Pratyusha Adibhatla

✉ pratyushaadibhatla3@gmail.com ☎ +1 (682) 374-3740 in LinkedIn GitHub

Experience

Research Assistant - Data Engineer, *UCSD Wireless Communication, Sensing, and Networking Lab* *San Deigo, CA*

- Architected an Azure-native ingestion framework using Python and SQLite to manage HDF5, CSV, DAT, and PCD datasets stored in Azure Blob Storage, reducing discovery time to under 3 minutes through indexed metadata queries.
- Engineered automated batch ETL workflows using Azure Pipelines to orchestrate metadata extraction, schema validation, and data quality checks for 5GB+ datasets per batch.
- Implemented data lineage and feature retrieval systems within the Azure ecosystem to enable version-controlled querying and downloading of datasets with comprehensive change tracking.

Research Assistant - Data Engineer – Retrieval Systems, *University at Buffalo* *Buffalo, NY*

- Engineered a vector search data pipeline to ingest and structure unstructured data (video, image, text) by converting video to transcripts, images to OCR text, and documents to cleaned text for low-latency retrieval utilizing RAG.
- Optimized data indexing workflows in ChromaDB by designing metadata schemas, regenerating embeddings, and refreshing vector indexes when new data was added to ensure consistency across modalities.

Student Assistant – Data Engineering & Analytics, *University of Texas at Arlington* *Arlington, TX*

- Architected a structured SQL database(Star Schema) to replace manual tracking sheets, implementing strict referential integrity constraints across 10+ tables to eliminate data redundancy and ensure consistent inventory reporting.
- Engineered Python-based batch ETL jobs to ingest and sanitize raw inventory data, implementing custom logic to handle schema variations and null values from legacy Excel inputs, automating what was previously a manual data entry workload.
- Developed complex SQL scripts using window functions to aggregate historical usage data across 200+ facility items, optimizing query performance to generate accurate demand forecasting reports for inventory procurement planning.

Programmer Analyst (Data & Application Support), *Cognizant (Client: Liberty Mutual Insurance)* *Coimbatore, India*

- Supported production batch jobs processing 100k+ transactions per month across UAT and Production environments hosted on AWS EC2, ensuring availability during business hours and resolving 15–20 weekly job failures through root cause analysis.
- Investigated upstream data stored in Amazon S3 by identifying schema mismatches, null values, and unexpected file changes; triggered and monitored AWS Lambda-based workflows to validate successful batch execution.
- Developed and maintained SQL queries to update and validate claims, payments, and policy data, performing source-to-target reconciliation for 200–300 financial transactions per batch to ensure data accuracy.
- Created runbooks and job monitoring dashboards documenting schedules, validation steps, execution logs, and recurring failure patterns to improve operational visibility and support efficiency.

Programmer Analyst Trainee *Cognizant* *Coimbatore, India*

- Built a Spring Boot (Java) + Angular full-stack application to display bus routes and live location status for public transport users, supporting 5–10 active routes in test and demo environments.
- Designed a relational database with 10+ route fields and REST APIs to handle 100–300 daily location updates, implementing GPS validation and normalization to reduce inconsistent records by 15–20%.

Projects

Job Recommendation System (LinkedIn-style)

- Built batch ETL pipelines in Apache Spark using DataFrame APIs, window functions, and Parquet to transform job posting and user interaction data into analytics-ready datasets.
- Implemented a batch recommendation workflow in Spark combining feature engineering and Spark ML components to generate Top-N job recommendations for users with limited interaction history in standalone mode.

Technologies

Languages: Python (Advanced), SQL (Advanced), Java.
Data Engineering: ETL/ELT Pipeline Design, Data Modeling (Star/Snowflake Schema),Azure Pipelines, Docker, REST APIs, Apache Spark / PySpark (academic& self-directed projects).
Databases & Vector Stores: MySQL, SQLite, ChromaDB.
Data Science & Tools: Pandas, NumPy, RAG, Git/GitHub, CI/CD, Power BI.

Education

Masters in Data Science | *University of Texas at Arlington Arlington, USA*

- Relevant Coursework: Machine Learning, Data Science, Data Analytics, Database Systems, Data Mining, Neural Networks, Probability and Statistics.

BTech in ECE | *Andhra University Visakhapatnam, India*