MANOJ KUDIKALA

Data Engineer

Fairfax, VA | +1(703)-755-0968 | manojkudikala6775@gmail.com | Linkedin | GitHub

SUMMARY

Data Engineer with 4+ years of experience designing and implementing data solutions across industries. Competent in managing the complete Software Development Life Cycle (SDLC) using Agile methodologies to deliver efficient, scalable data architectures. Expertise in developing complex data pipelines and algorithms using R, Python, SQL, and big data technologies such as Hadoop, MapReduce, Hive, Apache Spark, and Pig. Proficient in ETL processes with tools like SSIS and managing cloud-based data services on AWS and Azure platforms. Strong command of data analysis libraries, including NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, and TensorFlow, facilitating the creation of advanced predictive models and insightful data visualizations. Skilled in data reporting and visualization tools like Tableau, Power BI, and SSRS, enabling data-driven decision-making.

SKILLS

Methodology: SDLC, Agile
Programming Languages: R, Python, SQL, C, Java
Databases: NoSQL (MongoDB), SQL (MySQL, PostgreSQL)
Big data technologies: HDFS, Snowflake, Redshift, BigQuery
ETL Tools: Apache Airflow, Alteryx
AWS Technologies: EC2, S3, RDS, EMR (Elastic MapReduce), IAM, AWS Glue, Lambda, Kinesis, Redshift
Azure Technologies: ADF (Azure Data Factory), Azure Databricks, Azure Data Lake, Azure SQL Database
Packages: NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow
Visualization Tools: Tableau, Power BI, Quick Sight
IDEs: PyCharm, Jupyter Notebook, Visual Studio code
Operating Systems: Windows, Linux
Other Tools: Git, MS Office

EDUCATION

Master of Science in Data Analytics Engineering
George Mason University, Virginia, USA

WORK EXPERIENCE

Data Engineer, McKesson, VA **August 2023 - Current**

- Spearheaded adopting and implementing Agile methodologies, improving project flexibility and accelerating development cycles by 20%, reducing time-to-market for critical data-driven initiatives.
- Lead the design and deployment of data processing pipelines using Python, R, and SQL within the Hadoop ecosystem, including Hive, Pig, and MapReduce, achieving a 40% increase in processing efficiency.
- Refined and optimized big data applications using Apache Spark, reducing batch processing times and decreasing resource consumption, streamlining operations across projects.
- Architected sophisticated ETL processes utilizing SSIS, AWS Glue, and other tools to orchestrate data flows between SQL databases and cloud platforms like AWS and Azure, ensuring high data integration and consistency.
- Pioneered real-time data streaming solutions using Apache Kafka and Scala enhancing data feed reliability by 40%.
- Employed advanced data analytics libraries, including NumPy, Pandas, and SciPy, and machine learning frameworks like TensorFlow and Scikit-learn to perform deep statistical analyses and develop predictive models influencing business decision-making.
- Configured Amazon EC2 instances for deploying application servers and backend services, adjusting server settings to match workload demands, enhancing cost-effectiveness and scalability.
- Leveraged Amazon Athena to execute SQL queries on S3 data, streamlining data analysis and reporting tasks and reducing query response times by 20%.
- Created comprehensive data visualizations and interactive dashboards using Power BI, SSRS, and Tableau, which increase stakeholder access to and understanding of data, leading to a 25% improvement in the efficiency of data driven decision-making.

Data Engineer, Bajaj Finserv, India **February 2019 - July 2022**

- Directed the strategic integration of big data technologies such as Hadoop, Apache Spark, and Hive with SQL databases, doubling data warehousing efficiency and slashing financial data retrieval times by 50%.
- Crafted Python scripts for data ingestion and transformation, optimizing the management of complex financial datasets, which reduced processing time by 50% and enhanced data accuracy by 30% through efficient MapReduce jobs.

- Architected and sustained scalable cloud-based data pipelines using Azure and AWS, elevating data storage and computational capacity for complex financial models.
- Harnessed machine learning frameworks including TensorFlow, Seaborn, and Scikit-learn to conduct comprehensive financial analysis, resulting in 20% more accurate forecasts and a 10% rise in profit margins.
- Advanced data modeling and warehousing with innovative tools such as Snowflake and Google BigQuery, boosting financial data access and retrieval speeds by 30%.
- Constructed robust ETL pipelines, integrating diverse financial data sources into unified repositories via AWS and Azure, ensuring high data integrity and real-time accessibility for financial reporting.
- Streamlined deployment and scalability of financial applications by utilizing containerization technologies such as Docker and Kubernetes.
- Engineered real-time data monitoring systems using Apache Kafka, enabling immediate financial analysis and rapid market response, reducing error rate by 25%.
- Transformed data warehousing infrastructure by implementing Google BigQuery and Snowflake, leading to a 45% decrease in query response time and enhancing the firm's data handling efficiency.
- Administered AWS cloud technologies, including RDS, EMR, IAM, Lambda, Kinesis, and Redshift, refining data handling, processing, and security in cloud settings.
- Coordinated deploying and managing key Azure technologies like Azure Data Factory, Azure Databricks, Azure Data Lake, and Azure SQL Data Warehouse, driving seamless data integration and scalable analytics solutions.
- Developed advanced data visualizations with QuickSight and Matplotlib, delivering critical analytics that improved decision-making efficiency by 40% and enhanced strategic financial planning.

ACADEMIC PROJECTS

Graphical Insights using Generative AI

- Engineered a Streamlit interface for efficient CSV file uploads and processing of natural language queries, enhancing data interactions and accessibility.
- Executed API integrations with Hugging Face to harness LLMs like Code Llama 13B and Mistral 7B, automating Python scripting and Matplotlib visualizations.
- Deployed high-performance computing via Hopper cluster to handle datasets, achieving a 35% reduction in processing time and streamlining complex computational tasks for research efficiency.
- Applied scrum and agile practices to manage a team of six and project life cycles with YouTrack, focusing on sprint planning, user story development, and task management, while tracking progress through precise burn-down chart analysis.

Twitter Sentiment Analysis using Natural Language Processing (NLP)

- Performed sentiment analysis on over 600,000 Twitter posts using Python, leveraging libraries such as NumPy, Pandas, and NLTK for data manipulation and natural language processing, incorporating tokenization and stemming in the pre-processing.
- Applied machine learning algorithms, including Decision Trees, Logistic Regression, and XGBoost, to both TF-IDF and Bag of Words (BOW), achieving a maximum F1 score of 0.78 with Logistic Regression and TF-IDF.
- Generated interactive visualizations and word clouds to illustrate positive and negative hashtags extracted from tweets increasing analytical insights by 30%.

Image Recognition using OpenCV

- Automated the scraping of 250 images, ensuring 50 distinct shots per sports celebrity with Beautiful Soup and Selenium enhancing the accurate subsequent image processing.
- Designed the UI using HTML, CSS, and JavaScript and featured the backend using Python Flask server, enabling an SVM classifier to achieve an 85% recognition rate.
- Refined the data cleaning process for facial and eye detection, coupled with wavelet transform for feature extraction.

CERTIFICATIONS

- Certified in Azure Fundamentals (AZ-900)