KARAN PRATAP SINGH

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

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:	SQL, Python,R,SAS, SSIS, Linux Shell Scripting, SQL Server,Oracle,Mysql, Nosql.
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,GCP,Azure Data Factory, Azure Databricks, Snowflake,Informatica
Big Data Technologies:	Hive, Hadoop, HDFS, Splunk , Apache Spark, Apache Flume,Apache Kafka,HBase.
Data Visualization /Analysis:	Alteryx, Power BI, Visio, Cognos, 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 Software Engineer Apr 2021 - Jul 2022
Qualitest *Noida, India*

- Executed an end-to-end data pipeline, employing PySpark various workflows, and 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 speed of data analysis
- Optimized a data pipeline solution design by migrating from Hive QL to Apache Spark and performed debugging of migration algorithms resulting in 25% faster processing and improving system architecture.
- Implemented an ETL pipeline using AWS Lambda for event-triggered processing, AWS Glue for data preparation, and Amazon Athena for interactive querying. Enabled swift insights into credit sales trends and customer segmentation behavior, resulting in a 30% reduction in query response times for proactive decision-making.
- Implemented and automated data pipelines and data stores in Snowflake on service-oriented use cases, created test plans, and performed integration testing, increasing performance by 15%, and reducing ETL maintenance efforts by 20%.

Software Engineer Jun 2018 - May 2021
QA Infotech *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 the Snowflake database. Utilized Snow SQL to visualize comprehensive data lineage across all platforms, improving data transparency, and 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 distributed systems and application development, resulting in a 25% increase in data processing efficiency for the proof of concept project.
- Performed the Quality assurance Testing of the ETL pipeline Monitor processing latency and optimized query performance using techniques like partitioning and indexing.
- Oversaw the project documentation for Walmart's sales forecast initiative, encompassing business requirements, technical specifications, design documents, and test plans. Implemented a version control system resulting in a 20% reduction in document retrieval time, ensuring timely access to critical information for stakeholders, and enhancing project efficiency.

Data Analyst Intern Jun 2017 - Nov 2017
Nirwani Technologies Pvt. Ltd., India *Delhi, India*

- Created SQL queries and provided data solutions to the R&D team from customer usage reports, improving data quality and reducing final report generation time by 5%.
- Automated job failure handling with a scripting language like Python and created scripts for data flow decreasing manual intervention by 12%.
- Performed data wrangling and ensured data quality controls by validating data and using advanced problem-solving analytical skills and technical skills, leading to a 15% reduction in errors and improving accuracy by 10%.

PROJECTS

- **Conagra Hackathon (FMCG Data):** Led analysis of present FMCG market trends, providing growth recommendations for Conagra Brands. Identified opportunities in this segment using emerging technologies like Python and revealed a previously untapped market of $67 million, enhanced the product portfolio, and recommended effective cost-effective merchandising strategies. The analysis prompted the introduction of an additional product for substantial weekly sales enhancement
- **Stock Market Analysis** Implemented a full-stack real-time stock market analytics pipeline with stream ingestion with Alpha Vantage API, leveraging Apache Flume, Python, HDFS, and Spark RDDs for 5 stocks. (
- **Analyzing Powerlifting Dataset:** Used data analysis techniques and assessed the impact of different lifts on powerlifting competition outcomes. The project included machine learning predictive models committed to synthesizing data and highlighted the expertise in data analytics, statistics, RDBMS, and R.
- **Cab Fares Data Analysis :** Examined the data for Uber and Lyft rides in Boston to understand what factors affect cab prices using statistical modeling, business intelligence, and data quality management.
- **Truck Driver Hazard Identification :** Conducted comprehensive analysis of truck fleet data using Impala, resulting in precise transportation forecasts and the extraction of actionable insights to optimize operations. Employed Tableau for advanced data visualization, enabling the identification of patterns and the application of statistical innovation methods to identify high-risk truck drivers, thus significantly improving road safety measures