(973) 519-5397
Cleveland, Ohio
naveen.kannan12@gmail.com

Naveen Kannan

Data Engineer

Portfolio: Data Science Blog
github.com/naveenk2022
linkedin.com/in/naveenkannan1

Three (3) years expert skills in database development (PostgreSQL, MySQL, Hadoop) using ETL/ELT tools (SQLAlchemy, Alembic, Spark), developing production level REST APIs (FastAPI, Flask) and deploying scalable, containerized (Docker) workflows in HPC (High Performance Computing), AWS (EC2, RDS) and Azure (VM) cloud platforms.

Skills

Tools and Languages	Python, R, LaTeX, MarkDown, Bash, SQL
Database Software	PostgreSQL, MySQL, Hive, Hadoop
Web Frameworks	Flask, Django, FastAPI
Cloud Computing	AWS (EC2,RDS), Azure
Bioinformatics Libraries	Samtools, BCFtools, HTSlib, Tabix, Plink, Hail, Glow
Containerization	Docker
Visualization	ggplot2/tidyverse, matplotlib, Plotly
Data ETL/ELT	SQLAlchemy, Alembic, Pydantic
Source Control	git, GitHub, GitLab
Dashboard Software	Django, Shiny, Streamlit, PowerApps

Technical Experience

Data Engineer **July 2024 — Present**
Trailhead Biosystems *Cleveland, Ohio*

- Deployed a PostgreSQL 17 database on an Azure VM, and configured pgBackRest for automated daily backups.
- Designed and implemented a normalized data model using SQLAlchemy and Alembic and collaborated with cross-functional teams to integrate previously siloed data sources.
- Implemented secure Azure VM access by coordinating site-to-site VPN setup for on-premises connectivity and deploying Sophos endpoint VPN for remote access, replacing insecure public internet exposure.
- Deployed a FastAPI-based REST service using Gunicorn behind Nginx, with secure authentication via Microsoft Entra ID (Azure AD), supporting scalable and secure data access.
- Built and deployed a Django-based frontend with Gunicorn and Nginx, integrated with Entra ID for authentication, enabling secure intranet access to data entry and visualization tools.
- Migrated legacy Excel-based tracking workflows to a centralized relational database, enabling data traceability, and cross-team accessibility.
- Built interactive Streamlit dashboards to visualize key metrics and trends, enabling stakeholders to self-serve data insights.

Research Associate **May 2023 — July 2024**
Department of Population and Quantitative Health Sciences, Case Western Reserve University *Cleveland, Ohio*

- Deployed Docker containers in a HPC environment to annotate and score 362 million structural genetic variants associated with Alzheimer's disease for downstream analysis by ADSP (Alzheimer's Disease Sequencing Project) collaborators.
- Wrote custom Ansible roles and scripts to remotely automate PXE (Pre-Execution Environment) based installation of Linux OS Images (Ubuntu and RHEL) on an 18-node bare-metal server.
- Utilized PostgreSQL JDBC (Java Database Connectivity) integration with Hadoop/Spark to automate schema definition and ETL/ELT of 4 terabytes of large genomic flatfiles (VCF, BAM, CRAM) into the PostgreSQL server.
- Wrote custom Ansible roles and scripts to automated the installation of Apache Hadoop HDFS, YARN, MapReduce, and Spark across an 18-node cluster.
- Deployed bioinformatics pipelines via Docker containers within an AWS EC2 instance, connecting to an S3 bucket containing several hundred gigabytes of genomic data for data annotation.

Graduate Research Assistant **January 2023 — May 2023**
Department of Population and Quantitative Health Sciences, Case Western Reserve University *Cleveland, Ohio*

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

(973) 519-5397
Cleveland, Ohio
naveen.kannan12@gmail.com

Naveen Kannan

Data Engineer

Portfolio: Data Science Blog
github.com/naveenk2022
linkedin.com/in/naveenkannan1

Graduate Research Assistant **May 2022 — December 2022**
GIS Health and Hazards Lab, School of Medicine, Case Western Reserve University *Cleveland, Ohio*

- Developed Python scripts and YOLOv4 ML algorithms to automate the extraction of GPS coordinates from geospatial video feeds of refugee camps in the DRC after Mt. Nyiragongo eruption to generate of heat maps of tent distribution in refugee camps.

Junior Resident Doctor **July 2020 — September 2020**
Department of Psychiatry, Saveetha Medical College *Chennai, India*

- Evaluated patients and diagnosed psychiatric illnesses in an outpatient setting.

Junior Resident Doctor **March 2019 — March 2020**
Madras Medical College *Chennai, India*

- Rotated through the following departments: Internal Medicine, Pediatrics, General Surgery, Obstetrics and Gynecology, Community Medicine, Psychiatry, Emergency Trauma Ward, Labor Ward.

Education

MSc in Biomedical and Health Informatics , *Case Western Reserve University* January 2022 — August 2023
Bachelor of Medicine and Bachelor of Surgery, *Madras Medical College* September 2016 — March 2020

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.

- 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*.
https://downloads.hindawi.com/archive/2017/5247626.pdf

- 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.
https://hpbonline.org/article/S1365-182X(16)31428-9/fulltext

- Ramasamy, V., Vellaisamy, R., **Kannan, N.**, & Gounder, K. D. (2016). Refined technique of access loop in hepatobiliary surgery.*HPB, 18*, e593-e594.
https://hpbonline.org/article/S1365-182X(16)31598-2/abstract

- **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
https://www.hpbonline.org/article/S1365-182X(16)31013-9/fulltext