Bhuvan Mysore Sridhar

Chicago, Illinois, USA Bhuvanmysoresridhar@gmail.com +1 3125222130 in/LinkedIn

EDUCATION

Master of Science in Information Technology and Management | Illinois Institute of Technology | 3.83

Chicago, IL | 2026

- Relevant Courses: Big Data, Cloud Technologies, Data Engineering, Data Warehousing, Project Management, Advanced Topics in Data Management, Service Oriented Architecture, Human Computer Interaction, Vendor Management.

Bachelor of Engineering in Computer Science and Engineering | JSS Science and Technology University

Mysore, Karnataka | 2023

- Relevant Courses: Big Data Analytics, Data Mining, AI &ML , DBMS, Software Engineering, Networking.

EXPERIENCE

Junior Data Analyst | Torus Solutions | Mysuru, Karnataka | December 2022 – June 2024

- Designed and maintained scalable ETL/data pipelines using SQL and Python across distributed systems, ensuring data integrity and improving data quality for enterprise decision-making.
- Delivered automated reporting and real-time Power BI dashboards, streamlining workflows, uncovering key business patterns, and accelerating project timelines by resolving major bottlenecks.
- Executed comprehensive data analysis using SQL and data visualization tools, presenting clear patterns and summary dashboards to support cross-functional collaboration within teams.

SKILLS

Programming: Python, SQL, Scala, Spark, PySpark, Hadoop, Kafka, Kubernetes, PostgreSQL

Data Engineering: Apache Spark, Hadoop, Apache Kafka, Apache SeaTunnel, ETL Pipelines, Data Warehousing, Snowflake, Docker, Airflow, AWS Glue, S3, EMR, Kinesis, EventTriggers, Firehose, lambda, IAM roles and Permissions

Cloud: AWS, Azure, Oracle Cloud, Google Cloud, CI/CD Pipelines

Big Data & Analytics: Data Visualization, Data Governance, Data Quality Management, PowerBI , Tableau, Excel

Soft Skills: Problem-Solving, Teamwork, Leadership, Analytical Thinking, Collaboration, Resourcefulness

PROJECT

Real-Time Stock & Crypto Streaming Data Lakehouse on AWS | Self Developed | Github | September 2025 - Present

- Built a real-time streaming pipeline ingesting ~1,000+ records/sec of live crypto/stock prices into Apache Kafka, ensuring low-latency delivery for analytics.
- Designed a Bronze–Silver–Gold Delta Lake architecture on AWS S3 that processed and stored millions of events daily with 99.9% schema enforcement and ACID compliance.
- Implemented complex Spark transformations (rolling averages, Bollinger Bands, Top-5 gainers) reducing query latency from minutes to <10 seconds.
- Optimized storage and compute, achieving 40% faster query performance on Delta tables compared to raw Parquet.
- Delivered a cloud-native, scalable solution leveraging AWS S3 + Delta + Spark, showcasing production-grade design principles for streaming data engineering.

Event-Driven Delta Lakehouse with Real-Time Aggregations & Data Quality (PySpark, Delta Lake) | Self Developed|Github | June 2025 - June 2025

- Built a 3-layer, event-driven Delta Lakehouse (Bronze → Silver → Gold) with 5 Delta tables and 3 streaming jobs (Bronze ingest, Silver customers, Gold CDF→agg); 15s micro-batch trigger, exactly-once via checkpoints + idempotent MERGE; fact_orders partitioned by event_date and CDF enabled.
- Delivered real-time aggregates from CDF: backfilled once from fact_orders and then maintained agg_orders_daily incrementally; verified 100% parity between fact groupings and aggregate via a reconciliation script (no drift across runs).
- Classified and quarantined non-compliant records based on validation results, implementing automatic gating mechanisms to prevent the flow of invalid data into downstream Delta Lake tables.

Serverless Lakehouse using Polars , Duckdb and Delta- rs on AWS | Self Developed|Github| April 2025 - May 2025

- Achieved ~10× faster ETL performance using DuckDB and Polars compared to traditional Spark-based processing for small-to-mid sized datasets.
- Reduced infrastructure costs by 70% by eliminating the need for managed clusters (e.g., EMR/Spark), relying instead on serverless compute via AWS Lambda and lightweight Delta Lake storage on S3.
- Enabled real-time querying on Delta Lake tables with sub-second latency using DuckDB over S3-backed Parquet files, supporting interactive analytics without standing infrastructure.

CERTIFICATIONS

Google data analytics | Coursera | 2023

Python for Data analysis | Udemy | 2025

SQL for Data Analytics | Analyst Builder | 2024

Data Engineering Bootcamp | Udemy | 2025

Cloud Big Data Masters Program | TrendyTech | 2026