PRERNA SURESH
+1 (331) 2305411 | prernars14@gmail.com

Profile:

Skilled Data Engineer with expertise in designing, developing, and optimizing data infrastructure and pipelines to drive efficient data processing and business insights. Proficient in Python, SQL, and modern data technologies including Hadoop, Spark, Kafka, and Kubernetes. Experienced in building scalable, high-performance data systems, automating complex workflows, and modernizing legacy architectures. Strong background in extracting, transforming, and loading (ETL) large datasets to create actionable insights and support data-driven decision-making. Collaborative team player with a proven ability to align technical solutions with business objectives and deliver reliable, maintainable data systems. Eager to leverage technical expertise to solve complex data challenges and contribute to organizational success.

Education:

Master of Science, Data Science   **2024**
The University of Texas at Arlington, USA
Bachelor of Technology, Information Technology   **2020**
R.M.K Engineering College, Chennai, India

Professional Experience:

Tek Solve IT Solutions   Dallas, TX
Data Engineer   Jul 2024 - Present

Responsibilities:
- Designed and implemented scalable ETL workflows using **Python** and **SQL** to process patient data from multiple sources, reducing data processing time by 40%.
- Worked on real-time analytics pipeline using **AWS Kinesis** and **Python** to monitor patient vitals and detect anomalies, improving alert response times.
- Deployed the system on **AWS Lambda** for cost-effective, scalable execution, reducing fraudulent claims by 25%.
- Ensured data quality and integrity by implementing validation scripts in Python and conducting periodic audits on **SQL datasets**.
- Created interactive **Power BI** dashboards to visualize fraud patterns, enabling stakeholders to make data-driven decisions.

University of Texas at Arlington   Arlington, TX
Data Science Grader   Jan 2024 – May 2024

Responsibilities:
- Evaluated student submissions, including assignments, quizzes, and exams, ensuring alignment with grading rubrics and maintaining consistency in assessments.
- Delivered constructive feedback to students, highlighting areas for improvement and offering clarification to support their academic growth.
- Aided in the preparation and organization of course materials, while managing administrative tasks to facilitate the smooth operation of the class.
- Supported the professor in research activities, including data collection, analysis, and conducting literature reviews, contributing to ongoing academic projects

TCS   Chennai, India
Assistant System Engineer   Jan 2021 – Jul 2022

Responsibilities:
- Developed and optimized **Python-based ETL pipelines** to handle real-time data imports from multiple sources, ensuring seamless integration with Woolworths' systems and enhancing data accuracy by **30%**
- Utilized **Apache Kafka** to support high-throughput, low-latency streaming for inventory updates and order fulfillment, adjusting Kafka cluster performance for operational needs.
- Utilized **UC4** for job scheduling and workflow automation, managing complex data processing tasks and ensuring timely execution of ETL jobs across various environments.

- Automated monitoring and alerting systems with **Python** and **Azure Monitor**, reducing downtime and improving data reliability.
- Leveraged **Apache Airflow** to orchestrate real-time data processing workflows, automating complex DAGs and achieving a 15% boost in data pipeline efficiency and timeliness.
- Allocated resources effectively, including assigning tasks to team members and managing project timelines.
- Implemented **Informatica workflows** and utilized **Sqoop** for data extraction from Oracle, Teradata, and SQL Server to **HDFS**.
- Oversaw disaster recovery planning , implementing real-time backup and failover processes, contributing to **10%** improvement in data resilience and ensuring availability for critical operations
- Created and managed **SQL Script** to support database operations, including data extraction, transformation, and loading (ETL), as well as database performance tuning.
- Tracked and managed project tasks and time using **Harvest Bench**, ensuring accurate timekeeping and resource allocation for data engineering projects.
- Managed version control and collaboration using **Git**, including branching, merging, and maintaining code repositories for ETL scripts and data processing pipelines.

