**Karan Pratap Singh**

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

**SUMMARY**

Data Engineer with 4 years of experience in end-to-end data and testing solutions. Experienced in data pipelines, ETL processes, database architecture, and cloud data warehouses. Demonstrated success in delivering Data Integration and Data Modeling projects, overseeing initiatives that handled 1.5 million dollars in data migrations and collaborating with teams for project execution.

**EXPERIENCE**

**Senior Data Engineer** Apr 2021 - Jul 2022

Qualitest, India

* Executed an end-to-end data pipeline, employing PySpark and Airflow to streamline data and perform ETL, 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 Apache Spark resulting in 25% faster processing.
* Implemented and automated data pipelines and data stores in Snowflake, increased performance by 15%, and reduced ETL maintenance efforts by 20%

**Data Engineer** Jun 2018 - May 2021

QA Infotech, India

* Established 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 a self-serve platform using Python to perform data ingestion of 8 TB into Snowflake database. Utilized Snow SQL to visualize comprehensive data lineage across all platforms, improving data transparency and reducing troubleshooting time by 8%.
* 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, resulting in 15% improvement in query performance and scalability.

**Data Scientist Intern** Jun 2017 - Nov 2017

Nirwani Technologies Pvt. Ltd., India

* Created SQL queries for customer usage reports, reducing report generation time by 5%.
* Automated job failure handling with Python scripts, decreasing manual intervention by 12%.
* Built power bi dashboard for least replenished goods, enhancing decision-making and reducing analysis time by 8%.
* Ensured data quality by validating inconsistent data, leading to 15% reduction in errors, and collaborated with engineering for data reconciliation, improving accuracy by 10%.

**EDUCATION**

**Master of Science (M.S.) - Business Analytics** Aug 2022 - May 2024

The University of Texas at Dallas, Texas, TX

**Bachelor of Technology (B.Tech.) - Computer Science** Aug 2014 - Jun 2018

Dr. A. P. J. Abdul Kalam Technical University, Ghaziabad, U.P

**SKILLS**

* Programming Languages: Python, R, SQL, Linux Shell Scripting, NoSQL, Oracle PL/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
* Frameworks & Libraries: Data Warehousing, Data Modeling, REST API, CI/CD.

**LICENSES & CERTIFICATIONS**

**AWS Certified Solutions Architect – Associate** 2024

Amazon Web Services - 1FFP2MQD5EE11BS9

**Azure Data Engineer Associate** 2024

Microsoft