

# KARAN PRATAP SINGH

+1(469) 332-2869 • [karan.corporate25@gmail.com](mailto:karan.corporate25@gmail.com) • [LinkedIn](#) • [Github](#) • Dallas, TX

## EDUCATION

---

### Master of Science (M.S.) - Business Analytics

The University of Texas at Dallas, Texas, TX

Aug 2022 - May 2024

### Bachelor of Technology (B.Tech.) - Computer Science

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

Aug 2014 - Jun 2018

## SKILLS

---

### Programming Languages:

SQL, Python, R, Linux Shell Scripting, NoSQL, Oracle PL/SQL.

### Libraries:

Numpy, Pandas, Dplyr, Scikit-learn, Matplotlib, Ggplot2, Beautiful Soup.

### Cloud Platforms:

AWS S3, EC2, ELB, EBS, RDS, VPC, AWS Redshift, QuickSight, Lambda, AWS Athena, ADLS, Azure Data Factory, Azure Databricks, Snowflake

### Big Data Technologies:

Hive, Hadoop, HDFS, Apache Spark, Apache Flume, Apache Kafka.

### Data Visualization / Analysis:

Alteryx, Power BI, Visio, Tableau, MS Excel.

### Version Control:

Git, Jenkins, Docker.

### Statistics & AML:

Hypothesis Testing, Logistic Regression, ANOVA, K-Means, KNN, Random Forest.

### Frameworks:

Data Warehousing, Data Modeling, REST API, CI/CD, Agile methodology.

### Certifications:

AWS Solutions Architect Associate, Azure Data Engineer Associate, Google Analytics.

## EXPERIENCE

---

### Senior Data Engineer

Qualitest

Apr 2021 - Jul 2022

Noida, India

- Executed an end-to-end data pipeline, employing PySpark various workflows, data structures and used Airflow to streamline data, adding business value and 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 and data governance.
- Optimized a data pipeline solution design by migrating from Hive QL to Apache Spark and performed user acceptance testing for system analysis under regulatory compliance resulting in 25% faster processing
- Created Tableau reports from business requirements and data strategy for business intelligence reporting. Implemented and automated data pipelines and data stores in Snowflake for various use cases and business processes, increased performance by 15%, and reduced ETL maintenance efforts by 20%.

### Data Engineer

QA Infotech

Jun 2018 - May 2021

Noida, India

- Established a batch processing pipeline using PySpark to identify the replenishment data based on the purchase order transactions across all the channels and performed root cause analysis for management reporting, 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, application design and reducing troubleshooting time by 8%.
- Designed and developed a comprehensive solution architecture utilizing Azure Databricks, Azure Data Lake and Azure Data Factory for application development, resulting in a 25% increase in data processing efficiency for proof of concept project.
- Leveraged data analysis techniques to uncover valuable business insights and informed strategic decisions. Provided production support through meticulous data maintenance and effective issue management.
- 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 Analyst Intern

Nirwani Technologies Pvt. Ltd., India

Jun 2017 - Nov 2017

Delhi, India

- Created SQL queries and provided data solutions from customer usage reports, reducing report generation time by 5%.
- Automated job failure handling with Python scripts for data flow, system analysis and data collection, decreasing manual intervention by 12%.

- Built tableau dashboard for least sales, enhancing decision-making and reducing analysis time by 8%.
- Performed data wrangling and ensured data quality controls based on mentioned use case by validating and analyzing data, leading to 15% reduction in errors, and did multi-tasking by collaborating with engineering for data reconciliation, improving accuracy by 10%.
- Produced insightful data driven business tableau reports derived from thorough data analysis and decision making. Ensured data integrity and reliability through timely maintenance and proficient issue management in production environments.

## PROJECTS

---

- **Conagra Hackathon (FMCG Data):** Led analysis of FMCG market trends, providing growth recommendations for Conagra Brands. Identified opportunities in this segment using Alteryx and revealed a previously untapped market of \$67 million, enhanced the product portfolio, and recommended effective customer-centric merchandising strategies. The Alteryx analysis prompted the introduction of an additional product for substantial weekly sales enhancement
- **YouTube Analysis Project:** Extracted JSON and CSV files using AWS CLI, transformed them into tables via AWS Glue, and conducted table joins in Athena with Hive queries. Preprocessed the data innovation using AWS Lambda for data analytics.
- **Analyzing Powerlifting Dataset:** Used data analysis techniques and assessed the impact of different lifts on powerlifting competition outcomes. The project included predictive models to synthesize data and highlighted the expertise in data analytics, statistics, and R programming.
- **Cab Fares Data Analysis :** Examined the data for Uber and Lyft rides in Boston to understand what factors affect cab prices.