Casey Stella

CONTACT
INFORMATION

7229 Selworthy Ln
Solon, OH 44137

github: http://github.com/cestella
e-mail: cestella@gmail.com

PROFESSIONAL
EXPERIENCE

Stripe, San Francisco, CA

Staff Software Engineer **March 2019 – Present**

I am the tech lead for the Machine Learning Infrastructure organization at Stripe. My role encompasses the implementation and day-to-day operation of a custom platform supporting many hundreds of models with tight SLAs. This includes, but is not limited to, scalable scoring, training and feature generation/engineering infrastructure.

Finite State, Columbus, OH

Vice President of Data **August 2018 – January 2019**

As Vice President of Data at Finite State, a Series A stage cybersecurity startup, I was in charge of scalable data architecture and machine learning. I operated as both an individual contributor as well as architectural support.

Apache Software Foundation,

Vice President of Apache Metron **April 2017 – November 2020**

In my role as a committer and contributor to Apache Metron, funded by Hortonworks, I also performed the role of Vice President of the Apache Metron project for the Apache Software Foundation. During this role, I shepherded the project from incubation through to being a top level project. During this period, I managed releases, spoke about the project at public conferences and contributed heavily to its development and advancement.

A small selection of talks that cover my work on Apache Metron:
- Apache Metron as a Case Study of a Modern Streaming Architecture on Hadoop
- Model as a Service: Modern Streaming Data Science with Apache Metron

Hortonworks, San Jose, CA

Principal Software Engineer **December 2012 – July 2018**

As a principal engineer and data scientist on the Apache Metron team, I lead the advanced analytics team. I was tasked with developing the next generation data science and machine learning infrastructure in Apache Metron. Apache Metron is an open source, scalable cybersecurity analytics platform that is maintained within the Apache Software Foundation.

Principal Architect **November 2012 – December 2015**

As a Principal Architect in the consulting services organization at Hortonworks, I provide mentorship and guidance in operationalizing Hadoop and its ecosystem to solve business problems for our clients. I work with our clients to reduce risk and time to market by providing expertise in big data.

- Provide proof-of-concepts to reduce engineering churn
- Give extensive presentations about the Hadoop ecosystem, best practices, data architecture in Hadoop
- Provide mentorship and guidance to other architects to help them become independent
- Provide review and feedback for existing physical architecture, data architecture and individual code
- Debug and solve issues with Hadoop as on-the-ground subject matter expert. This could include everything from patching components to post-mortem analysis of errors.

I focus on issues around data science and especially natural language processing at scale. I have domain knowledge in medical/clinical informatics and oil/gas data analysis and signal processing.

Explorys, Cleveland, OH

"Big Data" Architect **January 2011 – November 2012**

I was a "big data" architect and, prior to that, a senior software engineer on the platform team at Explorys. The team was responsible for the creation, care and maintenance of the high performance indexing infrastructure. My job required a deep understanding of the Hadoop ecosystem. I designed the next generation data architecture for the unstructured data at Explorys as well as writing, debugging, and analyzing the performance of many map reduce jobs to realize that architecture.

- Devised and lead the implementation of the next generation architecture for more efficient data ingestion and processing.
- Proficiency with mentoring and on-boarding new engineers who are not proficient in Hadoop and getting them up to speed quickly.
- Experience with being a technical lead of a team of engineers.
- Proficiency with modern natural language processing and general machine learning techniques and approaches
- Extensive experience with Hadoop and HBase, including multiple public presentations about these technologies.
- Experience with hands on data analysis and performing under pressure.
- Designed and wrote a layer on top of MapReduce to make the task of writing MapReduce jobs easier and more safe for Junior Engineers.
- Contributed much of the code in our open source project:
  `http://github.com/ExplorysMedical/Apothecary`

Game Communications, Mayfield Heights, OH

Senior Engineer **November 2009 – January 2011**

Designed, implemented and integrated into legacy code a scalable network infrastructure layer for an instant message and VOIP network

- Created a software routing layer and implementing an ordered, reliable protocol on top of the raw network layer
- Integrated Cassandra as a distributed persistent metadata store to provide metadata resolution for network entities on the network
- Created low-level network libraries in C++ and C# to send messages across network
- Spearheaded institution of pair-programming and semi-agile practices

ION Geophysical, Houston, TX

Research Geophysicist **October 2008 – November 2009**

Member of the time processing R&D team, tasked to design, prototype and implement production-ready algorithms to use applied math and signal processing to reduce noise in seismic data. This was a telecommuting position from Cleveland, OH.

- Implemented multithreaded C++ applied linear algebra library consisting of efficient STL-compliant data structures and algorithms to assist in signal processing
- Assisted in the design and implementation of an internal flow-based map-reduce system for doing distributed scientific computation
- Used Tokyo Cabinet to provide an on-disk index of seismic metadata in an efficient way
- Implemented a multidimensional least squares adaptive filter that decreased wall-clock time for our most common seismic task by an order of magnitude
- Developed a fixed-precision algorithm in C for the embedded PowerPC 405 platform which uses first-order statistics to differentiate noise from initial signal in an accurate, efficient and robust way
- Spearheaded institution of pair-programming and semi-agile practices in a geographically dispersed environment

Oracle, Cleveland, OH

Senior Member of Technical Staff **October 2005 – October 2008**

Member of the Oracle Enterprise Repository team. We provided an enterprise J2EE application to organize and manage assets and their metadata.

- Worked as part of a geographically distributed team using agile practices
- Designed and implemented an enterprise build system around Maven, porting an existing heterogeneous build system from Ant and shell scripts to Maven
- Implemented multiple performance and scalability improvements, resulting in substantial benefits deriving from better caching strategies and algorithms with more favorable CPU/Memory complexity characteristics.
- Managed summer interns

CERTIFICATION	**Cloudera Certified Hadoop Developer**	**August 2012**
	Cloudera Certified HBase Developer	**August 2012**

EDUCATION **Texas A&M University**, College Station, TX

Masters of Science in Mathematics **Spring 2005**

- Emphasis in Computational Complexity and Theoretical Computer Science
- Advised by Dr. J. Maurice Rojas

University of Louisiana at Monroe, Monroe, LA

Bachelors of Science in Mathematics and Computer Science **Spring 2002**

SELECTED
PUBLICATIONS
Ofer Mendelevitch, Douglas Eadline, and Casey Stella
"Practical Data Science with Hadoop and Spark: Designing and Building Effective Analytics at Scale",
Book by *Addison-Wesley*, 2016.

J. Maurice Rojas, Frederic Bihan, and Casey Stella
"Faster Real Feasibility via Circuit Discriminants",
In *Proceedings of ISSAC 2009*, pp. 39-46, ACM Press, 2009.

PATENTS "Targeted Optical Character Recognition (OCR) for Medical Terminology",
Patent Issued June 7, 2016, Patent number 9361531, Optum Technologies

"SYSTEM AND METHOD FOR USING AN EDITABLE LIFECYCLE EVENT DISTRIBUTION LIST WITH A SERVICE METADATA REPOSITORY",
Patent Issued March 27, 2012, Patent number 8145680, Oracle Corporation

SELECTED
TALKS
I was a regular at the Hadoop Summit, Spark Summit EU and QCon London. You can find the slides for my various talks here.

"New Complexity Thresholds for Sparse Real Polynomials",
Sixth International Joint Meeting of AMS and SMM, Invited Talk

HONORS AND
AWARDS
Outstanding Teaching Assistant [2005]
National Science Foundation VIGRE Fellowship [2004]
AUF & Regents Fellowships for Oustanding Academic Achievement [2003]