Karan Pratap Singh

Dallas, Texas| 469-332-2869 | |[kxs220023@utdallas.edu](mailto:kxs220023@utdallas.edu) | [LinkedIn](http://www.linkedin.com/in/kps25/) | [GitHub](https://github.com/karankps25)

# SUMMARY

# Detail-oriented Data Engineer and Software Tester with over 4+ years of hands-on experience specializing in data integration and Testing. Proven track record of successfully designing, implementing, and optimizing end-to-end data pipelines. Experienced in various testing phases, including test planning, test case design, execution, and defect management.

# EDUCATION

Masters in Business Analytics and Information Systems, **The University of Texas at Dallas** **Aug 2022 – May 2024**

* Founder and President of “The Product League” club, having 192 members to date.
* Former Tech officer at theproductleague.com.

Bachelor of Technology, Computer Science, **Dr. A. P. J. Abdul Kalam Technical University** **Aug 2014 – Jun 2018**

# PROFESSIONAL EXPERIENCE



**Senior Software Engineer, Qualitest Group, India Apr 2021 – Jul 2022**

* Executed an end-to-end data pipeline, employing PySpark and Airflow to streamline data extraction, transformation, and loading processes, reducing data processing time by 30%.
* Tested the GLS Application using RPA for Inventory management in Warehouses for quality engineering.
* Developed and tested ETL data pipeline for data retrieval to extract sales data from Hive and send to target vendor for credit sales analysis, achieving a 20% increase in efficiency.
* Leveraged linux shell scripting for data seeding, deploying dummy data, and employed Postman scripts for rigorous API testing in the application.

# Software Test Engineer, QA Infotech, India Jun 2018 – May 2021

* Prepared Test Plan, test Cases, Test scripts and Test Metrics for the application as well as for the database verification based on the functional requirements, portal redesigned and test specs.
* Designed and implemented a self-serve platform using Python to perform data ingestion into Snowflake database. Visualized data lineage across various parameters using Snow SQL.
* Implemented a comprehensive solution architecture utilizing Azure Databricks, Azure Data Lake and Azure Data Factory, resulting in a 25% increase in data processing efficiency.
* Spearheaded UAT and conducted query analysis and index optimization to enhance overall database efficiency.
* Monitored and optimized the performance and scalability of data warehouse queries, identifying and addressing bottlenecks or inefficiencies in collaboration with the performance tuning team.

# Data Scientist Intern, Nirwani Technologies Pvt. Ltd., India Jun 2017 – Nov 2017

* Created complex SQL queries to generate reports associated with customers usage for their services.
* Created Python scripts to automatically address job failures, reducing manual effort by 30%.
* Developed a Tableau dashboard to understand the impact of least replenished goods.
* Ensured data quality by validating the inconsistent data and collaborate with engineering team for data reconciliation.

# SKILLS AND CERTIFICATIONS

* Professional Certifications: AWS Solutions Architect Associate, Azure Data Engineer Associate, Databricks Advance Engineer, Snowflake Hands on Essentials, Google Data Analytics,
* Programming Languages: Python, R, SQL, Linux Shell Scripting, NoSQL, Oracle PL/SQL , Hive
* Technologies: Big Data, Hadoop, HDFS, RPA, Apache Spark, Tableau, Splunk, Apache Kafka, Apache Airflow
* Software: Git, Jenkins, Docker, Microsoft Azure, AWS EC2, AWS S3, Amazon SQS, MS Excel, GCP configuration
* Frameworks & Libraries: Data Warehousing, Data Modeling, REST API, CI/CD, AWS Cloud Solution Architecture

# RELEVANT ACADEMIC PROJECTS

* **X Data pipeline using Airflow:** Data extraction from X (formerly Twitter) using Python, Airflow, AWS S3, EC2
* **Cab Fares Data Analysis:** Examined the data for Uber, Lyft rides in Boston to understand what factors affect cab prices.
* **Analyzing Powerlifting Dataset: The lift that favors winning using R:** Used ML techniques to evaluate the impact of each lift on the winning position and created data models to predict an athlete's chance of winning.