BHASKAR REDDY VANTEDDU

+1 (510) 306-4868 | bhaskar07v@gmail.com | Portfolio | LinkedIn | GitHub | San Francisco, CA

EDUCATION

University of Central Missouri **AUG 2023 – MAY 2025**
Master's in Computer Science

Jawaharlal Nehru Technological University
Bachelor's in Computer Science & Engineering

TECHNICAL SKILLS

Certifications: Combined Excellence Certification - DataEngineer.io (led by Zach Wilson)

Programming & Development: Python | AWS Services | Linux/Unix | Pandas | Docker | Flask | Git | CI/CD

Data Frameworks: PostgreSQL | PySpark/Spark | Iceberg | Trino | Apache Airflow | DBT | Databricks | Snowflake | Kafka | Tableau

Cloud Services: AWS Glue | AWS S3 | AWS RDS | AWS DynamoDB | AWS Redshift | AWS EMR | AWS Lambda | Boto3 | Astronomer

AI Frameworks: Agentic AI | CrewAI | MCP (Model Context Protocol) | LLM Integration | Predictive Analytics | Agent Orchestration

WORK EXPERIENCE

N Folks Solutions (Data Engineering Intern) **JAN 2023 – MAR 2023**

- Designed a real-time pipeline for an e-commerce platform processing ~1M user events/hour from Kafka using PySpark, enabling stakeholders to track advertising campaign performance.
  - Implemented offset-limiting for fault tolerance, reducing backfill errors by 80% while maintaining stability during peak traffic.
- Engineered a medallion architecture with bronze/silver/gold layers, ensuring high-quality data by standardizing formats and normalizing structures.
  - Built streaming analytics with tumbling and session windows that ensured 99% data completeness for accurate geo-analysis.
- Developed an IP-based location enrichment system with time-expiring cache that reduced IPinfo's API calls by 70% while maintaining data freshness.
  - Reduced latency by 6 sec during peak hours and saved over $500 monthly in API costs through intelligent caching.
- Created interactive dashboards visualizing regional user engagement with real-time metrics for session duration and conversion rates, enabling immediate optimization of targeted advertising campaigns.

Morse Team (Data Engineering Intern) **MAY 2022 – JUL 2022**

- Architected an SCD Type-2 framework for historical user analysis, replacing inefficient Type-1 storage to enable granular data tracking and idempotent processing.
  - Decreased backfill time by 60% while improving data consistency, enabling analysts to access reliable historical user trends.
- Implemented binary-array encoding to represent daily user activity metrics, transforming 30 days of binary events into a single numeric value for efficient storage.
  - Reduced table storage by 30x and improved query performance by 70% through efficient bitwise operations.
- Engineered cumulative data pipelines using Trino and SparkSQL to incrementally process and store user behavior data, solving historical trend analysis challenges.
  - Orchestrated workflows with Airflow DAGs and AWS Glue to ensure reliable, daily processing of user activity metrics.
- Documented database tables, columns and views with clear descriptions and usage examples, significantly reducing onboarding time for new team members.

PROJECTS

Hyper-ADS: An Agentic AI Ad Assistant

- Architected an AI-powered recommendation system helping small businesses create timely advertisements based on local events and real-time weather conditions.
  - Constructed specialized AI agents (Analyst, Weather, Marketing) that transform raw data into actionable recommendations.
- Engineered a complete cloud-based pipeline automatically collecting data from multiple sources, processing it through AI agents, and delivering insights.
  - Implemented a custom MCP server to standardize weather data access for AI analysis and recommendation generation.
- Developed a real-time visualization system using Server-Sent Events to show users the live thought process of AI agents as they work.
  - Maintained 99.5% uptime with auto-scaling up to 1k+ users while delivering 2,100+ personalized recommendations monthly.
- Deployed a production system that processed 15,000+ events monthly, boosted client engagement by 65%, and saved businesses $400/month in ad spend.

Stocks Analysis

- Pioneered parallel data ingestion from Polygon API using 4 concurrent Spark executors, achieving 4x faster processing and 10K+ calls/minute, while streamlining ETL workflows with Airflow DAGs.
  - Slashed AWS Glue processing costs by 40% through intelligent executor allocation and orchestrating weekly sequential ticker updates with daily parallel ingestion.
- Formulated a list-based temporal schema in Apache Iceberg with date-based partitioning for efficient storage and analysis, executing comprehensive quality checks using Pytest and Chispa.

Actors Historical Analysis

- Analyzed over a century of Hollywood actors' data (1914-2021) using a year-list structure that consolidated annual metrics into array elements, reducing storage by 25%.
  - Established data quality through DBT tests (not null, unique, accepted values) and WAP pattern, preventing anomalies.
- Modeled actor performance history with SCD type-2 framework in Trino with Iceberg, enabling detailed tracking of career trajectories and point-in-time analysis.
  - Organized interactive Tableau dashboards showing actor earnings trends, genre performance, and collaboration patterns.