

William Ronchetti

M.Eng., Computer Science • Back-End / Infrastructure Engineer

Alexandria, VA • (585) 831-0706 • wrr33@cornell.edu • github.com/willronchetti • linkedin.com/in/william-ronchetti-569839121

SUMMARY

Back-end and infrastructure software engineer with 7+ years building secure, scalable AWS data platforms in Python. Strengths in infrastructure-as-code, API and database performance optimization, container-based deployments, observability, and delivering FISMA-Moderate and GDPR-compliant systems that have passed multiple independent security audits.

TECHNICAL SKILLS

Languages: Python, C/C++, Java, OCaml, SQL

Infrastructure & IaC: AWS CloudFormation, VPC, EC2, ECS (Fargate), ECR, Elastic Beanstalk, Lambda, Batch, Step Functions, API Gateway, Route 53; Docker, Nginx, Apache

Data & Storage: PostgreSQL (RDS), Elasticsearch, DynamoDB, Amazon Athena, S3, Parquet

Observability & CI/CD: CloudWatch, Splunk (HEC), Fluent Bit / FireLens, Sentry, GitHub Actions, Travis CI, ReadTheDocs

Security: AWS KMS, Secrets Manager, Security Hub, Certificate Manager (ACM); FISMA-Moderate, GDPR

Scientific / Pipelines: Nextflow, Seqera

PROFESSIONAL EXPERIENCE

Novo Nordisk (via Kelly Services Global)

Remote • Sept 2024–Present

Data Product Engineer

- Spearhead development of the Omics Platform, an internal AWS-based system for managing petabytes of diverse omics data used by dozens of research teams for both cataloguing and analysis.
- Built a highly scalable data ingestion pipeline on AWS Lambda, Glue, and Batch for validation and transfer of raw data and discovery metadata, processing hundreds of datasets across several dozen vendors/CROs.
- Designed a metadata model for tracking diverse omics datasets using Parquet and Amazon Athena.
- Review code and mentor a small team of engineers on software, cloud-engineering, and security best practices to ensure GDPR compliance.

Harvard Medical School, Dept. of Biomedical Informatics

Remote / Boston, MA • Aug 2019–Present

Senior Software Engineer (since 2022; half-time since Jan 2025) — previously Software Engineer

- Lead back-end development in Python for the CGAP, 4DN, and SMaHT data platforms, supporting petabytes of raw data and associated metadata.
- Refactored complex, organically grown AWS infrastructure into reproducible infrastructure-as-code (CloudFormation), achieving FISMA-Moderate compliance and passing multiple independent security audits with workflows for promptly resolving vulnerabilities.
- Migrated the web platform from Apache on Elastic Beanstalk to Nginx on ECS (Fargate), reducing web portal compute costs by 50% and allowing for configurable autoscaling in response to network traffic.
- Implemented observability with structured event logging via a Splunk HEC pipeline and container log shipping through FireLens / Fluent Bit.
- Optimized Postgres and Elasticsearch query performance by resolving aggregation timeouts caused by large terms filters, cutting response time by 50%.
- Devised and implemented advanced security protocols (encryption, secrets management, access controls) and mentored junior engineers and domain experts on feature development and code review.

Duke University, Dept. of Electrical & Computer Engineering

Durham, NC • Feb 2019–Present

Associate in Research (Part-Time)

- Develop new assignments for an advanced C++ course and improve introductory C programming assignments on Coursera; support students with C and Python coursework.

Cornell University

Ithaca, NY • Aug 2016–Dec 2018

Teaching Assistant (Part-Time)

- Taught Operating Systems (CS 4410/4411) and System Security (CS 5430/5431) with practicum sections; built and improved C autograders and graded exams.

EDUCATION

Cornell University

Ithaca, NY

M.Eng., Computer Science (GPA 3.69), Dec 2018 • B.A., Computer Science (GPA 3.43), May 2018

Varsity Cross Country, Indoor & Outdoor Track; Association of Computer Science Undergraduates.

SELECTED PUBLICATIONS

Co-author on 5 peer-reviewed publications for data-platform and pipeline contributions with 267 citations, including:

- The Somatic Mosaicism across Human Tissues Network — *Nature* (2025). [nature.com/articles/s41586-025-09096-7](https://doi.org/10.1038/s41586-025-09096-7)
- An integrated view of the structure and function of the human 4D nucleome — *Nature* (2025). [nature.com/articles/s41586-025-09890-3](https://doi.org/10.1038/s41586-025-09890-3)
- Joint, multifaceted genomic analysis enables diagnosis of diverse, ultra-rare monogenic presentations — *Nature Communications* (2025). [nature.com/articles/s41467-025-61712-2](https://doi.org/10.1038/s41467-025-61712-2)
- The 4D Nucleome Data Portal as a resource for searching and visualizing curated nucleomics data — *Nature Communications* (2022). [nature.com/articles/s41467-022-29697-4](https://doi.org/10.1038/s41467-022-29697-4)

COMMUNITY INVOLVEMENT

- Assistant Cross-Country and Track & Field Coach, Arlington Public Schools (Yorktown).
- Volunteer, Virginia German Shepherd Rescue and the Lost Dog & Cat Rescue Foundation.