Pavan Teja Chilkuri

pavan.teja0511@gmail.com • +1 9017454564
www.linkedin.com/in/pavantch/ • Austin, TX, US

Data Engineer

Versatile and detail-oriented professional with diverse expertise in leading informatics projects, executing data analysis. Proven track record in implementing algorithms. Highly skilled in creating innovative data visualizations to enhance model effectiveness. Proficient in leveraging multiple systems to mine and evaluate large datasets, ensuring data integrity and collaborating effectively with various stakeholders. Strong analytical abilities, with a talent for establishing guidelines, applying methodologies, and setting goals to overcome challenges. Supports both business and technical teams by performing comprehensive data analytics using SQL, MS Excel, and statistical models.

Areas of Expertise

Big Data

Data Modelling

Data warehousing

Database Management

ELT/ETL systems

Agile/Scrum Methodologies

Scalable Systems Architecture

Software & Data Lifecycle

Test-Driven Development (TDD)

Microservices Architecture

Monitoring & Troubleshooting

Team Building & Leadership

Quality Assurance & Management

Cloud Computing & Source Control

Cross-functional Collaboration

Artificial & Business Intelligence

Education

Master of Science in Information Systems | 2022

University of Memphis, Memphis, Tennessee, USA

Bachelor of Technology in Mechanical Engineering | 2018

Jawaharlal Nehru Technological University, Hyderabad, Telangana, India

Certifications

Certified | AWS Cloud Practitioner 2023

Certified | Data Analytics for Technology
University of Memphis, 2022

Technical Proficiencies

Programming & Scripting Languages:	Python, JavaScript, HTML, CSS, SQL, R, DAX
Data Analysis Libraries:	Pandas, NumPy, Seaborn, Matplotlib, TensorFlow, Keras.
Databases & Storage:	Amazon S3, PostgreSQL, MySQL, Oracle, AWS Redshift, Teradata, SAS, MS SQL Server, HBase, MongoDB, Cassandra
Big Data & ETL:	AWS Glue, Apache Kafka, AWS Data Pipeline, Apache Airflow, HDFS, MapReduce, Spark, Hive, Flume.
Web Development:	Django, Flask, React.js, Angular JS, jQuery, Redis.
Cloud Technologies:	AWS (Glue, S3, EC2, DynamoDB, Redshift), GCP.
Data Visualization Tools:	Tableau, Power Bi.
Version Control:	Git, GitHub, Bitbucket.
Workflow & Project Management	JIRA, Agile methodologies.

Career Experience

Spark It Technologies, US **2022 – Present**
Data Engineer

- Designed and executed automated ETL pipelines with Directed Acyclic Graphs (DAGs) on AWS Glue, and streamed real-time transactional data using Apache Kafka, achieving sub-second latency with peak loads over 10,000 TPS, enhancing overall data processing efficiency and scalability.
- Designed and implemented a scalable data lake on Amazon S3, segregating raw, processed, and curated data using appropriate prefixes and bucket policies.
- Ensured data durability and availability by configuring cross-region replication and versioning on S3 buckets.
- Utilized Python's Pandas and NumPy for data processing, enhancing quality and integrity
- Automated the ETL process across billions of rows of data using AWS Data Pipeline, reducing manual workload by 26% monthly and improving operational efficiency.
- Successfully migrated Oracle databases to PostgreSQL via pgLoader and established real-time replication between PostgreSQL and MySQL, ensuring synchronized analytics with zero data loss and minimal downtime.
- Played a pivotal role in cross-functional teams, bridging the gap between data engineering, data science, and business stakeholders.
- Actively participated in agile sprints, ensuring timely delivery of all project milestones.

Environment: AWS Glue, Apache Kafka, Airflow, Amazon S3, Python (Pandas & NumPy), AWS Data Pipeline, pgLoader, PostgreSQL, MySQL, Oracle, and Agile methodologies.

University of Memphis, Memphis, TN **2021-2022**
Graduate Assistant - Data Analyst

- Processed and analyzed large datasets using Python, Pandas, and NumPy, uncovering key insights to support academic research objectives.
- Designed and developed interactive data visualizations using tools like Tableau and Matplotlib, presenting research findings to faculty and peers.
- Collaborated with research teams to gather data from multiple sources, ensuring completeness and accuracy. Implemented data cleaning techniques to address missing values, outliers, and inconsistencies, enhancing data reliability.
- Mentored 250+ undergraduate students in data analytics projects, fostering a deeper understanding of the subject matter.
- Utilized Jupyter notebooks for interactive data analysis, visualization, and model prototyping. Incorporated Seaborn and Matplotlib in Python scripts for in-depth exploratory data analysis and visualization of missing value patterns.
- Leveraged TensorFlow and Keras for deep learning-based regression analyses, enhancing predictive accuracy on non-linear datasets.

Environment: Jupyter, Python (Pandas, NumPy, Seaborn, Matplotlib, TensorFlow, Keras), and Tableau

GlobalLogic Technologies Pvt ltd., Hyderabad, India **2019 – 2021**
Analyst.

- Utilized Spark within a Python environment to distribute data processing tasks across large streaming datasets, achieving a 67% improvement in data ingestion and processing speed
- Optimized complex SQL queries for large datasets, enhancing database performance and reducing query execution time by 35%.
- Integrated Power BI and Excel optimized Dax queries for both the speed and accuracy of data-driven insights, enabling high-performance data analysis and reporting, resulting in a 40% improvement in efficiency, streamlining complex data analysis workflows.
- Applied data extraction by creating Views and Tables in Power BI to facilitate report and dashboard generation.
- Developed APIs for tracking projects completion rate and QA Data.
- Implemented QA data integration into Power Bi Dashboard for stakeholders to track and analyze project completion rates, error rates, and pass percentages.
- Collaborated on code using Git for version control, code reviews, and pull requests.
- Engaged in software enhancement by identifying opportunities for optimization and implementing effective solutions for improved code performance.

Environment: Python, Git, Power Bi, SQL, Spark.

Key Projects

Impact of COVID-19 and Climatic Conditions on U.S. Vehicle Accidents | Data Analysis

- The project's objective was to determine the impact of COVID-19 lockdowns on the number of accidents that occurred in the USA.
- Utilized SAS Enterprise Miner for decision tree construction, comparative analysis, and optimal model selection.
- Used regression models and neural networks to compare and determine the best model for data prediction.

E-commerce Marketplace | Web development & Analysis

- Created a robust e-commerce marketplace with product listings, secure payment processing, and seller ratings for seamless buying and selling.
- Accomplished website audit, optimized content/structure for better user experience, increased traffic, and analyzed user behavior/conversion rates via Google Analytics for actionable insights.

Analyzed Credit Card Risk Users | Data Analysis

- Used Python libraries, including pandas, matplotlib, and seaborn, for in-depth data analysis on credit card risk users. Generated visualizations to categorize risk based on worker category.
- Used Scikit-learn for risk prediction modeling and clustering of user data. Presented findings and recommendations, enhancing risk management through actionable policies and procedures using Power BI for dashboards.