

# Shana Hutchison

📞 862-226-2764 • ✉ shutchis@uw.edu • 🌐 [linkedin.com/in/shanahutchison](https://www.linkedin.com/in/shanahutchison)

Full-stack software engineer specializing in data management. On backends I build systems with strong mathematical foundations. On frontends I work with users to improve their time-to-insight on applications across the sciences.

## Education

<b>University of Washington</b> <i>M.S. in Computer Science &amp; Engineering (3/2017)</i> <b>Awards</b> NSF Graduate Research Fellow <b>Advisors</b> Bill Howe, Dan Suciu	<b>Seattle, WA</b> 9/2015–6/2021
<b>Stevens Institute of Technology</b> <i>M.S. in Computer Science, M.S. in Applied Mathematics, B.E. in Computer Engineering</i> <b>GPA</b> 4.00 Graduate, 3.97 Undergraduate <b>Thesis</b> <i>ModelWizard: Toward Interactive Model Construction</i> advised by David A. Naumann, Philippos Mordohai, Andrew D. Gordon <b>Awards</b> 2014 National Barry Goldwater Scholar, Association of Old Crows Scholar, Tau Beta Pi Scholar, Computing Research Association <i>Outstanding Undergraduate Researcher</i> Honorable Mention <b>Societies</b> <i>Tau Beta Pi</i> (Engineering), <i>Upsilon Pi Epsilon</i> (Computer Science), <i>Eta Kappa Nu</i> (IEEE)	<b>Hoboken, NJ</b> 8/2010–5/2015
<b>University of Edinburgh</b> <i>Study Abroad semester, 6 courses transferred</i>	<b>Edinburgh, UK</b> 1/2014–5/2014

## Experience

**Apache Accumulo – PMC Committer** 10/2015+  
○ Contribute code and discussion to the Apache Accumulo database community

### Laboratory & Industry

**Blue Origin – Software Engineer III** Kent, WA 11/2021+  
*Test & Flight Data Management*  
○ Developed the timeseries database hosting Blue's rocket sensor data, supporting near-real time analysis  
○ Specialized in front-end UI development with the React Typescript framework  
○ Contributed to *Manufacturing Resource Planning's* Java microservice analysis pipeline

**IBM Research Almaden – Research Intern** San Jose, CA 6/2017–12/2017  
*SystemML*; Mentors Matthias Boehm, Alexandre Evfimievski  
○ Developed cost-based optimization for Apache SystemML based on elementary rewrites over sum-product operators  
○ Achieved speedups as large as 20,000x on machine learning algorithm kernels

**MIT Lincoln Laboratory – Research Engineer** Lexington, MA 1/2015–9/2015  
*Computing and Analytics*; Advisors Jeremy Kepner, Vijay Gadepally  
○ Engineered Graphulo, a Java server-side matrix math library for the Accumulo database  
○ Recasted graph algorithms into the GraphBLAS standard; prototyped in Matlab

**Microsoft Research – Research Intern** Cambridge, UK 6/2014–8/2014  
*Programming Principles and Tools*; Advisor Andy Gordon  
○ Designed ModelWizard: a DSL in F# for interactive model construction targeting Tabular, a schema-based probabilistic programming language. Presented a concept poster at the Microsoft PhD Summer School

**Sandia National Laboratories – Technical Intern** Livermore, CA 5/2013–8/2013  
*Information Assurance*; Advisors Levi Lloyd, Tamara Kolda  
○ Pursued network anomaly detection via Accumulo schemas, machine learning and visualization  
○ Scaled LXCs (Linux Containers) with MiniMega, a mass distributed VM experiment platform

## MIT Lincoln Laboratory – Research Intern

Computing and Analytics; Advisor Jeremy Kepner

Bioengineering and Systems Technology; Advisor Darrell Ricke

- Integrated and benchmarked Accumulo distributed database features into D4M, a Matlab package delivering linear algebra and graph theory capabilities via Associative Arrays
- Applied D4M work to a DNA matching bioinformatics project, published in the *Lincoln Laboratory Journal*

