, J. LeFevre, C. Maltzahn, A. Uta "*Zero-Cost, Arrow-Enabled Data Interface for Apache Spark*", IEEE Big Data 2021.
- ♦ J. LeFevre, C. Maltzahn, "*Scaling Databases and File APIs with Programmable Ceph Object Storage*", Vault 2020.
- ♦ K. Dahlgren, J. LeFevre, A. Shirwadkar, K. Iizawa, A. Montana, P. Alvaro, C. Maltzahn, "*Towards Physical Design Management in Storage Systems*", IEEE/ACM Fourth International Parallel Data Systems Workshop (PDSW) 2019.
- ♦ J. LeFevre, N. Watkins, M. Sevilla, C. Maltzahn, "*Skybook: programmable storage for databases*", Vault 2019.
- ♦ M. Sevilla, R. Nasirigerdeh, C. Maltzahn, J. LeFevre, N. Watkins, P. Alvaro, M. Lawson, J. Lofstead, J. Pivarski, "*Tintenfish: File System Namespace Schemas and Generators*", Hot Storage 2018.
- ♦ M. Sevilla, N. Watkins, I. Jimenez, P. Alvaro, S. Finkelstein, J. LeFevre, C. Maltzahn, "*Malacology: A Programmable Storage System*", EuroSys 2017.
- ♦ J. LeFevre, R. Liu, C. Inigo, M. Castellanos, L. Paz, E. Ma, M. Hsu, "*Building the Enterprise Fabric for Big Data with Vertica and Spark*", SIGMOD 2016.
- ♦ S. Prasad, A. Fard, V. Gupta, J. Martinez, J. LeFevre, V. Xu, M. Hsu, I. Roy, "*Large-scale Predictive Analytics in Vertica: Fast Data Transfer, Distributed Model Creation, and In-database Prediction*", SIGMOD 2015.
- ♦ J. LeFevre, J. Sankaranarayanan, H. Hacigumus, J. Tatemura, N. Polyzotis, M.J. Carey, "*MISO: Souping Up Big Data Query Processing with a Multistore System*", SIGMOD 2014, pdf.
- ♦ J. LeFevre, J. Sankaranarayanan, H. Hacigumus, J. Tatemura, N. Polyzotis, M.J. Carey, "*Opportunistic Physical Design for Big Data Analytics*", SIGMOD 2014, pdf.
- ♦ J. LeFevre, J. Sankaranarayanan, H. Hacigumus, J. Tatemura, N. Polyzotis, "*Towards a Workload for Evolutionary Analytics*", 2nd Workshop on Data Analytics in the Cloud, (at SIGMOD 2013). (extended version).
- ♦ H. Hacigumus, J. Sankaranarayanan, J. Tatemura, J. LeFevre, N. Polyzotis, "*Odyssey: a multistore system for evolutionary analytics*", VLDB 2013.
- ♦ M.P. Consens, K. Ioannidou, J. LeFevre, N. Polyzotis, "*Divergent Physical Design Tuning for Replicated Databases*", SIGMOD 2012, pdf.
- ♦ I. Jimenez, J. LeFevre, N. Polyzotis, H. Sanchez, K. Schnaitter, "*Benchmarking Online Index-Tuning Algorithms*", IEEE Data Engineering Bulletin 34(4) 2011.
- ♦ J. Buck, N. Watkins, J. LeFevre, K. Ioannidou, C. Maltzahn, N. Polyzotis, S. Brandt, "*SciHadoop: Array-based Query Processing in Hadoop*", SC 2011.
- ♦ D. Kephart, J. LeFevre, "*CodeGen: The Generation and Testing of DNA Code Words*", IEEE Evolutionary Computation 2004, pdf.

PATENTS GRANTED

- ♦ 08-31-2021 US11106672B2 "Queries based on ranges of hash values."
- ♦ 02-02-2021 US10909119B2 "Accessing electronic databases."
- ♦ 12-10-2019 US10503718B2 "Parallel transfers of electronic data."
- ♦ 02-14-2017 US9569491B2 "Multistore online tuning system."
- ♦ 10-25-2016 US9477708B2 "System for multistore execution environments with storage constraints."
- ♦ 11-10-2015 US9183253B2 "System for evolutionary analytics."

LINKS

Academic page: users.soe.ucsc.edu, Google Scholar: scholar.google.com.

REFERENCES

Malu Castellanos, Oracle; Meichun Hsu, Oracle; Hakan Hacigumus, Confluent/Meta; Jagan Sankaranarayanan, Google; Philip Kufeldt, NVIDIA/Seagate; Peter Alvaro, UCSC; Michael J. Carey, UCI; Shel Finklestein, SAP/UCSC; Ike Nassi, SAP/UCSC.