

Professional Experience

Data Engineer (Consultant) Animal Cancer DX, LLC, Raleigh, NC
Jul 2021 - present

- Individually designed and configured Amazon Web Services (AWS) framework (Elastic Compute Cloud (EC2), Relational Database Service (RDS)) to host a Shiny application
- Individually designed and created an "Extract, Transform, and Load" (ETL) pipeline to aggregate, clean and warehouse experimental data on AWS RDS, automatically 1.) generate daily reports with statistical analyses and test results using RMarkdown, and 2.) distribute the RMarkdown PDF output to the end user via email

Senior Bioinformatics Data Scientist, Program Manager RTI International, Research Triangle Park, NC
GenOmics, Bioinformatics, & Translational (GOBOT) Research Center Jun 2015 - present

- Supervise multiple teams of data analysts
- Responsible for manipulating and cleaning large data volumes, designing and building ETL pipelines to aggregate data, enhance data interoperability & automate analyses
- Utilize Snowpipe to ingest Twitter data into Snowflake to directly query and transform JSON data
- Designed and created ETL pipeline to aggregate, clean and warehouse experimental data on AWS RDS, automatically 1.) generate daily reports with statistical analyses and test results using RMarkdown, and 2.) distribute the report to the end user via email
- Design and configure Amazon Web Services (AWS) Elastic Compute Cloud (EC2) & Relational Database Service (RDS) to host ETL pipelines and Shiny web applications
- Create meaningful visualizations from large data volumes (examples can be found in 2019 publication)
- Interface with both technical and non-technical stakeholders to resolve complex business and technical issues
- Migrated legacy Oracle systems to AWS EC2 and RDS (MySQL)

Bioinformatics Data Scientist Johns Hopkins Medical Institutions, Baltimore, MD
Sidney Kimmel Comprehensive Cancer Center Aug 2013 - Jun 2015

- Served as systems administrator for high-performance computing cluster
- Completed 2-year data mining project in 10 weeks using R, SQL, and bash
- Individually designed and created a complex ETL pipeline to populate a SQL database
- Designed and implemented analysis pipelines for large genomic datasets generated from clinical trials
- Incorporated machine learning methods to quality control processes where applicable
- Individually created novel visualizations from large genomic datasets (examples can be found in 2017 publication)

Bioinformatics Analyst Johns Hopkins Medical Institutions, Baltimore, MD
Pediatrics Infectious Diseases Oct 2011 - Aug 2013

- Individually designed and created an ETL pipeline, increasing the efficiency of the legacy workflow by over 700%

Bioinformatics Researcher Loyola University, Chicago, IL
Putonti Lab Jan 2010 - Jan 2011

- Created ETL pipeline to aggregate and analyze sequence data extracted from NCBI repository

Education

Johns Hopkins University Baltimore, MD
M.Sc. Bioinformatics Aug 2013

Loyola University Chicago, IL
B.Sc. Bioinformatics May 2011

Skills

Programming R (dplyr, tidyr, ggplot), SQL, Python, Perl, bash, Shiny, RMarkdown, knitr, Jupyter, HTML, CSS

Technologies Tableau, Snowflake, Hive, Jira, Confluence, AWS (EC2, RDS, S3), Git/Subversion, REDCap
Working knowledge of Airflow, Spark, Hadoop, AWS EMR

Published Visualizations

Gern, J. E., Jackson, D. J., Lemanske, R. F., Seroogy, C. M., Tachinardi, U., Craven, M., **Hwang, S. Y.**, ... Bacharier, L. B. (2019). The Children's Respiratory and Environmental Workgroup (CREW) Birth Cohort Consortium: Design, methods, and study population. *Respiratory Research*, 20(1), 115. doi:10.1186/s12931-019-1088-9

Vaz, M., **Hwang, S. Y.**, Kagiampaki, I., Phallen, J., Patil, A., O'Hagan, H. M., ... Baylin, S. B. (2017). Chronic cigarette smoke-induced epigenomic changes precede sensitization of bronchial epithelial cells to single-step transformation by KRAS mutations. *Cancer Cell*, 32, 360–376. doi:10.1016/j.ccell.2017.08.006

References