Lexington, MA

5/2012–8/2012

## Brown Brothers Harriman – Web Development Co-op

Business Application Development; Advisors John David, Steve Hansen

- Designed and developed front- and back-end web applications for financial reporting using SQL, C++, and jQuery

Jersey City, NJ

1/2012–5/2012

## Teaching

### University of Washington

Seattle, WA

Paul G. Allen School of Computer Science & Engineering: *Lecturer*

1/2020–3/2020

- CSE 414: Database Systems – taught data management to 150 students as the instructor of record

Paul G. Allen School of Computer Science & Engineering: *Teaching Assistant*

9/2018–12/2020,

- CSE 344: Intro to Data Management, CSE 414: Database Systems

4/2020–6/2021

### Stevens Institute of Technology

Hoboken, NJ

Computer Science Department: *Teaching Assistant*

8/2012–12/2013

- CS 506: Intro to IT Security, CS 135: Discrete Structures, CS 334: Automata and Computation

Academic Support Center: *Tutor*

8/2011–12/2013

- Teach individuals and groups in Mathematics, Computer Science, and Engineering

## Leadership & Activities

**Comprehensive Energy Psychology:** Completed CEP level 1, 11/2019.

**Touch for Health Energetic Kinesiology:** Completed training level 1 - 4, 8/2019.

**UAW 4121, UW's Union of Academic Student Employees:**

- Steward for Computer Science & Engineering, 2018–2021
- Chair of Election Committee, 2017–2018

**Program Committees:**

- Reviewer for the International Conference on Database Theory (ICDT), 2018
- Reviewer for the Information Systems Journal (Wiley), 2016–2017

**Global Grand Challenges Summit: 2017,** authored a Top 10 podcast with Sarah Engel on *The “Power” of the Internet*

▷ <https://www.ucl.ac.uk/steapp/professional-education/ggcs-how-to-change-the-world/podcasts/sustainability/sustainability-6>

**Undergrad Activities:** Organized seminars as VP of the *Graduate Computer Science Society*; hosted a road race as President of the *Cycling Club*; led operations for the 2012 *Castle Point Anime Convention*, attracting over 2100 people

## Publications

partly under former name D. Hutchison

### Conference Papers

- [44] Y. R. Wang, S. Hutchison, J. Leang, B. Howe, and D. Suciu, **SPORES: Sum-product optimization via relational equality saturation for large scale linear algebra**, *Proceedings of the VLDB Endowment*, vol. 13, no. 11, pp. 1919–1932, Sep. 2020, ISSN: 2150-8097. DOI: 10.14778/3407790.3407799. arXiv: 2002.07951 [cs.DB], Video: <https://www.youtube.com/watch?v=g-sUU0Xjbrs>.
- [43] M. Boehm, B. Reinwald, D. Hutchison, A. V. Evimievski, and P. Sen, **On optimizing operator fusion plans for large-scale machine learning in SystemML**, *Proceedings of the VLDB Endowment*, vol. 11, no. 12, pp. 1755–1768, Aug. 2018. DOI: 10.14778/3229863.3229865. arXiv: 1801.00829 [cs.DB].
- [42] H. Jananthan, Z. Zhou, V. Gadepally, D. Hutchison, S. Kim, and J. Kepner, **Polystore mathematics of relational algebra**, in *Big Data*, IEEE, Dec. 2017. DOI: 10.1109/BigData.2017.8258298. arXiv: 1712.00802 [cs.DB].

