

Parikshith Thriyambaka

Boston MA | 857-832-0786 | thriyambaka.p@northeastern.edu | [LinkedIn](#) | [Github](#)

Education

Northeastern University, Boston MA	Sep 2022 - April 2024
MS in Data Analytics Engineering Graduate Teaching Assistant Graduate Ambassador	GPA: 4.0/4.0
SVIT, Bengaluru KA	Aug 2016 - Sep 2020
BS in Computer Science and Engineering Best Outgoing Project Research Assistant	GPA: 3.7/4.0

Experience

Data Engineer II (HSBC Team), Digital API Craft	July 2021 – Aug 2022
<ul style="list-style-type: none">Led the development of a real-time data ingestion platform on GCP leveraging Dataproc, Pub/Sub and Apache Nifi, achieving an increase in data availability by 50GB per month and reducing latency by 2.5 seconds for financial transactionsImplemented CI/CD pipelines using Jenkins and Terraform, seamlessly integrated with Google Kubernetes Engine (GKE), leveraging Docker containers for efficient deployment, reducing financial data system deployment time to 30 minutesOrchestrated the migration of legacy ETL processes to Apache Spark, using agile methodologies, resulting in a 60% reduction in processing time for transactions, equivalent to saving 40 hours per cycle & enhancing system scalabilityOptimized Postgres database performance through strategic indexing, partitioning, and query refinement, cutting average query times by 70% to 15 seconds, resulting in significantly speeding up data retrieval processes for the business	
Data Engineer I (Khan Bank Team), Digital API Craft	Sep 2020 – June 2021
<ul style="list-style-type: none">Developed ML pipelines for an analytics platform using Databricks, Big Query and Big Query ML, processing 5TB of transaction data daily, resulting in a 55% increase in processing speed and annual operational savings of \$50,000Engineered a real-time data streaming system utilizing Apache Kafka and Cassandra, resulting in a daily processing time reduction of 3 hours and enabling the secure handling of over 10TB of transactional dataSpearheaded ELT pipeline management on GCP using Data Build Tool (dbt), Dataflow and Big Query which doubled data throughput and slashed operational costs by \$10,000, enhancing data management efficiencyAutomated data quality checks and monitoring solutions using Apache Airflow, resulting in a reduction of financial data pipeline failures by 100 incidents per week and enhancing data reliability for downstream processes	
Data Engineer Intern, BEL India	Jan 2019 – June 2019
<ul style="list-style-type: none">Designed high-performance schemas in 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 leadershipArchitected streaming pipeline architecture with Kafka, PySpark, and T-SQL integrating 30+ datasets, which doubled transaction processing capacity to 100,000 transactions per minute, significantly enhancing business agilityConducted 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 Analyst Intern, LWYD India	Jan 2018 – July 2018
<ul style="list-style-type: none">Developed predictive models using random forests, logistic regression, and ANNs to refine marketing strategies, increasing customer engagement by 20% and conversion rates by 15% through targeted analytics and personalized marketing effortsCollaborated with cross-functional teams to develop Tableau dashboards by consolidating data from over 120 sources, enhancing visibility and supporting informed decision-making across transactional servicesUtilized SQL and Python for data analysis, applying data normalization techniques that led to a 55% reduction in lead times and improved supply chain efficiency, boosting operational effectiveness	

Skills

Programming Languages: Python (Pandas, Tensorflow, Scikit-Learn), SQL, Scala, JAVA, R, C, C++, Javascript (ReactJS)
Databases and Data Warehouses : MySQL, PostgreSQL, MongoDB, DynamoDB, Redshift, Big Query
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, Jenkins, Git, Gitlab, JIRA, Confluence