Chaitanya Yadavalli

Data Engineer | PySpark · Databricks · AWS

ychaitu2@gmail.com | LinkedIn

PROFESSIONAL SUMMARY

Data Engineer with 2+ years of experience building scalable ETL pipelines and modernizing cloud data platforms for IoT telemetry systems. Expertise in PySpark, Databricks, AWS, and SQL. Proven track record in reducing data costs by 83%, cutting latency by 30%, and delivering zero-downtime migrations. Adept at automating data workflows, optimizing performance, and mentoring engineering teams.

WORK EXPERIENCE

Data Engineer   **July 2023 - Present**

HP   **Bangalore**

- Consolidated 40+ legacy pipelines into a unified PySpark framework, reducing operational overhead 75% and improving efficiency 15%.
- Automated onboarding with LLM-generated PySpark ETL, cutting setup time 2 days → <0.5 days.
- Reduced processing costs by 83% and improved runtime 90% via incremental data processing.
- Migrated Redshift → Databricks Unity Catalog with zero downtime, reconfigured 30+ Looker dashboards, and implemented IaC with Terraform.
- Optimized firmware update performance by 30% through Splunk telemetry analysis.
- Mentored 5+ interns in SQL and PySpark, enabling full productivity within one month.

Data Engineer Intern   **February 2023 - June 2023**

HP   **Bangalore**

- Built Databricks cost monitoring system with 95% accuracy, cutting spending by 10%.
- Developed CommonUtils library, reducing code duplication by 40% and improving maintainability.

PROJECTS

DocuBot – LLM-Powered Notebook Documentation Tool (Internal)

- Built an AI-powered assistant that automatically generates structured documentation for Databricks notebooks.
- Utilized Large Language Models (LLMs) to interpret code cells, extract logic, and produce human-readable explanations
- Enhanced maintainability and reduced onboarding time for data science and analytics teams.

Full-Stack News Website (Personal)

- Designed a decoupled architecture using Next.js (frontend) and Node.js (backend) hosted on AWS EC2 with serverless Lambda admin functions.
- Integrated GCP APIs to enrich content with real-time social media metadata.
- Executed AWS regional migration to improve availability and reduce latency(ap-south-1 to ap-south-2).

SKILLS

Programming Languages: Python, SQL, JavaScript, C++
Big Data : Apache Spark, PySpark, Databricks, Airflow
DevOps : Terraform, Docker, Git, Linux
Cloud : AWS (S3, EC2, Redshift, Lambda, IAM), Azure Data Factory
Data : ETL/ELT, Data Warehousing, Data Modeling, Data Quality

CONTRIBUTIONS & RECOGNITIONS

- HP Quarterly Award – Recognized for cost reduction and platform optimization.
- Founded HP's Data Engineering Community – Led seminars for 50+ engineers.

EDUCATION

B.Tech - Electronics and Communication Engineering   **2023**
SASTRA Deemed University   Thanjavur, TN
Relevant Coursework: Database Systems, Algorithms, Data Structures, Computer Networks