

Craig Andrew Struble, Ph.D.

(408)614-1088 | strubleca@yahoo.com | <https://www.linkedin.com/in/strubleca>

SUMMARY

A trained computer scientist specializing in bioinformatics, data science, and data engineering using modern cloud computing environments. I thrive in roles addressing complex data and computational challenges in biotechnology.

EXPERIENCE

PACIFIC BIOSCIENCES (PACBIO) | Director, Automation & Data Engineering
Sep. 2022 — Present | Menlo Park, CA

- Oversee development of telemetry gathering system for installed fleet of instruments.
- Transforming on-premise Elasticsearch-based telemetry system to cloud Snowflake-based system, improving stability, scalability, and performance.

INSCRIPTA | Senior Director, Computational Sciences and Engineering
Jul. 2017 — Sep. 2022 | Pleasanton, CA & Boulder, CO

- Department lead for 20 bioinformaticians, data scientists, data engineers, and software engineers.
- Responsible for computational infrastructure, algorithms, mathematical modeling, and statistical analysis supporting research and product development of CRISPR-based genome engineering products.
- Member of cross-functional product core teams for microbial and mammalian genome engineering platforms.
- Lead cross-functional core team through feasibility phase for a product extension of the Onyx™ microbial platform.

DELL EMC/EMC | Senior Principal Data Scientist | Principal Data Scientist
Feb. 2016 — Jul. 2017 | Feb. 2015 — Feb. 2016 | Santa Clara, CA

- Architect and developer for the Prometheus initiative, a microservice analytics platform to proactively manage EMC's deployed products.
- Developed services to identify common utilization patterns and outliers using time series clustering and Kalman filters. Developed models and services to predict risk of component failures leading to data loss.
- Mentored a team of 10+ developers in Shanghai. Oversaw the creation of continuous integration and deployment processes, API design, code review, and release planning.

ARIOSA DIAGNOSTICS | Director of Bioinformatics | Senior Scientist, Bioinformatics
Jan. 2012 — Feb. 2015 | Jan. 2011 — Jan. 2012 | San Jose, CA

- Managed a 6 person team of statisticians, data scientists and software engineers.
- Responsible for assay design, analytical pipelines, experimental design, laboratory troubleshooting and technical training, software development, system architecture, algorithm design, and release processes.
- Created and implemented the Fetal-fraction Optimized Risk of Trisomy Evaluation (FORTE™) algorithm, the core technology for Ariosa's Harmony Prenatal Test™ for genetic disorders. FORTE™ uses Bayesian analysis and machine learning.
- Developed quality control metrics and quality thresholds for sample data quality and oversaw the development and implementation in R of quality monitoring dashboards.

MARQUETTE UNIVERSITY | Associate Professor (tenured) | Assistant Professor
Aug. 2008 — Jun. 2011 | Aug. 2001 — Aug. 2008 | Milwaukee, WI

- Advanced high performance computing in Southeast Wisconsin. Obtained \$500,000 NSF major research instrument grant, for a computing cluster. Co-founder of the Southeast Wisconsin High Performance (SeWHIP) consortium.
- Architected **Pools of Virtual Boxes** a deployment of VirtualBox with Linux and HTCondor to support distributed computing on over 500 Windows lab machines.
- Architected a distributed protein docking platform enriching for drug candidates.
- Developed a named entity recognizer using a combination of n -gram modeling, abbreviation recognition, and Conditional Random Fields for the BioCreAtIvE II shared task for Gene Mentions.
- Performed bioinformatic analyses in support of several projects. Outcomes include: understanding impacts of dioxin on reproduction in zebrafish model systems; tandem SNP analysis for detection of trisomy 21; microbial population analysis in anaerobic digestion; discovery of a new microbial species.
- Introduced courses on data mining, text mining, and cloud computing. Taught bioinformatics, artificial intelligence, operating systems, programming languages, mathematical modeling, data structures and algorithms. Published 22 articles spanning bioinformatics, text mining, high throughput computing, case based reasoning and computer science education. Directed 13 master's student theses.

SKILLS

PROGRAMMING LANGUAGES

Python, R, Java, SQL, C#, C, Perl, Awk, Bash, LISP, PROLOG

BIOINFORMATICS

BioPython, samtools, bedtools, cellranger, Seurat, bwa, bowtie2, minimap, Illumina NGS, Oxford Nanopore, Sanger sequencing, microarrays

DATA SCIENCE

Neural networks, Bayesian networks, unsupervised learning, NLP, dimension reduction, Pandas, Numpy, scikit-learn, Tensorflow, Keras, dplyr, tidyverse

DATA ENGINEERING

PostgreSQL, Greenplum, Hadoop, Spark, Pig, Hive, Amazon Redshift, Elasticsearch, Snowflake, Kafka

CLOUD & DISTRIBUTED COMPUTING

Docker, Kubernetes, AWS (multiple services), Google Cloud, Microsoft Azure, HTCondor, SGE, Snakemake, CWL

MANAGEMENT

Agile/Scrum, Project Management, Product Development Process (PDP), Core Team Lead/Member, JIRA

EDUCATION

VIRGINIA TECH

Ph.D. in Computer Science and Applications

M.S. in Computer Science

B.S. in Computer Science, Math

ADDITIONAL INFORMATION

42 publications in fetal diagnostics, bioinformatics, data science, and computer science research areas; 8 issued patents. Detailed list available upon request.