

NIKHIL SANTHAPUR

Arlington,TX | Ph: 804-547-3858 | E-mail: nikhilsanthapur@outlook.com | Linkedin: [linkedin.com/in/nikhilsanthapur/](https://www.linkedin.com/in/nikhilsanthapur/) | Github: github.com/nikhilsanthapur

EDUCATION

The University of Texas at Arlington, Arlington, TX Master's Degree in Computer Science	Dec 2023
Jawaharlal Nehru Technological University, Hyderabad, India Bachelor's Degree in Computer Science	May 2020

COURSEWORK

Coursework: Data Structures & Algorithms, Software Development & Design patterns, Database Management Systems, Distributed Systems & System Design, Data Quality Management, Artificial intelligence, Big Data, Data mining, Data Warehousing, Machine learning, Business Intelligence & Statistics

SKILLS

Programming Languages: Python, Java, R

Databases: MySQL, Postgresql, MongoDB, MS SQL Server, NoSQL, Oracle SQL

Tools & frameworks: Power BI, Tableau, Alteryx, ETL, Informatica, Dell BOOMI, Airflow, Apache hadoop, Apache Spark, Apache Kafka, Apache Hive, Snowflake, SAS, Terraform, Matplotlib, NumPy, Pandas, NLTK, Seaborn, Pyspark, Docker, Kubernetes, LINUX/UNIX, Windows

Cloud Architecture & Services: Amazon web services(EC2,S3, Lambda,Glue, DynamoDB,EMR), Azure Datafactory, Azure Databricks, Azure DevOps

Cloud Certification: AWS Certified Solutions Architect (Validation ID: 6W26YH927EE4QLGM)

EXPERIENCE

Graduate Assistant, The University of Texas at Arlington	Jan 2022 - Dec 2023
---	----------------------------

- Analyzed 100,000 student health records using SQL Workbench in Azure Synapse analytics, developing efficient SQL queries for KPI analysis
- Designed 10+ interactive Power BI dashboards to provide insights to senior healthcare practitioners for actionable decisions based on data
- Worked on data preprocessing, cleaning & managing large, multi-modal datasets using Python libraries (pandas, NumPy, Matplotlib,seaborn)
- Cut data warehouse loading times by 50% by optimizing ELT/ETL processes with Talend for faster integration from diverse sources(XML, JSON)
- Designed and developed 12+ backend microservices based on SaaS using REST API in JSON and XML further processing it in Java and SQL.
- Troubleshoot and debug CI/CD pipelines using Git and Jenkins to ensure the quality and integrity of the code base.

Data Engineer, ShrisInfotech - Hyderabad,India	Jan 2020 - Aug 2021
---	----------------------------

- Developed a tender sourcing tool using Java SpringBoot, Hadoop improvising the existing search engine, reducing data latency by 17%.
- Contributed to data migration to Snowflake for a retail client, automating data workflows and reducing manual extraction time by 75% with optimized SQL views, streams, and Matillion job orchestration, resulting in a scalable data warehouse architecture for BI and analytics.
- Collaborated with a cross-functional team to build a customer-centric event data pipeline, involving data modeling using Databricks, Python, PySpark for front-store data analysis, driving data solutions & enhancing customer experience.
- Created customer segmentation models by managing 30+ data entities in a data lake using Databricks, Airflow & AWS.
- Worked closely with the DevOps team to improve workflow by hosting the developed AI models in Docker by wrapping them in Flask API.
- Utilized AWS EMR and Spark SQL jobs to efficiently process massive datasets stored in Amazon S3, facilitating seamless data transformation and movement between various AWS data stores and databases for data mart creation.

Intern, Risk Edge Solutions - Hyderabad,India	Sept 2019 - Jan 2020
--	-----------------------------

- Applied PyOD and Isolation Forest algorithms in Python, to analyze a dataset of 8M transactions using SQL, Excel achieving anomaly detection accuracy of **92%** and prepared a list of exhaustive data quality and standardization checks, improving 30% efficiency.
- Engineered an analytical dashboard using R-Shiny and ReactJS to display stock price updates aiding user satisfaction of 75%.
- Mastered Exploratory Data Analysis, reviewing 10 TB datasets to identify patterns, and outliers ensuring data governance.
- Designed SAP Smart Forms for invoices, credit memos, remittance advice and payment vouchers that employ financial metadata.
- Owned and composed 3 Audit Power BI Dashboards for 20+ users to track data governance metrics utilizing 2 years of data.

PROJECTS

Deciphering birth weight patterns	Dec 2023
--	-----------------

- Conducted analysis using statistical modeling in R and advanced analytics with AWS Glue and Amazon SageMaker to investigate factors affecting newborn weight, providing actionable insights for health care decision-making, improving maternal and child health outcomes.

Mobile Application for Image Analysis and Classification	Aug 2023
---	-----------------

- Developed a mobile application for binary classification of handwritten numbers. Built with Flask and Python (Agile methodologies), the model achieves an impressive accuracy of 86%. By implementing rigorous unit and integration testing, ensured seamless backend interaction between APIs and the Flask server, containerized with Docker images further facilitated efficient deployment and scalability.

Lawn Mower Simulator	Dec 2022
-----------------------------	-----------------

- Developed an application using agile methodologies for a Lawn Mower simulator using Java and Spring Boot, applying the State and Observer design patterns to manage mower behavior and update lawn visualization. Containerized microservices (Mower control, Notification) with Docker and orchestrated them with Kubernetes to achieve enhanced efficiency, scalability, and manageability.