- [41] D. Hutchison, **Distributed triangle counting in the Graphulo matrix math library**, in *High Performance Extreme Computing (HPEC)*, IEEE, Sep. 2017. DOI: 10.1109/HPEC.2017.8091041. arXiv: 1709.01054 [cs.DC], Graph Challenge Honorable Mention; Poster: <https://github.com/dhutchis/documents/blob/master/posters/2017-09-Graphulo-TriangleCounting-poster.pdf>.
- [40] L. Milechin, V. Gadepally, S. Samsi, J. Kepner, A. Chen, and D. Hutchison, **D4M 3.0: Extended database and language capabilities**, in *High Performance Extreme Computing (HPEC)*, IEEE, Sep. 2017. DOI: 10.1109/HPEC.2017.8091083. arXiv: 1708.02934 [cs.DB].
- [39] D. Hutchison, B. Howe, and D. Suciu, **LaraDB: A minimalist kernel for linear and relational algebra computation**, in *SIGMOD Workshop on Algorithms and Systems for MapReduce and Beyond (BeyondMR)*, ACM, May 2017. DOI: 10.1145/3070607.3070608. arXiv: 1703.07342 [cs.DB]. Online: <https://youtu.be/d-ZY8lIs5Pc?t=2m45s>, Slides: <https://sites.google.com/site/beyondmr2017/program>.
- [38] J. Wang et al., **The Myria big data management and analytics system and cloud service**, in *Conference on Innovative Data Systems Research (CIDR)*, Jan. 2017. Online: <https://homes.cs.washington.edu/~magda/papers/wang-cidr17.pdf>, Slides: <http://cidrdb.org/cidr2017/slides/p37-wang-cidr17-slides.pdf>.
- [37] D. Hutchison, J. Kepner, V. Gadepally, and B. Howe, **From NoSQL Accumulo to NewSQL Graphulo: Design and utility of graph algorithms inside a BigTable database**, in *High Performance Extreme Computing (HPEC)*, **Best Student Paper**, IEEE, Sep. 2016. DOI: 10.1109/HPEC.2016.7761577. arXiv: 1606.07085 [cs.DB], Slides: <https://github.com/Accla/graphulo/blob/master/docs/presentations/2016-09-HPEC-Graphulo-Algorithms.pdf>.
- [36] J. Kepner et al., **Associative array model of SQL, NoSQL, and NewSQL databases**, in *High Performance Extreme Computing (HPEC)*, IEEE, Sep. 2016. DOI: 10.1109/HPEC.2016.7761647. arXiv: 1606.05797 [cs.DB].
- [35] T. Weale, V. Gadepally, D. Hutchison, and J. Kepner, **Benchmarking the Graphulo processing framework**, in *High Performance Extreme Computing (HPEC)*, IEEE, Sep. 2016. DOI: 10.1109/HPEC.2016.7761640. arXiv: 1609.08642 [cs.DB].
- [34] A. Chen, A. Edelman, J. Kepner, V. Gadepally, and D. Hutchison, **Julia implementation of the dynamic distributed dimensional data model**, in *High Performance Extreme Computing (HPEC)*, **Best Paper**, IEEE, Sep. 2016. DOI: 10.1109/HPEC.2016.7761626. arXiv: 1608.04041 [cs.MS].
- [33] J. Kepner et al., **Mathematical foundations of the GraphBLAS**, in *High Performance Extreme Computing (HPEC)*, IEEE, Sep. 2016. DOI: 10.1109/HPEC.2016.7761646. arXiv: 1606.05790 [cs.MS].
- [32] D. Hutchison, J. Kepner, V. Gadepally, and A. Fuchs, **Graphulo implementation of server-side sparse matrix multiply in the Accumulo database**, in *High Performance Extreme Computing (HPEC)*, **Best Student Paper Finalist**, IEEE, Sep. 2015. DOI: 10.1109/HPEC.2015.7322448. arXiv: 1507.01066 [cs.DB], Slides: <https://github.com/Accla/graphulo/blob/master/docs/presentations/2015-09-HPEC-Graphulo-MatrixMultiply.pdf>.
- [31] V. Gadepally, J. Bolewski, D. Hook, D. Hutchison, B. Miller, and J. Kepner, **Graphulo: Linear algebra graph kernels for NoSQL databases**, in *International Parallel & Distributed Processing Symposium Workshops (IPDPSW)*, IEEE, May 2015. DOI: 10.1109/IPDPSW.2015.19. arXiv: 1508.07372 [cs.DS], Slides: <https://github.com/Accla/graphulo/blob/master/docs/presentations/2015-05-GABB-GraphuloInGraphBLAS.pdf>.
- [30] D. Hutchison and S. Kleinberg, **Causal inference under uncertainty via adjustments and SOPDs**, in *Causality and Experimentation in the Sciences*, Paris, France, Jul. 2013. Online: <http://caeits.sciencesconf.org/conference/caeits/hutchison-caeits2013.pdf>.

