MADHU SUDHAN

Dallas, TX | +1(469)999-1480 | yasamadhu9699@gmail.com | LinkedIn

PROFESSIONAL SUMMARY

- Results-driven Data Engineer with 6+ years of experience architecting and optimizing enterprise-scale data platforms across airline, healthcare, and technology domains.
- Skilled in developing real-time and batch pipelines using Spark, Kafka, Flink, Databricks, Airflow, and dbt, and delivering cloud-native solutions on AWS, Azure, and GCP.
- Proven success in large-scale data migrations, advanced SQL tuning, and ETL/ELT orchestration, achieving 30-40% faster processing and 20-25% cost savings.
- Hands-on expertise in Snowflake, Redshift, BigQuery, Synapse, and modern data warehouses to support scalable analytics.
- Experienced in MLOps (MLflow, TensorFlow, PyTorch), DataOps, CI/CD, Terraform, and Kubernetes, ensuring secure, automated, and compliant pipelines aligned with HIPAA, GDPR, and SOC2.
- Adept at enabling decision-making through BI dashboards (Power BI, Tableau, Looker) and bridging engineering with business analytics for measurable impact.

TECHNICAL SKILLS

Programming & Scripting: Python (Pandas, NumPy, PySpark), SQL (ANSI, T-SQL, PL/SQL), Scala, Java, Shell (Linux/UNIX)

Data Engineering & ETL: Apache Spark, PySpark, Apache Flink, Kafka, Databricks, Apache Airflow, dbt, Azure Data Factory (ADF), AWS Glue, Talend, Informatica PowerCenter, Apache NiFi

Databases & Data Warehousing: Snowflake, Azure Synapse Analytics, AWS Redshift, Google BigQuery, Oracle Data Warehouse, PostgreSQL, MySQL, SQL Server, DB2, MongoDB, Cassandra, Cosmos DB

Cloud Platforms: AWS (S3, EMR, Lambda, Glue, Redshift, Athena, DynamoDB), Azure (Data Lake, Blob Storage, Synapse, Databricks, Monitor, IAM), GCP (BigQuery, Dataflow, Pub/Sub, Composer), Cloudera Hadoop

Business Intelligence & Visualization: Power BI, Tableau, Looker, Amazon QuickSight

DevOps, DataOps & Orchestration: Git/GitHub/GitLab, Jenkins, Azure DevOps, Docker, Kubernetes, Terraform, CI/CD Pipelines, MLOps (MLflow, Kubeflow)

Data Governance & Security: Data Quality, Data Lineage, IAM, RBAC, HIPAA, GDPR, SOC2 Compliance, Data Catalogs

PROFESSIONAL EXPERIENCE

Data Engineer   **Feb 2023 - Present**
American Airlines | Dallas, TX

- Engineered real-time and batch data pipelines with Apache Spark, Flink, and Kafka, enabling continuous ingestion of flight and passenger data and improving throughput by 35% for operational dashboards.
- Delivered high-availability ETL/ELT pipelines in Azure Data Factory and Databricks that integrated structured and unstructured sources, sustaining 99.9% uptime for flight operations and baggage tracking analytics.
- Increased the speed of analytics queries by 40% in Snowflake and Synapse through partitioning and clustering strategies, which provided faster access to flight schedules and passenger information.
- Reduced infrastructure provisioning times by 60% by automating deployment of Databricks, Synapse, and Blob Storage environments with Terraform, ensuring consistent and scalable airline data systems.
- Developed machine learning workflows with MLflow, TensorFlow, and PyTorch to predict flight delays, achieving 25% higher accuracy in forecasting and improving resource allocation across operations.
- Enhanced interoperability by creating RESTful APIs aligned with FHIR and HL7 standards, integrating third-party weather and booking platforms and providing near real-time updates for decision-making.
- Accelerated release cycles by 95% through Jenkins and Azure DevOps pipelines that automated testing, deployment, and monitoring of airline data workflows, minimizing production defects.
- Built executive-level Power BI dashboards that visualized ticket sales, customer satisfaction, and on-time performance, giving leaders real-time KPIs to guide operational and commercial strategies.
- Strengthened compliance with HIPAA, GDPR, and SOC2 by applying IAM policies, RBAC, and encryption in Azure, safeguarding passenger records and maintaining regulatory standards across platforms.

