

Pavan Babu Bheesetti

pavanbabu2k@gmail.com | (+1) 818-332-6178 | United States | [linkedin.com/in/pavanbheesetti/](https://www.linkedin.com/in/pavanbheesetti/) | [GitHub](https://github.com/pavanbabu2k)

SUMMARY

Software engineer with 2+ years of experience building scalable data platforms, ETL pipelines, and financial data infrastructure in fintech environments. Skilled in Python, SQL, Airflow, and cloud-native systems on AWS, with experience designing high-throughput data pipelines, analytics layers, and reliability-focused production systems supporting treasury and settlement operations.

EDUCATION

University of Texas at Arlington Jan 2024 – Dec 2025
Master's in Mathematics & Computer Science (Data Science) GPA: 3.7/4.0

Relevant coursework: Distributed Systems, Cloud Computing, Machine Learning, Statistics, Data Structures, OOPS

Andhra University Jul 2018 – May 2022
Bachelor of Engineering, Electrical Engineering GPA: 3.3/4.0

WORK EXPERIENCE

Data Engineer Aug 2022 – Jan 2024
Transcend Street

- Architected AWS-based financial settlement ingestion platform (Python, SQL, EC2, S3) processing **200K–1M+** daily transactions for treasury reconciliation workflows, maintaining **99.9%** data availability during market hours.
- Designed a collateral optimization engine (liquidity waterfall + what-if simulations) using Python/C++ services and SQL models, enabling treasury desks to minimize funding usage and generating **\$100K–\$300K** quarterly savings.
- Built replayable ETL pipelines with Apache Airflow orchestration, implementing checkpointing and deterministic keys to handle late-arriving trades and historical backfills, eliminating duplicate settlements and reducing manual corrections.
- Implemented transfer pricing calculations (benchmark, security class, haircut, FX) through Python services and REST APIs with validation layers, reducing pricing disputes by **30%** and shortening resolution cycles to under 1 hour.
- Integrated Triparty and CCP settlement flows (BNY Mellon, Euroclear, J.P Morgan) using secure API and file interfaces with automated reconciliation checks, reducing manual bookings by **40–60%** and settlement breaks by **20–35%**.
- Built intraday liquidity analytics pipelines (SOD, intraday, EOD views) using SQL aggregation services and dashboard feeds, enabling faster funding decisions and lowering financing costs by **3–5%**.
- Partnered with treasury, risk, and engineering teams to build data infrastructure supporting large-scale settlement operations, enabling faster reconciliation and more reliable financial reporting.

AI Research Engineer – (Volunteer) Jan 2026 – Present
SEAR Lab

- Created an internal knowledge assistant using Google Apps Script and Google Drive API to consolidate **5,000+** research files into a centralized, searchable system, reducing manual lookup time by **60%**.
- Developing automated content processing workflows (Apps Script, Python) to convert diverse file formats into structured JSON with enriched metadata for indexed retrieval.
- Engineered semantic search with vector embeddings and ChromaDB, enabling relevance-based ranking with metadata filters and improving response accuracy by **25%**.

TECHNICAL SKILLS

- **Programming Languages:** Python, C, C++, JavaScript, Java
- **Databases & Analytics:** SQL, Power BI, Tableau, Advanced Excel, Forecasting, Oracle, PostgreSQL
- **Data Engineering:** Spark, Airflow, Kafka, Data Warehousing, Distributed Systems, ETL/ELT
- **Cloud & DevOps:** AWS, Docker, Kubernetes, Jenkins, GitOps, CI/CD, Monitoring & Observability
- **AI/ML:** TensorFlow, CNN, LLM, AI models, TPU, LangChain, ChromaDB, Ollama, Pandas, NumPy
- **Tools & Methodologies:** Agile, Scrum, Jira, JUnit, dbt, Linux, Git, Databricks

RESEARCH & PROJECTS

Capstone Project Architect – (ASPIRA AI) Aug 2025 – Dec 2025
University of Texas at Arlington

- Developed a career decision support system that matches users to suitable roles based on profile and behavioral signals, using Azure Data Factory, Databricks, Python, SQL, and PySpark.
- Aggregated multi-source user and career datasets, performing cleaning, deduplication, and feature transformation to create training-ready data and increase recommendation accuracy by **33%**.
- Built predictive recommendation models combining semantic similarity and structured scoring, evaluated with Precision@3, nDCG, and MRR, achieving **0.80** Precision@3 for data-driven career decision support.
- Developed KPI dashboards and reports in Azure Synapse to track model performance and recommendations across **50+** scenarios.

RAG DocuChat – Machine Learning

- Designed a document processing pipeline for PDF parsing, text normalization, and semantic chunking with controlled overlap, generating **384-dimensional** embeddings using Sentence Transformers (PyTorch, all-MiniLM-L6-v2).
- Achieved **85%+** retrieval relevance and consistent context preservation, validated through grounding checks and retrieval accuracy evaluation across diverse document queries.

CERTIFICATION & PUBLICATIONS

- **AWS Certified AI Practitioner:** Validated AI/ML and generative AI concepts with AWS services (SageMaker, Bedrock).
- **Research Publication:** Built LSTM models achieving **92%** forecasting accuracy on historical Spotify datasets.