NIKHIL KUTTAN

kuttan.n@northeastern.edu| 857-301-5495 | www.linkedin.com/in/nikhil-kuttan

Education

Northeastern University, Boston, MA   **Sep 2021 – Dec 2023**
Masters of Science in Information Systems

Mumbai University, Mumbai, India   **Jun 2012 – May 2015**
Bachelors of Computer Engineering

Technical Skills

Programming Languages: SQL, PL/SQL, Python, Shell Scripting
Python Libraries: PySpark, NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, PyTorch, TensorFlow
Tools: Power BI, Tableau, Hitachi Pentaho, Alteryx, Talend, GIT, Airflow, Mage, ADF(Azure Data Factory), Crontab, ERP
Databases: Oracle, Microsoft SQL Server, MySQL, PostgreSQL, MongoDB, Cosmos DB
Cloud Technologies: Azure Datalake Gen 2, CosmosDB, Databricks, Synapse, KeyVault, AWS(S3, Athena, Glue, EMR), GCP

Professional Experience

Peapod Digital Labs, Quincy, MA   **Jan 2023 – Jun 2023**
Data Engineer Co-Op
- Devised single master pipeline in **Pentaho ETL** to purge data across all **5 PDL** brands, accelerating the process by **50%**
- Extracted **20 million** records from Azure Datalake by employing robust ETL pipeline with **Pyspark** on Azure Synapse performed transformation, enrichment on Business data and ingested in **CosmosDB** for application team
- Architected **Jenkins CI/CD** pipelines and developed end to end implementation of analytical initiatives for PDL brands
- Created **Kafka** topics using Confluent Kafka from preprocessed files for different business cases
- Collaborated with Data Scientist to develop **Spark** application using **Spark-SQL** in Databricks for data extraction from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns
- Suggested team adopt **ADF** to automate migration of data sources boosting execution by **30%** and fixed data silos
- Coordinated with off-shore team in different time zone & monitored daily jobs, supported QA, cross functional users

Macleods Pharmaceuticals Ltd, Mumbai, India   **Apr 2019 – Aug 2021**
Data Engineer
- Spearheaded interactive KPI trends implementation using Power BI for forecasting corporate goals and executed validations in system achieved through **PL/SQL** as per India Pharmacopoeia and US Food and Drug Administration
- Designed **Airflow** pipelines to orchestrate the Spark jobs in **Databricks** and to send job alerts via email
- Led team of 3 to implement ETL pipelines for processing data into big data system, reducing integration time by **15%**
- Built Python scripts that obtained data from APIs, transformed & ingested into **EDW** boosting execution by **25**%
- Implemented **JWT authentication** on APIs to ensure privacy, security of highly sensitive pharmaceutical datasets

In-Solutions Global Ltd, Mumbai, India   **Nov 2015 – Dec 2017**
Data Analyst
- Developed **Talend** jobs to extract data from multiple sources into data marts improving data retrieval performance
- Gathered information via **SQL** queries from multiple production databases for research and reporting purposes
- Conducted performance tuning of application stored procedures, functions, triggers enhancing performance by **50%**
- Designed and constructed TTUM (Transfer Text Upload Maintenance) on **CBS** to automate credit - debit adjustments

Academic Projects

IMDb Data Analysis: Data Warehousing & BI for IMDb **(Talend, Tableau, Power BI, Snowflake)**
- Performed **ETL** on **100 million** rows, designed **dimensional model**, scraped & cleaned data to load in **Snowflake** for analysis using **Talend** and optimized the job performance by **35%** to execute in 1 hour 30 minutes and visualized

Restaurant Inspection New York City **(Hadoop, Hive, Presto, PIG)**
- Executed MapReduce on New York City Restaurant Inspection dataset leveraging **Apache Hadoop**. Extracted data from HDFS through queries using **Hive QL**, **Presto** and **Pig** to perform transformations and aggregations to explore data