Karan Ajay Pisay

Austin, TX (Willing to relocate) | (443)-593-8409 | karanpisay.careers@gmail.com | [GitHub](#) | [LinkedIn](#) | [Portfolio](#)

Summary

Highly skilled **Data Engineer** with over 5 years of experience designing, implementing, and optimizing ETL/ELT pipelines and big data solutions. Expertise in leveraging tools like Apache Spark, Apache Airflow, and Apache NiFi for large-scale data processing and workflow automation. Proficient in cloud platforms, including AWS and Microsoft Azure, with a strong background in machine learning, data analytics, and risk management within fintech and energy domains. Demonstrated ability to drive business value through advanced data visualization using Tableau, Looker, and Alteryx, coupled with robust experience in SQL, Python, and Snowflake for scalable data storage and analysis. Adept at collaborating with cross-functional teams to deliver high-impact solutions that enhance operational efficiency and strategic decision-making.

Skills

Programming Languages: Python, R, Java, Scala, C++, JavaScript
Database Technologies: Snowflake, SQL, PostgreSQL, MySQL, NoSQL (MongoDB), AWS S3, Neptune, Hadoop, BigQuery
Statistical Analysis: Hypothesis testing, Descriptive/Predictive Analytics, ANOVA, A/B Testing, Chi-Square
Analytical Tools: Alteryx, Microsoft Excel, Tableau, VBA, Power BI, Qlik, Splunk, Tableau Prep, Looker, DBT
Machine Learning (AI/ML) Algorithms: Regression, Classification, Clustering, Decision Trees, Random Forest, XGBoost, LLM
Libraries: Pandas, Seaborn, NLTK, Beautiful Soup, Scrapy, OpenCV, Scikit-learn, PyTorch, TensorFlow, SQLAlchemy, Flask
DevOps and Tools: JIRA, CI/CD, Asana, HubSpot, Git, GitHub, Databricks, Workday, Snowflake, API, SSMS, Docker, Airflow
Big Data Technologies and Cloud Platforms: AWS, Microsoft Azure, Google Cloud Platform (GCP)

Work Experience

Tri-County Electric Cooperative, Inc.   Texas, USA
Enterprise Applications Data Analyst   Sep 2023 - Present

- Developed ETL pipelines using **Apache Nifi** and **Apache Airflow** for integrating energy consumption data enabling real-time monitoring and reducing processing times by 65%.
- Leveraged **Apache Spark** and **Snowflake** to process and store energy and finance data enhancing system performance by 45% and ensuring data accuracy for financial reporting and compliance.
- Leveraged **LookML** for Data Modelling, leading to **OLAP and OLTP** operations for ad-hoc analysis strategic decision-making.
- Enhanced financial forecasting accuracy by 40% by building data cleaning pipelines using **Alteryx** and analytical dashboards using **Looker**, enabling optimized budget allocation and resource management, and improving financial efficiency.
- Integrated data pipelines within a cloud environment, utilizing Snowflake for scalable storage and Apache Airflow for workflow automation, improving data wrangling and data normalization speeds by 25% and ensuring data compliance with GDPR.
- Collaborated with cross-functional teams for data collection and cleaning using **SQL** and **Pyspark** to ensure data integrity.

LeveragAI   California, USA
Marketing Data Analyst   Jun 2023 - Sep 2023

- Developed and deployed predictive models on **AWS Cloud** using **S3**, **Lambda**, and **SageMaker** to enhance campaign targeting and customer segmentation, improving engagement by 25% and conversion rates by 15%.
- Created comprehensive marketing dashboards and real-time performance reports with **Excel** and **AWS QuickSight**, facilitating data-driven decision-making that increased marketing ROI by 30%.
- Automated data extraction and reporting processes using **Excel** and AWS tools, reducing manual efforts by 60% and increasing the accuracy of key marketing performance metrics by 40%.
- Analyzed customer behavior and campaign performance, utilizing machine learning models to provide actionable insights that informed marketing strategies, resulting in a 20% uplift in customer acquisition and retention.
- Partnered with cross-functional teams to optimize marketing strategies and enhance customer acquisition through effective data analysis and active communication.

Triangular Automation   India
Data Engineer   Jul 2018 - Jun 2021

- Designed and implemented ELT pipelines by leveraging tools like Apache Airflow and **Apache Spark** to ingest, process, and store large volumes of customer credit data, ensured data governance, and enhanced data integration and data quality by 40%.
- Developed risk management models with **XGBoost** and other machine learning techniques, enhancing credit risk assessments by 35% and supporting strategic financial decisions with a 20% improvement in accuracy.
- Built interactive dashboards using **Tableau** to provide actionable insights into **Key Performance Indicators** (KPI), financial risks, ARR, and MRR, leading to a 30% increase in data-driven decision-making efficiency for stakeholders.
- Implemented big data solutions with **Databricks** and **Azure Data Warehouse**, optimizing data preprocessing pipelines, reducing latency by 40%, and enabling real-time analytics for risk strategy development.
- Communicated results to management, summarizing key insights and recommendations and facilitating analytical thinking.

Education

University of Maryland, Baltimore County   Maryland, USA
Master of Professional Studies, Data Science   Aug 2021 - May 2023
Pune University   Pune, India
Bachelor of Engineering, Computer Science   Aug 2016 - Jun 2019

Projects

Real-time Sentiment Analysis on Cryptocurrency Market | AWS (SageMaker, EMR, Lambda), Python, PowerBI, LangChain
- Engineered a web application to preprocess and ingest data from Twitter and Google News, achieving a 91% accuracy rate with NLTK's VADER and fine-tuned BERT Transformer.
- Leveraged AWS SageMaker, EMR, and Lambda for efficient data handling and processing.
- Deployed advanced Natural Language Processing models and dynamic dashboards using AWS S3 and PowerBI, significantly enhancing the sentiment analysis and insights into cryptocurrency market trends.

Real-time Data Pipeline for Retail Analytics | Apache Kafka, Apache Spark, PySpark, Python, MySQL (OLTP, OLAP), Tableau
- Designed and implemented a real-time data pipeline using Apache Kafka and Spark, ensuring seamless data ingestion, low-latency processing, and data storage in MySQL, resulting in a 70% improvement in data processing speeds.
- Created interactive Tableau dashboards for real-time retail analytics, enhancing decision-making, improving inventory management accuracy by 25%, and increasing overall sales by 10% through actionable insights into customer behavior and sales performance.

Automotive Customer Insights and Sales Forecasting | Snowflake, Principal Component Analysis, EDA, Jupyter Notebook
- Conducted extensive time series analysis and feature engineering using Snowflake and Prophet, delivering actionable insights into customer behavior, vehicle preferences, and market trends, leading to a significant increase in customer retention and sales performance.

Certification

PMI: Disciplined Agile SCRUM Master (DASM)
Alteryx Machine Learning Fundamentals
AWS Associate Solutions Architect
AWS Cloud Practitioner
Microsoft Essentials in Data Analysis
Certified Associate Project Management (CAPM)