Data Engineer   **Mar 2020 - Aug 2021**
Optum Health Care | Bangalore, India

- Built large-scale pipelines in PySpark and Scala on Databricks to process petabyte-level claims data, cutting batch runtimes by 30% and ensuring timely insights for provider performance.
- Increased efficiency of healthcare reporting by creating a Snowflake warehouse connected with Talend and Azure Data Factory, which shortened ingestion time by 25% and centralized STAR rating datasets.
- Improved reliability of EHR and claims processing by developing ADF and Synapse pipelines, boosting throughput while reducing ingestion failures by 20% across clinical workflows.
- Strengthened fraud detection by introducing real-time streaming with Azure Stream Analytics and Kafka, which lowered detection latency by 40% and minimized financial exposure.
- Enhanced query responsiveness for operational dashboards by tuning MySQL with indexing and partitioning, achieving a 35% faster response time for compliance-related reports.
- Delivered compliance-ready visualizations in Tableau, presenting STAR ratings and risk adjustment metrics that improved reporting accuracy by 30% and guided executive strategies.

- Introduced automated data quality checks using PySpark and SQL validation rules, maintaining 99% accuracy across clinical and claims pipelines while reducing manual review cycles.
- Partnered with healthcare analysts and compliance teams to map regulatory needs into ETL workflows, ensuring data pipelines consistently aligned with HIPAA and CMS reporting standards.

Data Analyst   **May 2017 - Feb 2020**
Splunk | Hyderabad, India
- Processed more than 10TB of daily log data by building scalable ETL workflows in Informatica, NiFi, and Spark, giving downstream teams faster access to cleansed and reliable datasets.
- Migrated enterprise workloads from Teradata to GCP (BigQuery, Dataflow, Composer), which cut query times by 20% and lowered overall data warehouse costs by 25%.
- Enabled near real-time monitoring by implementing data pipelines on AWS Glue, Lambda, Redshift, and EMR, reducing system latency by 30% and improving visibility into IT operations.
- Unified data from Oracle, SQL Server, PostgreSQL, MongoDB, and APIs into BigQuery, increasing reporting accuracy and creating a single source of truth for business intelligence teams.
- Shortened deployment cycles by 40% through CI/CD automation in Jenkins and Kubernetes, which standardized releases and reduced manual intervention in analytics workflows.
- Delivered actionable insights with Power BI dashboards that visualized log trends and error events, enabling IT teams to resolve incidents 25% faster and maintain higher uptime.
- Raised confidence in reporting by introducing automated testing frameworks in PyTest and JUnit, expanding data validation coverage by 35% and reducing ETL-related errors.
- Protected sensitive log data by embedding encryption and governance controls into Hadoop and GCP environments, ensuring compliance with SOC2 and GDPR requirements.

PROJECTS
Real-Time Flight Operations Data Platform
- Designed a real-time ingestion framework with Azure Data Factory, Kafka, Spark Streaming, and Synapse, capturing flight, baggage, and customer activity with 99.9% uptime across airline operations.
- Implemented a Databricks Lakehouse architecture combining batch and streaming pipelines, enabling sub-second query performance and unified analytics at scale.
- Developed Power BI dashboards integrated with Synapse to provide live KPIs for flight control teams, while deploying ML pipelines in MLflow and TensorFlow that improved flight delay prediction accuracy by 20%.

Healthcare Risk Adjustment Analytics
- Created a Snowflake warehouse connected through Talend and Azure Data Factory, centralizing claims and clinical datasets for efficient healthcare compliance reporting.
- Automated STAR ratings analysis with PySpark-based risk adjustment models, reducing reporting cycles by 30% and ensuring timely delivery of regulatory metrics.
- Designed interactive Tableau dashboards for patient risk stratification, enhancing visibility into high-risk groups and improving compliance reporting accuracy by 25%.

Cloud Data Migration & Analytics
- Migrated 10TB+ enterprise datasets from Teradata into GCP BigQuery using Cloud Composer and Dataflow, lowering processing costs by 25% while improving scalability.
- Built real-time monitoring pipelines in AWS Glue, Kafka, and Redshift, enabling immediate visibility into system reliability and accelerating incident response.
- Automated data validation workflows with PyTest and SQL checks, improving dataset reliability by 35% during and after cloud migration.

EDUCATION
Master of Science in Business Analytics   **Aug 2021 - Dec 2023**
The University of Texas at Dallas | Dallas, TX

Bachelor of Technology in Mechanical Engineering   **Sep 2013 - May 2017**
Jawaharlal Nehru Technological University | Hyderabad

CERTIFICATIONS
- Microsoft Azure Data Engineer Associate (DP-203) - **Microsoft**
- Databricks Lakehouse Fundamentals - **Databricks Academy**
- Apache Spark Developer Associate - **Databricks**
- Google Cloud Professional Data Engineer - **Coursera**
- IBM Data Engineering Professional Certificate - **Coursera**
- Data Engineering with Python - **DataCamp**