## Journal Papers.....

- [29] M. J. Smith *et al.*, **Monitoring internet trade to inform species conservation actions**, *Endangered Species Research*, Mar. 2017. DOI: 10.3354/esr00803.
- [28] J. Kepner, D. Ricke, and D. Hutchison, **Taming biological big data with D4M**, *Lincoln Laboratory Journal*, vol. 20, no. 1, 2013. Online: [https://www.ll.mit.edu/publications/journal/pdf/vol20\\_no1/20\\_1\\_6\\_Kepner.pdf](https://www.ll.mit.edu/publications/journal/pdf/vol20_no1/20_1_6_Kepner.pdf).

## Thesis.....

- [27] D. Hutchison, **ModelWizard: Toward interactive model construction**, M.S. Thesis, Stevens Institute of Technology, May 2015. arXiv: 1604.04639 [cs.PL].

## Book Chapters.....

- [26] V. Gadepally, J. Bolewski, D. Hook, S. Hutchison, B. Miller, and J. Kepner, **Graphulo: Linear algebra graph kernels**, in *Massive Graph Analytics*, D. Bader, Ed., Taylor & Francis, 2022, ISBN: 9780367464127.

## Preprints.....

- [25] D. Hutchison, B. Howe, and D. Suciu, **Lara: A key-value algebra underlying arrays and relations**, Apr. 2016. arXiv: 1604.03607 [cs.DB].

## Posters.....

- [24] D. Hutchison, M. Boehm, A. Evfimievski, B. Reinwald, P. Sen, and F. Ozcan, **Holistic sum-product optimization for large-scale machine learning**, Aug. 2017, Poster presented at IBM Intern Symposium.
- [23] D. Hutchison, B. Howe, V. Gadepally, and J. Kepner, **In-database vs. external system analytics on a key-value store**, Jan. 2017. Online: <https://github.com/Accla/graphulo/blob/master/docs/posters/2017-01-Graphulo-MapReduce.pdf>, Poster presented at North East Database Day (NEDB).
- [22] D. Hutchison, S. Jain, B. Howe, and D. Maier, **Ocean genomic analysis with Myria**, Aug. 2016. Online: <https://github.com/uwdb/related-work/blob/master/oceanography-istc/2016-08-Myria-ocean-ISTC.pdf>, Poster presented at summit for the Intel Science and Technology Center (ISTC) for Big Data.
- [21] D. Hutchison, V. Gadepally, J. Kepner, and B. Howe, **Graphulo: Native linear algebra in a NoSQL DB**, Aug. 2016. Online: <https://github.com/Accla/graphulo/blob/master/docs/posters/2016-08-Graphulo-algorithms-ISTC.pdf>, Poster presented at summit for the Intel Science and Technology Center (ISTC) for Big Data.
- [20] D. Hutchison, B. Howe, D. Suciu, and Z. Tatlock, **Polystore optimization via program expression graphs**, Jan. 2016. Online: <https://github.com/dhutchis/documents/blob/master/posters/2016-01-21-polypeg-poster.pdf>, Poster presented at North East Database Day (NEDB).
- [19] D. Hutchison, J. Kepner, and V. Gadepally, **Graphulo: Graph processing for Accumulo databases**, Jan. 2016. Online: <https://github.com/Accla/graphulo/blob/master/docs/posters/2015-09-Graphulo-ISTC-Poster.pdf>, Poster presented at North East Database Day (NEDB) and Sep. 2015 retreat for the Intel Science and Technology Center (ISTC) for Big Data.
- [18] L. Battle *et al.*, **Technologies for visualization of big medical text data**, Aug. 2015. Online: <https://github.com/Accla/graphulo/blob/master/docs/posters/2015-08-BigDAWG-MIMIC-Topic-Modeling.pdf>, Poster presented at Very Large Databases Conference (VLDB).
- [17] D. Hutchison, E. Cherin, X. Li, and H. Yang, **HBaaS: Heterogeneously-accelerated bioinformatics-as-a-service**, Apr. 2015, Poster and demo presented at Stevens Institute Senior Design Expo.
- [16] D. Hutchison and the Microsoft Research Tabular Team, **Structural clustering**, Jul. 2014. Online: <https://github.com/dhutchis/documents/blob/master/posters/2014-07-structural-clustering.pdf>, Poster presented at Microsoft PhD Summer School.

