Poonam Parag Thakur

Chicago, IL | (303) 408-2937 | poonamofficial98@gmail.com | poonam-parag-thakur | poonamthakur08 | Portfolio

EDUCATION

Master of Science in Data Science, University of Colorado Boulder, GPA: 3.8/4.0   **Aug 2021 - May 2023**
Coursework - Statistical Methods and Applications, Data Mining, Machine, Data Center Scale Computing, Structured Data Analysis and Modeling, Text Mining

TECHNICAL SKILLS

Languages	Java, Python, R, Typescript, SQL, Bootstrap (HTML/CSS), pySpark
Databases	PostgreSQL, Oracle, MongoDB, Amazon Redshift, Amazon S3, MariaDB, Redis, MySQL
Frameworks	NumPy, Pandas, Matplotlib, Seaborn, Scikit Learn, BeautifulSoup, Spring Boot, Angular 7, Ruby on Rails
Tools	Git, Tableau, MS Office Suite, VS Code, Jupyter Notebook, Docker, Kubernetes, AWS Glue, Apache Airflow, Kafka, GCP

WORK EXPERIENCE

Amazon   **May 2022 - Jul 2022**
Data Engineer | SQL, Redshift, pySpark, python, AWS Glue, Apache Airflow   Seattle, WA

- Collaborated with cross-functional team pipeline migration of a massive legacy & complex dataset comprising 240 million records (terabyte to petabyte scale), maintaining datasets during the transition, to facilitate a seamless shift from the deprecated old data source EDX to the Data Lake, ensuring uninterrupted data flow.
- Implemented CI pipeline to enhance data quality and consistency by utilizing AWS Glue and Apache Airflow for monitoring, alerting, ingesting, cleansing data, and executing an ETL process to update a (Data Warehouse) Redshift cluster after crawling S3 data.
- Developed a pseudonymization ETL data pipeline that combined data from variety of sources with Amazon tools and pySpark, ensuring privacy for stakeholders, and implemented a streamlined CD pipeline.
- Validated data on Redshift, EDX, and Data Lake (Andes) with Datagrip and Cradle, ensuring 98% data accuracy and completeness of data, and implemented and automated a 100% compliant pipeline, surpassing the previous solution, resulting in a streamlined workflow.

National Stock Exchange (NSEIT)   **Jun 2019 – Jul 2021**
Full Stack Software Developer | Java, PostgreSQL, Oracle, RESTful APIs, Angular, Agile, Database management   Mumbai, India

- Collaborated on secure access management and financial product development, achieving compliance accuracy, a 50% decrease in unauthorized access incidents, and adherence to regulations and standards utilizing strong cross-discipline communication skills.
- Developed a responsive Angular-based (Typescript) portal with modular, reusable, and testable design to ensure cross-browser compatibility and scalability. Enhanced functionality and performance by integrating robust RESTful APIs built with Java and Spring Boot (using Spring Tool Suite IDE) for seamless integration with backend services.
- Optimized database design by incorporating Triggers and Procedures, leveraging (Relational Database) PostgreSQL and Oracle, leading to a 40% reduction in human interaction, and a 50% process time reduction.
- Delivered advanced solutions utilizing a comprehensive technology stack, following Agile Scrum methodology and MVC architecture, achieving top performance ratings as the youngest team member for 2 years.

PROJECT EXPERIENCE

Nike Footwear Market Analysis | Python, beautifulsoup, machine learning algorithms, APIs   **Jan 2023 - Apr 2023**
- Conducted data-driven analytics by gathering data from diverse sources, including two APIs (Newsapi and Twitter) and web scraping using beautifulsoup, to extract valuable insights into customer preferences and purchasing behavior.
- Prepared data and interpreted trends and patterns using predictive modeling machine learning algorithms techniques such as clustering (Kmeans and Hierarchical), association rule mining (ARM), and Latent Dirichlet Allocation (LDA) to uncover valuable insights into customer preferences and purchasing behavior. Designed interactive dashboards to visually display valuable insights using Tableau.

Event Seating and Parking Management System | MySQL, MariaDB, tableau   **Oct 2022 - Dec 2022**
- Engineered database solutions by designing optimal Entity Relationship Diagrams (ERD), constructing robust database designs, and executing precise relational schema mapping utilizing the power of MariaDB and MySQL RDBMS technologies.
- Built Tableau dashboard to visualize core business key performance indicators (KPIs) (e.g., Monthly Recurring Revenue), saving 3-4 hours/week of manual reporting work and providing valuable business intelligence insights by establishing connections with MariaDB.

Youtube Mix Playlist Identifier | Python, NoSQL, Kubernetes, Docker, Flask   **Sep 2022 - Dec 2022**
- Engineered a highly scalable, secure web application using Flask Server and Bootstrap framework with HTML/CSS, deployed it on Google Cloud Platform (GCP) using Docker and Kubernetes with REST API, Minio, and (NoSQL or non-relational databases) Redis for seamless data storage, achieving 50% faster server response time.
- Utilized Git for version control to manage the code changes, collaborate with team members, and maintain code quality.
- Architected comprehensive services with a user-friendly Flask frontend, resulting in a seamless and efficient user experience.

Real-Time Song Sentiment Analysis | Python, NoSQL, AWS EC2, MongoDB, Plotly Dash, Kafka, Spark, Airflow   **Feb 2022 - Mar 2022**
- Built, orchestrated, and maintained optimal data pipelines, utilizing AWS EC2 (Cloud platform), Kafka, Apache Spark, Apache Airflow, and MongoDB, for real-time song sentiment analysis of high-volume streaming data and workflow management.
- Designed an interactive dashboard with Plotly Dash to display real-time sentiment analysis results, providing users with actionable insights into music trends for over 500 songs.