Akshay Kohchade – Data Engineer

Professional Summary

Result-oriented Data Engineer with experience in building scalable batch ETL pipelines, real-time data ingestion, and production support in healthcare data environments. Skilled in PySpark, Databricks, and AWS Glue with strong exposure to schema validation, data quality checks, and Delta Lake transformations. Recently extended capabilities to include observability tooling, instrumentation using OpenTelemetry, and Grafana/Prometheus-based monitoring. Proficient in SQL, Python, and data processing frameworks, with an interest in applying ML techniques for anomaly detection and SLA forecasting. Experienced in working with Google Cloud Platform (GCP) and BigQuery for scalable analytics.

Technical Skills

- **Languages:** Python, SQL
- **Big Data Frameworks:** Apache Spark, PySpark, Databricks, Apache Kafka, Apache Airflow
- **ETL Tools:** AWS Glue, Azure Data Factory, DBT, OpenTelemetry (OTel)
- **Monitoring & Observability:** Grafana, Prometheus, Azure Monitor, Splunk
- **Cloud Services:** AWS (S3, Glue, Redshift), Azure (ADF, Monitor), Google Cloud Platform (GCP)
- **Databases:** MySQL, Redshift, Delta Lake, BigQuery
- **Machine Learning:** Anomaly Detection, Forecasting (Basics)
- **Workflow & Integration:** Data Validation, Schema Management, RCA
- **Others:** Power BI (Integration & Validation Support), Git

---

Work Experience

Data Engineer – LogIQ Labs Private Limited | Bengaluru, Karnataka
May 2025 – Present

- Provided production support for UnitedHealth Group (UHG) by proactively monitoring FHIR-based healthcare data pipelines and Delta Lake tables.
- Ensured data availability via SLA monitoring using **Azure Monitor**, while handling ingestion failures, schema mismatches, and alert investigations.
- Recently implemented **OpenTelemetry instrumentation** for tracing ingestion latency and added **Grafana dashboards using Prometheus metrics** to visualize job performance.
- Coordinated with Power BI and analytics teams to resolve data refresh issues and validate healthcare metrics.
- Introduced **anomaly detection logic** for schema drift and delayed job failures.
- Gained experience using **Splunk** for log monitoring and advanced troubleshooting.

Data Engineer – Zetaion System Private Limited | Bengaluru, Karnataka
Dec 2022 – Apr 2025

- Developed batch ETL pipelines using PySpark in Databricks to process healthcare data.
- Used **AWS Glue** for orchestration, staging in S3, and loading to **Redshift** and Delta Lake.
- Applied business rules for deduplication, schema checks, and transformation.
- Built internal Grafana-based dashboard for pipeline monitoring, leveraging **Prometheus exporters** from Glue job metrics.
- Integrated **OpenTelemetry** into Spark jobs to trace delays and failures in real time.
- Participated in a PoC using **anomaly detection models** (Python + scikit-learn) to forecast SLA breaches and alert fatigue reduction.
- Worked on **Google Cloud Platform (GCP)** to orchestrate batch pipelines and leverage **BigQuery** for ad hoc analytics.
- Built and scheduled pipelines using **Apache Airflow** to coordinate GCP and AWS data workflows.

GIS & Data Engineering Intern – PST IT Solutions | Aurangabad, Maharashtra

Feb2020–Oct2022

- Contributed to TULIP GIS project by creating maps from complex spatial data for urban planning.
- Supported ETL development and monitoring for pipeline integrity and system health.

Key Projects

Production Support for Healthcare Data Pipelines (Optum – UnitedHealth Group)

- Monitored Delta Lake tables built on FHIR and JSON structures using **Grafana** dashboards integrated with **Prometheus**, **Azure Monitor**, and **Splunk**.
- Validated record completeness, handled schema mismatches, and coordinated Power BI reporting.
- Integrated **OpenTelemetry** into ingestion pipeline for latency and RCA tracking.
- Built an **anomaly detection module** (PoC) for identifying silent data issues and schema drift.

Cloud-Native Data Orchestration Pipeline

- Used PySpark and AWS Glue to automate ingestion and transformation.
- Enforced business rules, schema validations, and consistency checks in Delta Lake.
- Leveraged **Prometheus + Grafana** for monitoring ingestion durations and alerting on failures.
- Explored **AWS IoT Core and Kinesis** for real-time data ingestion (20% of data volume).
- Supported the integration of **ML-based early warning alerts** into the monitoring layer.
- Built batch pipelines on **GCP**, leveraging **BigQuery** for analytics and **Apache Airflow** for scheduling and orchestration.

Education

Bachelor of Science (B.Sc.)
Dr. Babasaheb Ambedkar Marathwada University – Aurangabad, Maharashtra
Percentage: 64.78%

Additional Highlights

- Hands-on setup of **Grafana dashboards with Prometheus exporters** for Spark/Glue jobs.
- Trained in **OpenTelemetry** basics and integrated into job-level instrumentation.
- Built simple **anomaly detection Python scripts** to predict job delays and ingestion failures.
- Experience working on **Google Cloud Platform (GCP)** using **BigQuery** and **Apache Airflow**.
- Gained familiarity with **Splunk** for log-based insights and incident resolution.
- Collaborated with cross-functional teams including BI analysts, product, and ops for RCA and SLA compliance.