Trios Technology
Data Analyst Intern

Chennai, India
Sep 2019 - Nov 2020

Responsibilities:
- Supported the **development** of workflows for extracting, transforming, and loading (ETL) data from various sources into a centralized repository, ensuring accuracy and consistency in data processing.
- Utilized **Python and SQL** to clean and transform data, including addressing missing values, normalizing data formats, and enriching datasets.
- Employed **Informatica** and **AWS** tools to integrate data from various sources, providing a unified view for analysis and reporting, improving data accessibility by 20%.
- Implemented **Data quality checks** and validation processes, reducing errors by 15%, and ensured reliable data output by monitoring and resolving issues.
- Developed reports and visualizations using **Tableau**, enhancing decision-making efficiency by 15% through clear, actionable insights for stakeholders.

MRV InfoTech
Software Engineer Intern

Chennai, India
May 2018 - Jun 2019

Responsibilities:
- Assisted in building and optimizing core application features using Java and Spring Boot, collaborating with senior engineers to implement and test new functionalities.
- Integrated third-party APIs using **Python and Node.js** to enable better communication between internal systems and external data sources, contributing to a 20% increase in system efficiency and data flow.
- Developed automated test scripts with **Selenium and JUnit** to ensure feature functionality, reducing production errors.
- **Maintained and optimized the codebase**, resolving bugs and improving application speed by 10%, while supporting CI/CD processes.

Technical Skills:
- **Development & Programming:** Python, SQL, Java, C++
- **Big Data & Cloud Technologies**: HDFS, Hadoop, Spark, Kafka, Airflow, AWS, Azure, Databricks.
- **Data Integration, Containerization & Visualization**: Informatica, Oozie, SQL Server, MongoDB, Spark SQL, Docker, Kubernetes, Jenkins, Tableau, Power BI.

Project Highlights:

Unraveling Climate patterns and Forecasting futures   **Jan'24 - May'24**

- Utilized NOAA's weather datasets, addressing missing values by imputing with **TMAX/TMIN** and adjacent data points to ensure data integrity and reliability.
- Developed and implemented **ARIMA** and **SARIMA** models for annual weather forecasting, achieving an RMSE of 2.89°C to ensure predictive accuracy.
- Cleaned, transformed, and pre-processed data to meet modeling requirements, ensuring consistent and high-quality data for robust forecasting.

- Designed and deployed an interactive dashboard with an intuitive interface, enhancing user accessibility to historical and forecasted weather data for better decision-making.

Mask R-CNN for Image Segmentation   **Aug '23- Dec '23**

- Implemented the **Mask R-CNN** image segmentation model on the COCO dataset, utilizing ResNet-50 for effective feature extraction and accurate object segmentation.
- Applied data preprocessing techniques including resizing, normalization, and augmentation to the **COCO dataset**, ensuring a diverse and representative training set.
- Trained and fine-tuned the Mask R-CNN model, optimizing hyperparameters to improve segmentation accuracy by **15%**.
- Evaluated model effectiveness by comparing predicted masks with ground truth and calculating **Intersection over Union (IoU)** metrics, demonstrating Mask R-CNN's potential for applications in autonomous driving and medical imaging**.**

Analyzing and Visualizing IMDB database using SQL   **Feb '23- Apr '23**

- Utilized XAMPP (Apache, MySQL, PHP, Perl) to retrieve and manipulate data, improving data processing efficiency while developing a user-friendly interface with HTML/CSS and interactive JavaScript visualizations.
- Engineered SQL queries and integrated dynamic visualizations to display trends and patterns, enhancing user interaction and insight discovery by 25%.
- Performed data analysis on movie genres, actor popularity, and rating correlations, uncovering key insights that improved user comprehension.
- Developed and presented a visually engaging report to communicate findings, demonstrating expertise in DBMS and data visualization.

Achievements and Research:

- Published a research paper in **Journal of Xidian University** on "Criminal Record Intendance" in Vol. 14, Issue 5, 2020.
- Contributed an innovative solution at the **Smart India Hackathon** 2019, IIT Kanpur, focusing on real-world challenges and showcasing strong problem-solving and technical expertise.
- **AWS Certified Solutions Architect** with expertise in designing, deploying, and managing secure, scalable, and cost-effective applications on the AWS cloud platform.