

MAHESH SINGH

ms562738@gmail.com | +91-9521044136

Summary

Data engineer with extensive experience at Tata Consultancy Services, specializing in the development of efficient ETL pipelines using PySpark and Apache Airflow. Achieved significant improvements in data accuracy and reporting speed through strong analytical skills and expertise in Snowflake and AWS. Focused on delivering impactful data solutions and promoting effective team collaboration.

Skills

- ETL and ELT processes
- Data pipelines and transformation
- Data modeling techniques
- Apache Airflow orchestration
- DBT for data transformation
- Kafka messaging system
- Snowflake data warehousing
- AWS cloud services
- S3 storage solutions
- Athena query service
- MySQL database management
- PostgreSQL database systems
- Oracle database solutions
- SQL Server administration
- Python programming language
- PySpark for big data processing
- Data visualization tools
- Power BI analytics platform
- Splunk for log analysis
- GitLab version control
- Bitbucket repository management
- VS Code IDE usage
- PyCharm development environment
- SSMS for SQL management
- Agile project methodology
- Cloud infrastructure management
- Data mapping strategies
- Operational Data Store (ODS)

Experience

Data Engineer | Tata Consultancy Services - Gurgaon | 05/2022 - Current

- Engineered robust E2E ETL pipelines with PySpark and Apache Airflow, successfully processing over 1TB of data daily from diverse sources, leading to a 30% increase in data accuracy and reporting speed.
- Designed and implemented real-time data processing pipelines with Apache Kafka and AWS Lambda, enabling event-driven ingestion and low-latency processing of streaming data.
- Conceived and constructed data ingestion pipelines for raw and processed data in Amazon S3, optimizing storage and reducing downstream analytic query times.
- Streamlined and automated data loading processes into Snowflake, improving data processing efficiency by 30% for scalable warehousing solutions.
- Leveraged DBT for data transformation and model creation in Snowflake, enhancing data lineage tracking from lake house to Snowflake and resulting in 30% improvement in transparency or processing efficiency.
- Containerized ETL jobs using Docker and orchestrated them via Kubernetes, ensuring scalable and resilient data workflows.
- Implemented GitLab CI/CD pipelines to automate data pipeline deployments and version control, resulting in a significant increase in release reliability.
- Implemented DevOps best practices for continuous integration, testing, and deployment of data pipelines, enhancing deployment speed and reliability.

Data Analyst | Tata Consultancy Services - Bangalore | 07/2021 - 05/2022

- Analyzed feedback from over 500 customers to identify trends and developed actionable recommendations that drove a 15% increase in customer satisfaction.

- Designed and implemented a reporting system delivering real-time insights into customer satisfaction and product performance.
- Proactively monitored system performance, resolving 95% of potential issues before impacting 1000+ customers.
- Developed DAX measures to enhance model efficiency by 30% and created calculated column to streamline data categorization and analysis.
- Utilized data modeling techniques to establish complex table relationships (1-N, N-1, N-N), enabling actionable business insights and enhancing decision-making and operational efficiency.
- Created interactive business-focused charts to reveal key insights, enabling data-driven decision-making and enhancing reporting efficiency.

Education

Global institute of technology | Jaipur | 05/2021

Bachelor of Technology: Computer science & engineering

Websites, Portfolios, Profiles

- www.linkedin.com/in/mahesh-dataeng

Certifications and Achievement

- Client Satisfaction Appraisal Letter
- Best Team Award x 2
- Google Cloud Certified Associate (GCP)
- Alteryx Machine Learning Fundamentals
- Alteryx Foundation Micro-credential