

Ryan P Smith

linkedin.com/in/rpseq

github.com/rpseq

Salt Lake City, UT

ryan.smith.p@gmail.com

+1-319-899-0190

PROFILE

Bioinformatics scientist turned cloud data and infrastructure engineer. 7 years of bioinformatics computing work in academia and 2.5 years of professional software engineering on a Data Engineering team at Amazon.

TECHNICAL SKILLS

- **Languages:** Python, Bash, R (ggplot), awk, SQL, C++, golang, L^AT_EX, Mathematica, Java, Scala
- **Technologies:** Linux/Unix, Kubernetes, Helm, AWS (S3, EC2, Redshift, EKS, EMR), GCP(GKE), Grafana, Prometheus, Terraform, Looker, git, Docker, Airflow, CircleCI, GitLab CI, Databricks, Spark, HBase, Pinot, Druid, Avro/Parquet, Kafka, GATK, bedtools, bwa-mem

EXPERIENCE

- **eero, an Amazon company** San Francisco, CA
Software Development Engineer, Data January 2019 – Present
 - **Data Infrastructure and Operations:** Design, provision, maintain, and scale k8s clusters to power most of our data systems. All cloud infrastructure is provisioned with Terraform and deployments automated with GitLab. Databricks and Spark for custom batch and streaming transformations. Metrics, monitoring and alerting using Prometheus and Grafana. Build and maintain a collection of services using GitLab and Helm deploying to k8s clusters: Apache Airflow, Apache Pinot, Apache Superset, Kafka Connect sinks, Scala stream-processing apps.
 - **Real-time Data Systems:** Build and maintain data systems powering customer-facing application features. These features provide customers with visual dashboards for observing device bandwidth usage as well as security and content filtering events across a fleet of >1.5 million eero routers. Kafka for event streaming and Kafka Connect running in k8s to sync up to 60k records per second to a large HBase cluster, while ALSO serving on average 60k customer queries per second. Also design and build all of the cloud infrastructure for a new Pinot cluster running in k8s. This powers sub-second queries over very large datasets with continuous realtime ingestion of fleet data from Kafka. Apache Superset is used as the data visualization and exploration tool.
- **Girihlet, Inc., Oakland Genomics Center** Oakland, CA
Bioinformaticist + IT Engineer Aug 2018 – Oct 2018
 - **Bioinformatics:** Provide a quality control dashboard for Illumina short read data. General Bash, Python, and Perl scripting to automate mitochondrial genome variant calling.
 - **Information Technology:** Upgrade and maintain the Oakland Genomics Center networks. Maintain two on-premises CentOS servers for data processing and web hosting (Apache, NGINX)
- **The McDonnell Genome Institute, Washington University** St Louis, MO
Graduate Research Scientist, Ira Hall Lab, Computational Genetics Feb 2015 – April 2018
 - **Distributed Computing, Bioinformatics:** Process and analyze population-level whole-genome sequencing data to improve our understanding of structural variation in human genomes. Computation is done in a Linux environment with Bash pipelines, distributed across a high-performance computing cluster. Pipelines often use tools combining Bayesian statistics and machine learning approaches, packaged and deployed via Docker containers.
 - **Single-cell Sequencing:** Develop novel computational and molecular biology methods for whole-genome sequencing of single mammalian neurons. Process up to 10 TB of raw Illumina sequencing data in 36 hours.
 - **Teaching Assistant:** Advise students in characterizing the genome sequences of unknown bacteriophages. Lab course using bioinformatics tools to identify genes, predict their function, and classify novel phages into existing phylogenetic groups.
- **University of Iowa** Iowa City, IA
Undergraduate Research Fellow, Adam Dupuy Lab, Cancer Genetics Aug 2010 – May 2014
 - **Genome Editing:** Designed methods for viral genetic engineering in mouse models of human cancers and a sequencing method for detecting resulting transgene insertions.
 - **Bioinformatics:** Processed high-throughput genome sequencing data using Linux bioinformatics tools followed by ad-hoc statistical analyses and data visualization in Python and R.

EDUCATION

- **Washington University** St Louis, MO
MA, Molecular Genetics and Genomics Aug. 2014 – May 2018
- **University of Iowa** Iowa City, IA
BS, Microbiology and Informatics Aug. 2009 – May 2014