Parikshith Thriyambaka

Boston MA | 857-832-0786 | thriyambaka.p@northeastern.edu | LinkedIn | Github

Education

Northeastern University, Boston MA – MS in Data Analytics Engineering — Sep 2022 - April 2024
Relevant Courses: Big Data Management for Analytics, Machine Learning, Operations Research, Statistics — **GPA: 4.0/4.0**

SVIT, Bangalore India – BS in Computer Science and Engineering — Aug 2016 - Sep 2020
Relevant Courses: Big Data Analytics using Spark, Cloud Computing, Database Management Systems — **GPA: 3.7/4.0**

Experience

Data Engineer II (HSBC Team), Digital API Craft — Dec 2021 – Aug 2022

- Engineered comprehensive data pipelines utilizing Python, SQL, and Spark to manage over 10TB of REST API data, enhancing data modeling and ETL processes that contributed to a $15,000 revenue increase for HSBC
- Innovated real-time streaming data solutions on GCP with Dataproc, Dataflow, and Pub/Sub, processing over 10 million daily records to accelerate decision-making for HSBC UK by providing immediate insights into market trends' KPIs
- Directed the agile migration of data storage from PostgreSQL to BigQuery using cloud services, achieving a seamless transition with reduced downtime from 10 minutes to 3 seconds, ensuring uninterrupted data access
- Optimized Cloud SQL performance through strategic indexing, partitioning, and query refinement, cutting average query times by 15 seconds and significantly speeding up data retrieval processes for the business

Data Engineer (Khan Bank Team), Digital API Craft — July 2021 – Dec 2021

- Led ELT pipeline deployment on Google Cloud Platform using DBT and Google Cloud Dataflow, which doubled data throughput and slashed operational costs by $10,000, enhancing data management efficiency
- Automated data workflows with GCP's Dataflow and Cloud Composer, cutting time-to-insight from 20 minutes to 6 seconds, significantly enhancing decision-making speed at Khan Bank
- Enhanced data processing efficiency through the application of advanced data structures and object-oriented programming, reducing resource utilization by 5000 MB, thus elevating the data analysis user experience
- Conducted comprehensive unit and regression testing across 10+ critical projects, achieving 95% code coverage, which surpassed team quality benchmarks and secured data integrity for precise insights

Data Engineer Intern, BEL India — Jan 2019 – June 2019

- Implemented streaming pipeline architecture with Kafka and PySpark, integrating 30+ datasets, which doubled transaction processing capacity to 100,000 transactions per minute, significantly enhancing business agility
- Orchestrated high-performance Snowflake cloud data warehouse, achieving sub-5-second response times for ad-hoc analysis queries, thereby accelerating insights delivery and bolstering the company's industry leadership
- Developed automated data quality checks and monitoring using Apache Nifi and Python, reducing data errors by 100 instances weekly from an initial 400, markedly improving data reliability and accuracy

Data Analyst Intern, LWYD India — Jan 2018 – July 2018

- Collaborated with cross-functional teams to develop Tableau dashboards by consolidating data from over 120 sources, enhancing visibility and supporting informed decision-making across transactional services
- Utilized SQL and Python for data analysis, applying data normalization techniques that led to a 15% reduction in lead times and improved supply chain efficiency, boosting operational effectiveness
- Developed predictive models with machine learning algorithms (including random forests, logistic regression, and ANNs) to refine marketing strategies, achieving a 20% increase in customer engagement and a 15% rise in conversion rates

Skills

Programming Languages: Python (Pandas, Tensorflow, Matplotlib, Seaborn), SQL, Scala, JAVA, R, C, C++, Javascript (ReactJS)
Databases and Data Warehouses : MySQL, PostgreSQL, MongoDB, DynamoDB, Redshift, BigQuery
ETL Tools & Data Visualization: Fivetran, Informatica, dbt, Tableau, PowerBI, Apache Superset, Looker
Cloud Technologies: GCP (Dataflow, Dataproc, Bigquery, EMR) , AWS (Glue, S3, Athena, Lambda, Redshift), Docker, Kubernetes
Data Processing Tools: Apache kafka, Spark, Splunk, Hadoop (including HDFS, MapReduce, YARN), Apache Nifi
CI/CD & Agile: Apache Airflow, Git, Gitlab, JIRA, Confluence