Praneeth Chavva

(801)-462-4151 • praneethchavva06@gmail.com • linkedin.com/in/praneethchavva/ • github.com/praneeth0810

Data Engineer with **3 years experience** building **cloud-native data pipelines** on **AWS** and **GCP**, handling **1TB+ data/month** and **15K+ events/sec** for real-time and batch use cases. Skilled in **Airflow**, **Glue**, **Spark**, **Kafka**, and **dbt**, with focus on **pipeline reliability**, **data quality (SCD2, Great Expectations)**, and **CI/CD**. Known for **cross-functional collaboration**, owning end-to-end workflows, and mentoring engineers to build systems that reduced data latency by 70% and enabled real-time business decisions.

SKILLS

Tech Stack: Python, SQL, Airflow, Glue, Spark, Kafka, dbt, Hive, Git, Bash, Python APIs, CI/CD workflows
Data Modeling & Quality: Star/Snowflake schemas, SCD Type 2, schema versioning, dbt tests, Great Expectations
Cloud & Database: AWS (S3, Glue, Redshift), GCP (BigQuery, Composer), MySQL, PostgreSQL
Visualization: Tableau, Streamlit, Power BI
Certifications: AWS Certified Data Engineer – Associate

PROFESSIONAL EXPERIENCE

University of Utah, Salt Lake City, Utah — January 2024 – April 2025
Research Assistant – Data Engineering

- Resolved slow dashboard loads by refactoring Tableau + dbt models on PostgreSQL, applying **STAR/Snowflake schema** and **indexed views**—cut query time **by ~35%**.
- Built **modular Airflow DAGs** with schema drift detection, retry logic, and Slack alerting, improving resilience during source changes.
- Engineered **SCD Type 2 logic** via **dbt snapshots** to maintain **historical features** across Machine Learning retrains—ensured experiment reproducibility.
- Automated data validation using **Great Expectations + dbt tests**, reducing manual QA and boosting data reliability for research.
- **Collaborated with** 5+ ML researchers to define fairness metrics and refactor feature engineering pipelines, cutting **data prep by 40%** per iteration.

Cognizant Technology Solutions, Hyderabad, India — January 2021 – June 2023
Data Engineer

- Owned 12+ Airflow DAGs for batch ETL pipelines (**~100K records/day**) with dynamic task mapping, retries, and SLA monitoring—**reduced failures by 30%.**
- Tuned Spark and AWS Glue jobs **processing 1TB+/month** by tuning executor memory, partitioning, and shuffle strategy, **cutting job runtime by ~17%**.
- Improved Kafka streaming pipeline **(15K+ events/sec)** by optimizing **partition keying and consumer group tuning**, reducing lag and message drop rate.
- Built internal Python APIs to expose Airflow DAG metrics (failures, retries, duration), helping QA team **cut debug time by 32%.**
- Embedded Great Expectations in pipelines for null checks, row count validation, and schema conformance, **reducing QA escalations by 40%.**
- **Added unit/integration tests** for dbt + Python transforms using pytest and mocked inputs, ensuring regression coverage.
- Collaborated with Product Managers and Data analysts to redesign Power BI schemas, shifting to denormalized views and **improving dashboard refresh time by ~ 70%** and reducing manual joins.
- **Mentored** 3 junior engineers on Airflow, Spark optimization, and CI/CD best practices, boosting team code quality and release confidence.
- Set up CI/CD with GitHub Actions + pre-commit hooks for DAG tests and staging deploys—**cut manual effort by 60%.**

PROJECTS

Uber Trip Data Analytics Pipeline - BigQuery, Composer, Looker Studio (Personal Project)

- Built a GCP pipeline using **BigQuery** and **Composer** to process 1M+ trip records. Modeled with **STAR schema**, Enabled daily revenue forecasting and anomaly detection on tips and surge pricing via **Looker** dashboards.

Health Analytics-Apache Spark, Hadoop, Hive, AWS S3, Predictive Analysis (Academic Project)

- Processed 500K+ health records via **Spark on Hadoop**, engineered predictive models for patient volume, and managed **resource provisioning** on AWS S3. Navigated **HIPAA-compliant storage constraints** and **batch scheduling challenges** with Hive—emulating near-prod scalability in academic context.

EDUCATION

University of Utah, Salt Lake City, UT — August 2023 – April 2025
Master of Science in Computer Science, GPA: 3.8/4.0

Vellore Institute of Technology, India — July 2017 - April 2021
Bachelor of Technology in ECE | Specialization in IoT and Sensors