MANOJ KUMAR BELLAMKONDA
DATA ENGINEER

214-831-6540 | manojkumar.b@mycvtalent.com | Dallas, TX

SUMMARY

- 4 years of experience designing and building scalable data solutions. Leveraged Cloud Data Engineering tools (Azure, AWS) to deliver Data Warehouses, Hadoop ecosystems, Big Data analytics pipelines, and Data Visualization and Reporting solutions.
- Established a real-time data ingestion pipeline using Kafka, enabling near real-time analytics for a business use case.
- Implemented Apache Airflow to orchestrate a complex data pipeline with 10 dependent tasks, resulting in a reduction in operational costs and ensuring successful data pipeline runs.

SKILLS

Programming Language:	Python, R, SQL
IDE's:	PyCharm, Jupyter Notebook
Big Data Ecosystem:	Hadoop, Hive, Apache Airflow, Apache Kafka, Apache Spark, Apache Flink, DataBricks
Cloud Technologies:	AWS (EC2, S3, RDS, Lambda, Glue, Athena, AWS Pipeline, Redshift), Azure (Azure Data Factory, Data Lake, Synapse, Blob Storage, Azure DevOps, Databricks)
Visualizations:	Tableau, Power BI, Excel
Packages & Data Processing:	NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, PySpark, Data Pipelines, Jenkins
Version Control & Database:	GitHub, Git, SQL Server, PostgreSQL, MongoDB, DynamoDB, MySQL, Snowflake

EDUCATION

Master of Science in Computer Science   **May 2023**
University of Houston-Clear Lake, Houston, TX

Bachelor in Computer Science and Engineering   **May 2020**
Gokaraju Rangaraju Institute of Engineering and Technology, Telangana, India

EXPERIENCE

Northern Trust, TX   **January 2023 – Present**
Azure Data Engineer

- Optimized data pipelines using ADF data compression and partitioning strategies, resulting in a 25% decrease in data processing time and storage costs.
- Successfully integrated Azure Data Lake Storage and Data Lake Analytics to streamline data processing workflows, resulting in a 15% reduction in data processing times.
- Implemented automated data quality checks within Databricks notebooks, reducing data inconsistencies by 15% and saving 2 hours of manual data cleaning weekly.
- Developed and optimized SQL queries for data extraction, transformation, and loading (ETL) processes, achieving an average 15% reduction in query execution time.
- Established a CI/CD pipeline in Azure DevOps, resulting in a 10% decrease in deployment failures for data pipelines.
- Reduced data pipeline downtime by 60% through robust error handling and retry mechanisms in Airflow workflows.

Vision Healthcare Solutions, India   **January 2019 - July 2021**
AWS Data Engineer

- Leveraged Databricks, a cloud-based platform, to build and deploy Spark applications for data analytics, resulting in a 20% cost reduction compared to on-premise infrastructure.
- Built and deployed ETL (Extract, Transform, Load) jobs using Glue crawlers and Spark scripts, automating data ingestion from various sources into a data warehouse, saving 10 hours of manual workload per week.
- Implemented Snowflake's User-Defined Functions (UDFs) and stored procedures for complex data transformations, achieving a 10% improvement in data processing efficiency.
- Created a cost-effective data analysis platform using Athena for ad-hoc queries on a massive log dataset, reducing data analysis costs by 50%.
- Empowered data analysts with a 15% reduction in Redshift query execution time during a major marketing campaign, enabling accelerated insights for data-driven marketing decisions.
- Utilized Power BI data quality checks to identify and resolve inconsistencies, resulting in a 30% increase in data accuracy for dashboards presented to stakeholders.
- Configured Kafka topics and partitions, achieving a 20% increase in message throughput to meet the data ingestion needs of downstream applications.
- Exploited PySpark's distributed processing power to streamline data processing, resulting in a 25% efficiency gain.

CERTIFICATIONS

- AWS Certified: Data Analytics – Specialty
- Microsoft Certified: Azure Data Engineer Associate