Naveen Kannan

DATABASE ENGINEER

Cleveland, Ohio | United States Permanent Resident

☐ +1 973 519 5397 | ✉ naveen.kannan12@gmail.com | 🏠 naveenk2022.github.io/quarto_blog | 🐙 naveenk2022 | 💼 naveen-kannan-177120118

Experienced Bioinformatics Data Engineer with 3+ years of expertise in designing scalable ETL pipelines, integrating large-scale biological datasets, and building robust data infrastructure to support cutting-edge research. Proven track record in establishing best practices for data architecture, collaborating cross-functionally to align systems with analytical goals, and driving innovative solutions for biological discovery and therapeutic development.

Work Experience

Database Engineer   *Cleveland, Ohio*

TRAILHEAD BIOSYSTEMS   *JULY 2024 - Present*

- **HD-DoE Database Design and Implementation**
    – Automated extraction of structured data from **20 gigabytes** of proprietary reports into PostgreSQL using Python, reducing manual processing time by **90 percent**.
    – Architected **normalized database schema** enabling complex queries across **3 years of historical data**.
    – Optimized database performance through indexing, partitioning, and query tuning, resulting in **40 percent** faster query response times.
    – Built interactive dashboard **visualizing key metrics and trends**, enabling stakeholders to self-serve data insights.
    – Transformed **siloed report data** into **queryable format** supporting company-wide aggregate analysis.
- **Internal Banked Cell Inventory Tracking**
    – Implemented normalized database schema to **replace legacy Excel tracking system** for banked cell inventory.
    – Developed **PowerApps interface** for streamlined data entry of bioreactors metrics, QC data and banked cell vial counts.
    – Created **relational data model** connecting bioreactor,cell vial and cell count data, enabling granular inventory tracking.
    – Transformed manual spreadsheet tracking into structured database system, improving data accuracy and reducing entry errors by **90 percent**.

Research Associate   *Cleveland, Ohio*

POPULATION HEALTH AND QUANTITATIVE HEALTH SCIENCES, SCHOOL OF MEDICINE, CASE WESTERN
                                                                                                                                                           MAY 2023 - JULY 2024
RESERVE UNIVERSITY

- **Automated Large-Scale Cluster Deployment and Management**
    – Orchestrated the redeployment of a high-performance computing cluster with 23 nodes.
    – Configured a **PXE server** on the service node, enabling **fully automated, unattended OS provisioning** across the cluster.
    – Automated network configuration via a **DHCP server**, ensuring consistent IP address allocation across the cluster.
    – Developed **Ansible scripts** to intergrate with **iDRAC**, facilitating remote, automated PXE-based cluster reprovisioning.
- **PostgreSQL Database Administration**
    – Configured and deployed a **PostgreSQL cluster**, along with schema design and data modeling.
    – Integrated PostgreSQL with Spark and automated ETL and schema generation for **4 terabytes of large genomic flatfiles**.
    – Implemented **comprehensive security measures**, including user role management, privilege controls, firewall rules, and network security for secure database operations.
- **Automated HDFS and Spark Deployment**
    – Developed **Ansible scripts** to automate installation and configuration of **Apache Hadoop HDFS, YARN, MapReduce, and Spark** on a HPC cluster with 18 worker nodes and 1 master node.
- **Docker/Singularity Container Deployment**
    – Developed Mamba-based Docker/Singularity containers for reproducible scientific pipelines, reducing container build times by 300 percent, and deployed **RStudio and JupyterLab servers** through **Docker** containers.
    – Transferred Docker storage cache to a physical hard drive, eliminating storage errors and preserving existing images.
- **Genomic Data Analysis**
    – With the **Alzheimer's Disease Sequencing Project**, annotated and scored **362 million genetic variants** associated with Alzheimer's disease.
    – Developed **Generative Adversarial Neural Network** to create synthetic human genomic data trained on 1000 Genomes data, incorporating quality checks for Linkage Dysequilibrium and Hardy Weinburg Equilibrium.

Graduate Research Assistant   *Cleveland, Ohio*

POPULATION HEALTH AND QUANTITATIVE HEALTH SCIENCES, SCHOOL OF MEDICINE, CASE WESTERN
                                                                                                                                                           JANUARY 2023 - MAY 2023
RESERVE UNIVERSITY

- Engineered Docker containers for streamlined deployment of specialized R packages and Python modules, optimizing analyses for single-cell RNA datasets.