- [15] D. Hutchison and D. Kleszyk, **Investigating the influence of infrastructure on the load response of stevens course servers**, Dec. 2011, Course project for CPE 345 Modeling and Simulation.

## Other Talks and Demos.....

- [14] D. Hutchison, **Smart contracts**, Jan. 2018. Online: <https://www.youtube.com/watch?v=GMYkt7Rhmcg>, UWDB Seminar; Slides: <http://bit.ly/2DG9k87>.
- [13] D. Hutchison and M. Ahmad, **SystemML: Large-scale ML, DL, and optimization**, Oct. 2017, UWDB Seminar.
- [12] D. Hutchison, **Holistic sum-product optimization for large-scale machine learning**, Aug. 2017, IBM Project Presentation.
- [11] D. Hutchison, **In-database analytics for NoSQL key-value stores**, Dec. 2016. Online: <https://www.cs.washington.edu/qualsexam/dhutchis>, Slides: <https://github.com/Accla/graphulo/blob/master/docs/presentations/2016-12-Key-Value-Analytics.pdf>.
- [10] B. Howe, D. Hutchison, and S. Jain, **Demo of myria as a federated database**, Oct. 2016. Online: [https://github.com/uwescience/raco/blob/SPJA-federation/HPDA\\_review.ipynb](https://github.com/uwescience/raco/blob/SPJA-federation/HPDA_review.ipynb), Presentation and demo to government sponsors.
- [9] D. Hutchison, J. Kepner, and V. Gadepally, **Lara: A language of linear and relational algebra for polystores**, Dec. 2015. Online: [http://db.cs.washington.edu/events/database\\_day/2015/database\\_day\\_2015.html](http://db.cs.washington.edu/events/database_day/2015/database_day_2015.html), UWDB Database Day industry talk.
- [8] D. Hutchison, J. Kepner, and V. Gadepally, **Graphulo use and design**, Aug. 2015, Presentation and demo to government sponsors.
- [7] V. Gadepally, L. Edwards, D. Hutchison, and J. Kepner, **Using d4m for rapid prototyping of analytics for apache accumulo**, Apr. 2015. Online: <http://accumulosummit.com/program/talks/using-d4m-for-rapid-prototyping-of-analytics-for-apache-accumulo/>, Presentation at Accumulo Summit.
- [6] D. Hutchison, A. Gordon, and C. Russo, **ModelWizard: Interactive model construction for Tabular**, Aug. 2014. Online: <https://github.com/dhutchis/documents/blob/master/presentations/2014-08-modelwizard-tabular.pdf>, Presentation and demo to Programming Principles and Tools at Microsoft Research.
- [5] D. Hutchison and L. Lloyd, **Network anomaly detection**, Aug. 2013. Online: <https://github.com/dhutchis/documents/blob/master/presentations/2013-08-network-anomaly-detection.pdf>, Presentation at an intern symposium at Sandia National Laboratories.
- [4] D. Hutchison, **Our aims as modelers: Toward better predictions, explanations, interventions**, Aug. 2013. Online: [github.com/dhutchis/ModelTalk](https://github.com/dhutchis/ModelTalk), Presentation at a technical seminar at Sandia National Laboratories.
- [3] D. Hutchison, **Uncertain? Represent your belief with SOPD**, Jan. 2013. Online: <https://github.com/dhutchis/documents/blob/master/presentations/2013-01-omm-sopd.pdf>, Minute madness talk at Programming Languages Mentoring Workshop (PLMW) at the Principles of Programming Languages Conference (POPL).
- [2] D. Hutchison, **Accelerating bioinformatics with big data technologies**, Aug. 2012. Online: <https://github.com/dhutchis/documents/blob/master/presentations/2012-08-d4m-bioinformatics.pdf>, Presentation at an intern symposium at MIT Lincoln Laboratory.