

NAGA SHRIKANTH AMMANABROLU

New York, NY • +91-9819045480 • naga.93@aol.com • LinkedIn • Website • GitHub

PROFESSIONAL EXPERIENCE

Memorial Sloan Kettering Cancer Center, New York, NY

Sept 2020 – Present

Data Engineer II

- Architected, implemented, and maintained a high-performance data lakehouse platform on AWS, utilizing PySpark, Glue, Athena, and Apache Hudi; transition from IBM Cloud cut data ingestion latency by 60% and realized over \$250,000 in infrastructure cost savings.
- Lead the development of the Pathology Data Mining project, constructing an infrastructure that enables data interoperability, minimizes data copies, and provides IRB-compliant access to pathology images, compute, and ML workflows on Databricks, using S3 Access Grants, IAM, and MLFlow.
- Enhanced data pipeline efficiency by setting up Glue job alerts, reducing data processing errors by 20% through timely intervention; and built Grafana dashboards for monitoring performance metrics of Hive-based tables.
- Spearheaded the creation of MSKPyDeID, an internal Python library, streamlining data de-identification processes for secure data loading in lower-level environments; mitigated PII exposure risks and ensured regulatory compliance.

Icahn School of Medicine at Mount Sinai, New York, NY

Apr 2019 - Sept 2020

Data Engineer II, Scientific Computing

- Managed the full cycle of extracting SNOMED terms from clinical notes utilizing Apache cTakes and Clinithink, processing 10,000+ patient records weekly; this initiative saved 100+ hours of manual review time and allowed for context-aware searching for IRB-approved research studies.
- Orchestrated the development and maintenance of real-time streaming HL7 data processing systems with the Iguana interface engine; executed successful prototyping of pipelines on the open-source Mirth engine.
- Piloted cutting-edge data ingestion pipelines leveraging HIPAA-compliant Microsoft Azure services including Databricks, HDInsight, and Data Factory; demonstrating streamlined data processing operations.
- Facilitated the building, maintenance, and tuning of open-source Healthcare Common Data Model systems like I2B2 and OMOP. Improved performance of the I2B2 PostgreSQL instance by 60x.

Future plc (formerly Purch Group Inc), New York, NY

Mar 2017 - Apr 2019

Associate Data Scientist (Mar 2017 - Apr 2019)

- Developed and deployed forecasting algorithms utilizing Generalized Additive Models to predict KPIs for publisher services; improved decision-making accuracy by 20%.
- Engineered and maintained BI AWS ETL infrastructure for Data Lake post-acquisition, automating data ingestion through partner network APIs and web scrapers, resulting in a yearly reduction of 7000+ man hours.
- Led the transformation of legacy data pipelines on AWS Lambda and Glue to leverage S3 and Amazon Athena, culminating in a 75% decrease in Amazon Redshift footprint, saving the organization more than \$120,000 annually.

Data Analyst Intern (Jun 2016 - Aug 2016)

- Led the 'Categorization of Purch Websites' project, developed web scrapers and designed database structures; Applied TF-IDF and Topic Modeling ML algorithms to categorize digital publications within Purch Group, optimizing content organization and streamlining operations.

EDUCATION

Georgia Institute of Technology, Remote, USA

Jan 2021 – Dec 2023

Master of Science in Computer Science

University of Illinois at Chicago, Chicago, IL, USA

Aug 2015 – Dec 2016

Master of Science in Industrial Engineering

University of Mumbai, Mumbai, MH, IND

Aug 2011 – May 2015

Bachelor of Engineering in Electronics Engineering

SKILLS

Programming Languages: Python, Java, SQL, R

Cloud Technologies: AWS, Azure, IBM Cloud

Machine Learning: Time series analysis & forecasting, Natural Language Processing (NLP), Deep Learning

Tools: Databricks, Glue, Spark, Databases (PostgreSQL, SQL Server, MySQL), Terraform, Flask, MLFlow