Graduate Research Assistant   *Cleveland, Ohio*

GIS HEALTH AND HAZARDS LAB, SCHOOL OF MEDICINE, CASE WESTERN RESERVE UNIVERSITY   *MAY 2022 - DECEMBER 2022*

- **Mapping human movement patterns after a natural disaster in the Democratic Republic of Congo.**
    – Developed Python scripts to extract GPS coordinates from geospatial video feeds of refugee camps in DRC after Mt. Nyiragongo eruption, and implemented ML algorithms (YOLOv4) for tent detection for each frame of the video feed.
    – Generated heat maps of tent distribution based on GPS coordinates.
    – Wrote up the findings as a part of a manuscript for a peer-reviewed publication.

Junior Resident Doctor *Chennai, India*
DEPARTMENT OF PSYCHIATRY, SAVEETHA MEDICAL COLLEGE *JULY 2020 - SEPTEMBER 2020*
- **Diagnosing and managing patients with psychiatric illnesses**
    - Learned to evaluate and diagnose psychiatric illnesses in stressful environments, and was assigned and achieved a goal of management of five patients per day.

Junior Resident Doctor *Chennai, India*
MADRAS MEDICAL COLLEGE *MARCH 2019 - MARCH 2020*
- **Rotated through the following departments:** Internal Medicine, General Surgery, Obstetrics and Gynecology, Community Medicine, Psychiatry, Emergency Trauma Ward, Labor Ward.

Education

MSc in Biomedical and Health Informatics *Cleveland, Ohio*
CASE WESTERN RESERVE UNIVERSITY *JANUARY 2022 – AUGUST 2023*
- GPA-3.9

Bachelor of Medicine and Bachelor of Surgery *Chennai, India*
MADRAS MEDICAL COLLEGE *SEPTEMBER 2016 - MARCH 2020*
- GPA-4.0

Skills

Operating Systems & Environments
DEBIAN, UBUNTU, FEDORA RED HAT LINUX, RASPBIAN LINUX, WINDOWS 11

Programming & Scripting
PYTHON, R, BASH, SQL

Data Engineering & Analytics Tools
HADOOP/SPARK (PYSPARK, HIVE), JUPYTER, GEOPANDAS, SF (R), QGIS

Development & Automation
DOCKER, SINGULARITY, SLURM, ANSIBLE, GIT

Machine Learning & Libraries
TENSORFLOW, PYTORCH, SCIKIT-LEARN, PANDAS, NUMPY, SCIPY

Databases & Storage
HDFS (HADOOP DISTRIBUTED FILE SYSTEM), POSTGRESQL SERVER

Bioinformatics Tools
SAMTOOLS, BCFTOOLS, HTSLIB, TABIX, PLINK

Markup & Documentation
LATEX, MARKDOWN, R MARKDOWN

Teaching Experience

Teaching Assistant *Case Western Reserve University*
A DATA DRIVEN INTRODUCTION TO GENOMICS AND HUMAN HEALTH (PQHS 451) *AUGUST 2023 - DECEMBER 2023*
- Held weekly office hours, developed and graded assignments.

Leadership

Vice President *Cleveland, Ohio*
STEINER HOUSE INTERNATIONAL CO-OP *MAY 2022 - OCTOBER 2022*
- **Community Leadership**
    - Facilitated the integration of new residents, fostered community through monthly social events, and collaborated with the executive board to ensure effective communication.

Admissions Manager *Cleveland, Ohio*
STEINER HOUSE INTERNATIONAL CO-OP *MAY 2022 - SEPTEMBER 2022*
- **Residential Admissions and Selections**
    - Successfully admitted and onboarded a diverse graduate student cohort, including students from 8 different countries.

Ordering Manager *Cleveland, Ohio*
STEINER HOUSE INTERNATIONAL CO-OP *MAY 2022 - SEPTEMBER 2022*
- **Inventory Management**
  - Managed inventory for Steiner House, with a capacity of 21 students.

