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 Data 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%.
* 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.
* Optimized a data pipeline solution by migrating from Hive QL to Spark Scala resulting in 25% faster processing.
* Developed and automated data pipelines and data stores in Snowflake: including designing, implementing, testing and debugging.

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

* Implemented a batch processing pipeline using pyspark to identify the replenishment data based on the purchase order transactions across all the channels to provide an omnichannel 360 view for a customer using Janus Graph, resulting in a 30% improvement in real-time insights.
* Designed and implemented a self-serve platform using Terraform and Python to perform data ingestion into Snowflake database. Visualized data lineage across various parameters using Snow SQL.
* Designed and developed a comprehensive solution architecture utilizing Azure Databricks, Azure Data Lake and Azure Data Factory, resulting in a 25% increase in data processing 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.
* Worked on RPA to automate the GLS for Inventory management in Warehouses under WMs for quality engineering.

# 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, T-SQL, Hive
* Technologies: Big Data, Hadoop, HDFS, 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
* **Serverless webapp using AWS:** AWS Lambda, Amazon API Gateway, AWS Amplify, Amazon DynamoDB, Cognito
* **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 data analysis techniques to evaluate the impact of each lift on the winning position and created data models to predict an athlete's chance of winning.