Anish Nair

Bay Area, CA | +13154201928 | aknair@syr.edu | linkedin.com/in/anknair

A self-starter and highly collaborative professional focusing on process improvement and strong problem solving skills with an ability to adapt to change. With experience in Big Data Engineering, seeking a Data Engineer opportunity to utilize analytical skills and present information in a dynamic environment.

EDUCATION

MS INFORMATION MANAGEMENT 2018
Syracuse University, Syracuse, NY (STEM)
GPA – 3.88 (Minor - **Data Science**)
BE INFORMATION TECHNOLOGY   2016
University of Mumbai, Mumbai, India
GPA – 3.67

BUSINESS TECHNOLOGY SKILLS

- ➢ Relational Databases
- ➢ Data Engineering
- ➢ Data Migration
- ➢ Business Intelligence
- ➢ Big Data Analysis
- ➢ Big Data Engineering
- ➢ Scripting & Automation
- ➢ Parallel Processing
- ➢ Data Visualization
- ➢ Data Architecture
- ➢ Data Investigations & Improvements

MANAGEMENT SKILLS

- ➢ Process Improvements
- ➢ Quality Assurance (QA)
- ➢ System Analysis
- ➢ Project Management
- ➢ Business Process Improvements
- ➢ Strategic Analysis & Planning
- ➢ Technical Documentation
- ➢ Quantitative Analysis

TECHNICAL SKILLS

Skill	Rating
➢ Bash/Shell Scripting	●●●●○○
➢ Scala/Spark	●●●●○
➢ R programming	●●●○○
➢ MS Excel	●●●●○
➢ AWS/GCP	●●●○○
➢ Machine Learning	●●●○○
➢ Airflow	●●●○○
➢ Python Programming	●●●●●
➢ PySpark/Hadoop	●●●●●
➢ Looker/Tableau	●●●●●
➢ SQL/Oracle	●●●●○

SIDE PROJECTS

UBER
FINANCIAL MANAGEMENT
CHATBOT
RECOMMENDATION SYSTEMS
WEB DEVELOPMENT

PROFESSIONAL EXPERIENCE

Data Engineer, *Lingo, Abbott, Bay Area, CA*   **April 2023 – Present**
- Designed and implemented a robust data ingestion system by integrating multiple source systems via APIs using Python, SQL, Spark, and Databricks, enhancing data processing
- Designed and developed a data warehouse strategy in a lake house architecture for better analytics and quicker data access to build a medallion architecture
- Implemented automated testing and DevOps practices using Databricks asset bundles, streamlining deployment processes and ensuring data pipeline reliability and efficiency.

Lead Data Engineer/Software Engineer, *FORME, Bay Area, CA*   **June 2022 – April 2023**
- Developed, Designed and Architected a modern ELT solution using **Fivetran**, **BigQuery**, **Looker** and **Airflow** to become a data driven organization.
- Developed and deployed an **algorithm** for predicting weights a user should be recommended for a specific exercise based on the user's demographic information
- Deployed **Airbyte** and **Airflow** on Kubernetes to migrate from **Fivetran** for larger data sets and reduce costs whereas Airflow for automating data jobs.

Senior Data Engineer, *Optum, UnitedHealth Group, Boston, MA*   **Aug 2021 – June 2022**
- Migrated ID management system API from an on-premise **Hadoop** ecosystem to an AWS cloud solution using **Athena**, **Glue**, **S3**, **SQS** and **API Gateway**
- Developing, deploying and testing the migration for ETL legacy pipeline to a complete cloud native tech-stack on **AWS** using **Lambda, Step Functions, Athena, Glue, SQS, SNS**
- Collaborating in a cross-functional team of 9 to ensure data quality is maintained in the effective transition of the pipeline from on-premises to the AWS cloud successfully

Data Engineer, *Optum, UnitedHealth Group, Boston, MA*   **April 2019 - Aug 2021**
- Designed, Developed and tested an API for the ID management system on Hadoop using **Scala, Spark, Phoenix** to cater for multiple pipelines and support dynamic ID generation
- Migrated legacy system from Oracle to a Big Data Ecosystem for unique ID generation in a distributed environment for new patients using **Spark, Scala, Python, Hbase/Phoenix**
- Automating Extraction Transformation and Loading for legacy pipelines by using **Bash, Python and SQL** to deliver 20 TB structured deidentified data in 20 days
- Open source contribution to PySqoop(python module) and implementing **PySqoop** into pipeline to convert data from Oracle to Hadoop Distributed File System

Associate Data Engineer, *Optum, UnitedHealth Group, Boston, MA*   **Aug 2018 - April 2019**
- Integrated Data Pipelines to migrate processes from Oracle to **Hadoop** and schedule jobs using **AirFlow** to eliminate manual intervention and improve **data processing** and quality
- Transforming data in **PySpark** and transferring 89 Mil rows from **Hadoop** to NFS mount, compressing data in **Linux** for faster transfer using **SFTP** to execute data partnerships
- Aided Data Engineers in automating **ETL** using **Python** to enhance data by increasing number of patients by transforming claims into clinical data to meet business objectives

Data Analyst, *Optum, UnitedHealth Group, Boston, MA*   **June 2018 - Aug 2018**
- Automated and developed scripts to improve **data quality** checks of monthly and weekly data by 5 days for multiple clients using **API's, Python, Bash, SQL and crontabs**
- Implemented a data quality framework which included **deidentification** of data, whitelisting data and testing the data for PHI and PII using **Python(regex)** and **SparkSQL**

Faculty Research Assistant, *Syracuse University, Syracuse, NY*   **Sep 2017 - May 2018**
- Assisting the professor to analyze **complex data** by using multiple distribution techniques and **fuzzy match algorithm** to perform disambiguation by using **SQL server**, **R** and **Python**

Business Function Intern, *Lam Research Corporation, Bay Area, CA.*   **May 2017 - Aug 2017**
- Created **Data Visualization** for reporting in **Tableau** to communicate **Business Intelligence** and **insights** to support business decisions and identify key **metrics.**

ACADEMIC PROJECTS

Natural Language Processing, nltk Movie Reviews   Python, nltk, MS LUIS
- Evaluated Movie Reviews using Python and nltk to process data for analysis
- Illustrating product review to be good or bad and deploying classification algorithms
- Designed a chatbot for an online store for training intents and utterances using MS LUIS