Projects

Interactive Geospatial Dashboard Creation *Case Western Reserve University*
MAPPING THE ENVIRONMENTAL IMPACT OF THE NORFOLK SOUTHERN TRAIN DERAILMENT IN OHIO *2023*
- `https://naveen-kannan.shinyapps.io/final_east_palestine_dashboard_2`
  - Coded scripts to scrape publicly available air and water quality data from USGS(United States Geological Survey) and AirNow AQI for counties within a 30-mile radius of the Norfolk Southern train derailment in February 2023.
  - Implemented an interactive Shiny dashboard in R, offering users the ability to explore and analyze daily trends in Ohio, West Virginia, and Pennsylvania counties throughout January and February.

Data Science Blog *Cleveland, Ohio*
TUTORIALS ON DATA ENGINEERING AND ARCHITECTURE *2023*
- `https://naveenkannan.netlify.app/blog_main.html`
  - Authored tutorials on topics ranging from DevOps to data engineering tutorials.

Data Analytics and Visualization *Cleveland, Ohio*
ANALYSIS AND VISUALIZATION OF PUBLIC HEALTH DATASETS *2022*
- `https://naveenkannan.netlify.app/portfolio.html`
  - Performed statistical analysis of large public health datasets, including the Population Assessment of Tobacco and Health (PATH) Study, and the NHANES (National Health and Nutrition Examination Survey).
  - Interpreted and presented the findings in a scientific manner, alongside data visualization to further enhance the impact of the findings.

Publications

- Ajayakumar, J., Curtis, A. J., Maisha, F. M., Bempah, S., Ali, A., **Kannan, N.**, Armstrong, G. & Morris Jr, J. G. (2024). Using spatial video and deep learning for automated mapping of ground-level context in relief camps. International Journal of Health Geographics, 23(1), 23.

- Ruksakulpiwat, S., Thongking, W., **Kannan, N.**, Wright, E., Niyomyart, A., Benjasirisan, C., …& Still, C. H. (2024). Understanding the Relationship Between Comorbidities, Medication Nonadherence, Activities of Daily Living, and Heart Condition Status Among Older Adults in the United States: A Regression Analysis and Machine Learning Approach. Journal of Cardiovascular Nursing, 10-1097.

- **Naveen Kannan**[1], Nicholas Wheeler[1], Genome Center for Alzheimer's Disease, Li-San Wang[2], Yuk Yee Leung[2], William S. Bush[1],[1]Cleveland Institute for Computational Biology, Department for Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA;[2] Department of Pathology and Laboratory Medicine, Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA, Quality Control metrics of individual genetic variants in the Alzheimer's Disease Sequencing Project are associated with FAVOR annotations.

- **N. Kannan**[1], N. Wheeler[1], Genome Center for Alzheimer's Disease, L-S. Wang[2], Y. Leung[2], W. S. Bush[1]; [1]Cleveland Inst. for Computational Biology, Dept. for Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [2]Dept. of Pathology and Lab. Med., Penn Neurodegeneration Genomics Ctr., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA., Annotation and scoring of the deleteriousness of individual genetic variants in the 4th release of the Alzheimer's Disease Sequencing Project;(PB4451). Presented at the Annual Meeting of The American Society of Human Genetics, November 3, 2023, in Washington DC.

- Segamalai, D., Abdul Jameel, A. R., **Kannan, N**., Anbalagan, A., Duraisamy, B., Raju, P., & Devy Gounder, K. (2017). Mediastinal pseudocyst: varied presentations and management—experience from a tertiary referral care centre in India. *HPB Surgery*, *2017*.

- Vellaisamy, R., **Kannan, N**., Anbalagan, A., Raju, P., Duraisamy, B., Murugesan, C. S., & Gounder, K. D. (2016). Endoscopic access to hepatic duct through duodenum during follow up–after primary surgery for hepatolithiasis. *HPB*, *18*, e530.

- Ramasamy, V., Vellaisamy, R., **Kannan, N**., & Gounder, K. D. (2016). Refined technique of access loop in hepatobiliary surgery. *HPB*, *18*, e593-e594.

- **Kannan, N.**, Vellaisamy, R., Govindarajan, M., & Gounder, K. D. (2016). Pellagra following pancreaticoduodenectomy for malignant pancreatic carcinoid with pluripotent hormonal potential. *HPB*, *18*, e381-e382.