

Rajashekar Raja

https://www.linkedin.com/in/rajashekar-raja-16b197169/

630.408.6699
rajashekar_raja@yahoo.com

Senior Data Engineer

Senior Data Engineer | Spark · Kafka · Airflow · AWS · GCP · Databricks

Senior data engineer and acting tech lead with 13+ years building production data platforms on Spark, Kafka, Airflow, AWS, and GCP, including Databricks for analytics and ML data. Recent work: Snowplow-style behavioral pipelines (enrichment, incremental processing, data quality checks and alerting), Airflow orchestration at scale, and OpenTelemetry observability with cost-aware trace sampling. Strong in data quality (Great Expectations), governance-sensitive workflows, and partnering with product and ML teams on reliable datasets. Hands-on IC who mentors and sets standards; seeking senior/staff data engineering roles.

Areas of Expertise:

- Behavioral event pipelines & Snowplow-style analytics ingestion
- Spark, Kafka, Airflow on AWS/GCP/Databricks
- Data quality, governance, and reliable consumer datasets
- Observability (OpenTelemetry) and production pipeline operations

Technical Skills

Core:	Python, SQL, PySpark, Apache Spark, Apache Kafka, Apache Airflow
Cloud:	AWS (S3, EMR, Lambda, Glue, ECS), GCP (GCS)
Platforms:	Databricks, Delta Lake
Quality / Ops:	Great Expectations, OpenTelemetry, Grafana, CloudWatch, Prometheus
Cloud / Containers:	Kubernetes, Docker, GCP (GCS), AWS (EC2, EMR, ECS, Glue, Athena, S3, Lambda)
AI Tools:	Cursor/Codex, GitHub Copilot, custom AI agent skills, GenAI pipelines

Work Experience

EVERYDAY HEALTH GROUP, OR (Remote) - Full-Time 2024 - Current

Senior Data Engineer (Acting Tech Lead)

As acting Data Engineering Tech Lead, led modernization of the open-source Snowplow pipeline with seamless version upgrades. Built enrichment data-quality checks with alerting so teams can correct issues quickly. Designed frameworks and enforced best practices across Airflow and AWS-based ingestion workflows. Collaborated with product and ML teams to deliver business-critical training datasets.

- Led Snowplow pipeline modernization and Airflow DAG re-architecture, improving performance and scalability for real-time behavioral tracking.
- Implemented incremental processing with lookback for late-arriving events and validation on enrichment failures, reducing silent reporting issues and speeding remediation (PySpark/Python).
- Created and enhanced AI Skills for team-wide use, raising code quality and shortening time to production.
- Enhanced the GenAI Community Insights data collection platform with moderation and interest-topic support, improving clustering accuracy.
- Designed a Grafana token-usage dashboard using OpenTelemetry and custom CloudWatch metrics, giving real-time visibility into token consumption.
- Improved observability for the multi-agent chatbot with OpenTelemetry metrics, strengthening token-usage tracking across the application.
- Built data quality and validation pipelines for Snowplow enrichment failures on web and mobile (Python/PySpark), isolating bad events and speeding up bug fixes.
- Built AWS Lambda functions for real-time ingestion and reinforced data governance practices to support GDPR compliance.

ACTIVISION SHARED SERVICES, OR (Remote) - Full-Time

2022 - 2024

Senior Data Engineer

Core member of the data platform team: ingestion APIs, data quality, observability, and governance-aligned pipelines for high-volume player-related workflows.

- Designed and executed OpenTelemetry trace implementation across data platform services, improving application visibility and speeding issue resolution.
- Implemented dynamic sampling for trace collection, reducing storage cost in the collector.
- Built a data quality library with Great Expectations and PySpark, automating tests on static data via Apache Airflow so consumers receive reliable data for modeling.
- Developed Type 1 and Type 2 datasets to store and manage data quality results, preserving audit history in GCS and the Databricks catalog for end-user analysis.
- Migrated the Ingestion API onto Istio service mesh with Argo CD rollouts, improving request control/filtering and enabling zero-downtime deployments.

NIKE INC, Beaverton, OR (Contract)

2017 - 2022

Senior Data Engineer

As an independent engineer, played an important role in facilitating the data acquisition process for the data science team. Used advanced big data technologies and tools in data retrieval, transformation and delivery, directly contributing to success of machine learning initiatives. Optimized data pipelines running on Apache Airflow to finish data processing within SLA timelines. Modernized existing pipelines by rewriting jobs with unit test cases and canary checks.

- Built efficient data processing pipelines delivering high-quality training data for recommendation engines on the Nike app.
- Used Apache Spark and Python on Databricks to process data efficiently and meet ML training delivery timelines.
- Developed a static aggregated sales estimation table in PySpark to help analysts plan store inventory from four years of sales history.
- Used Apache Airflow to schedule ingestion, processing, and aggregation jobs loading to S3 for Tableau dashboards.

VISA INC, Foster City, CA (Contract)

2016 - 2017

Data Engineer

Led migration of current data processing system to Spark with Scala, spearheading both onsite and offshore teams in planning and executing backlog tickets through Scrum ceremonies. Orchestrated design of system and delegated tasks to ensure timely delivery of products to stakeholders. Facilitated onboarding of new team members and provided mentorship to expedite integration into project deliveries. Proactively addressed challenging issues and alleviated blockers by implementing effective solutions, fostering collaborative and productive work environments.

- Migrated data processing jobs from DB2 to Spark, improving performance and reducing cost.
- Created RDD lookups and broadcast joins to minimize network data transfer and speed jobs.
- Implemented real-time ingestion with Kafka, demonstrating how streaming improved downstream products and user recommendations.

Education, Bengaluru - KA, India

2004 - 2008

- **Bachelor of Engineering (BE)**, Electrical and Electronics, Visveswaraya Technological University.