

# Zachary Dawson

zachwdawson@gmail.com | 508 446 3322 | zachwdawson.github.io

## Education

### Georgia Institute of Technology

Aug 2024 – May 2026

MS in Computer Science | GPA: 4.0

- Relevant Coursework: Deep Learning, Machine Learning, Bayesian Statistics, Natural Language Processing, Software Development Processes, Computer Networks

### Northeastern University Honors Program

Sept 2017 – May 2021

BS in Computer Science, Minor Mathematics, magna cum laude

## Experience

### Senior Data Engineer, Global Atlantic Financial Group – Brighton, MA

Aug 2022 - May 2024

- Introduced the company's first ML model into a data pipeline to identify duplicate contacts 10% more accurately than previous rules-based model.
- Converted pipeline of 200+ jobs from Control-M to AWS Glue using Terraform.
- Extended the annuity CRM data warehouse for life insurance data by updating mastering logic for all key business objects, facilitating the tracking and analysis of \$1 billion in business.
- Led planning and development of splitting policies among various salespeople, resulting in \$5 million in commission to be split among various participating salespeople.
- Created integration for reading for and writing to Salesforce via Salesforce Bulk API, adding 5000 new prospects to Salesforce and the enterprise data warehouse.
- Mentored and onboarded interns and new-grad hires onto EDH projects.

### Data Engineer, Global Atlantic Financial Group – Brighton, MA

Sept 2021 – July 2022

- Collaborated with business stakeholders and technology leaders to migrate the existing annuity CRM system to in-house data warehouse, enabling reporting on 700,000 policies and \$40 billion worth of transactions.
- Recreated the territory management process for life and annuity business, saving users 1 hour per day.
- Developed a file-type agnostic ingestion process to Hive and Redshift for core ETL architecture.
- Designed and implemented mastering process for contact information from seven data sources.

### Bioinformatics Co-op, MIT BioMicro Center – Cambridge, MA

Jan 2020 – June 2020

- Ensured timely data delivery and analysis of new NovaSeq sequencing data, averaging 2-4 terabytes a week.
- Migrated legacy FASTQ generation and format to industry standards, decreasing FASTQ generation time for NextSeq and NovaSeq flow cells by over 50
- Added Novaseq and 10X Cellranger processing into the existing Illumina pipeline, saving biologists days of running and monitoring scripts.

### Server/SRE Co-op, Skillz Inc. – San Francisco, CA

Jan 2019 – June 2019

- Led planning and implementation of new internal load testing infrastructure, enabling up to ten times standard production load in staging.
- Created weekly company-wide presentations on system health and performance using metrics from monitoring tools in staging and production environments.

## Technical Skills

**Languages:** Python, SQL, Scala, Java, Bash

**Frameworks and Libraries:** Spark, Hadoop, Terraform, Pandas, NumPy, PyTorch, Scikit-Learn, SciPy, Spring Boot, Terraform, Gatling, Ruffus

**Tools:** Tools: Git, GitLab, Linux, AWS (S3, EC2, Redshift, Glue, Lambda), MySQL, Maven, ControlM Workflow Automation, Slurm Workload Manager, Jenkins, RESTful APIs, New Relic, Jira, Jupyter Notebook, Next Generation Sequencing