Priya Raman | Senior Data Engineer | Cloud Data Platforms & Real-Time Pipelines

2210 Fremont Avenue N, Seattle, Washington, United States, 98103
(206) 555-0182 • priya.raman@outlook.com • https://www.priyaraman.dev
@priyaraman • @priyaraman • @priya-raman

Basics

Data engineer with 7+ years of experience designing, building, and operating batch and streaming data pipelines. Skilled in cloud-native data platforms, data modeling, and data quality. Proven track record of translating business requirements into reliable, scalable data products.

Education

Seattle Polytechnic Institute **https://www.seattlepoly.edu**
Master, Data Engineering, Score: 3.9 *Sep 2015–Jun 2017*

- Focused on distributed data processing, storage systems, and cloud-native architectures
- Completed a capstone on real-time streaming pipelines for urban mobility data
- Built hands-on experience with Spark, Kafka, and columnar data warehouses

Courses: Distributed Systems, Database Systems, Data Warehousing, Cloud Computing, Machine Learning for Data, Data Modeling, Stream Processing, Information Retrieval

Cascade State University **https://www.cascadestate.edu**
Bachelor, Computer Science, Score: 3.7 *Sep 2011–Jun 2015*

- Developed a strong foundation in algorithms, databases, and software engineering
- Completed coursework in statistics and data analysis
- Served as a teaching assistant for an introductory database course

Courses: Data Structures, Algorithms, Operating Systems, Computer Networks, Software Engineering, Statistics

Work

Aurora Data Labs **https://www.auroradatalabs.com**
Senior Data Engineer *Jan 2021–Present*

- Designed and maintained batch and streaming pipelines using Apache Spark, Kafka, Airflow, and dbt
- Led migration of legacy ETL workflows to a cloud-native lakehouse on Snowflake and Databricks
- Built data quality checks and lineage metadata to improve incident detection and root-cause analysis
- Partnered with analytics and product teams to define data models, SLAs, and delivery timelines
- Mentored junior engineers on pipeline design, testing, and operational best practices

Keywords: Apache Spark, Apache Kafka, Apache Airflow, dbt, Snowflake, Data Modeling, Data Quality, ETL/ELT

Pulse Retail Group **https://www.pulseretail.com**
Data Engineer *Jul 2017–Dec 2020*

- Built and operated ELT pipelines for retail sales, inventory, and customer analytics
- Modeled warehouse layers in Amazon Redshift and automated transformations with Airflow
- Integrated third-party marketing and POS data into a unified reporting environment
- Improved pipeline reliability through monitoring, alerting, and automated data validation
- Supported Tableau dashboards used by merchandising and finance teams

Keywords: Python, SQL, AWS Glue, Amazon Redshift, Apache Airflow, Tableau, Data Warehousing

Cascade Health Systems **https://www.cascadehealth.org**
Data Engineering Intern *Jun 2016–Aug 2016*

- Assisted with building SQL-based ETL jobs for clinical and operational reporting
- Profiled source data and documented data quality issues for remediation
- Automated recurring data extracts and validation checks with Python

Keywords: SQL, Python, ETL, Data Profiling

Languages

English: Native or Bilingual Proficiency

Tamil: Native or Bilingual Proficiency

Spanish: Limited Working Proficiency **Keywords**: DELE B2

Skills

Data Engineering: Expert **Keywords**: Apache Spark, Apache Kafka, Apache Airflow, dbt, Snowflake, Databricks, Data Modeling

Cloud Platforms: Advanced **Keywords**: AWS, Azure, Google Cloud, AWS Glue, Azure Data Factory, Google BigQuery

Programming: Advanced **Keywords**: Python, SQL, Scala, Java, Bash

Databases: Advanced **Keywords**: PostgreSQL, Amazon Redshift, Snowflake, MongoDB, Apache Cassandra

Streaming & Messaging: Intermediate **Keywords**: Apache Kafka, Spark Structured Streaming, RabbitMQ, Avro, Schema Registry

Analytics & BI: Intermediate **Keywords**: Tableau, Power BI, Looker, dbt, Data Quality

Awards

Data Engineering Excellence Award

Aurora Data Labs *Mar 2023*

Recognized for leading the migration of legacy ETL workflows to a cloud-native lakehouse, reducing batch processing time by 40%.

Outstanding Graduate Capstone

Seattle Polytechnic Institute *Jun 2017*

Awarded for a capstone project on real-time streaming pipelines for urban mobility data.

Certificates

Google Cloud Professional Data Engineer **https://cloud.google.com/certification/data-engineer**

Google Cloud *Apr 2022*

AWS Certified Data Analytics - Specialty **https://aws.amazon.com/certification/certified-data-analytics-specialty/**

Amazon Web Services *Nov 2021*

Databricks Certified Data Engineer Professionalhttps://www.databricks.com/learn/certification/data-engineer-professional

Databricks *Mar 2023*

Publications

International Journal of Data Engineering and Analyticshttps://www.ijdea.org/articles/scalable-real-time-ingestion-pipelines

Scalable Real-Time Ingestion Pipelines for Retail Analytics *Oct 2020*

- Presents a reference architecture for low-latency ingestion using Kafka and Spark Structured Streaming
- Discusses schema evolution, exactly-once semantics, and operational monitoring
- Includes benchmarks from a production retail deployment

Projects

OpenLineage Data Quality Toolkit **https://github.com/priyaraman/openli**

An open-source toolkit for automated data lineage and quality checks across cloud data warehouses. *Jan 202*

- Built a Python library that integrates with Airflow and dbt to capture lineage metadata
- Added configurable data quality rules for freshness, volume, and schema drift
- Used by multiple internal teams to reduce incident detection time

Keywords: Data Quality, Data Lineage, Python, Airflow

Retail Sales Lakehouse **https://github.com/priyaraman/retail-lakehou**

A reference lakehouse implementation for retail sales analytics using Delta Lake and Spark. *May 2022–Sep 20*

- Designed bronze, silver, and gold layers for ingestion, cleansing, and business aggregates
- Implemented incremental processing and partition pruning for large fact tables
- Published dashboards in Power BI for sales and inventory KPIs

Keywords: Delta Lake, Spark, Lakehouse, Analytics

Interests

Open Source: Apache Airflow, dbt, Data Engineering

Outdoor Activities: Hiking, Cycling, Photography

Music: Jazz, Guitar

Volunteer

Civic Data Alliance **https://www.civicdataalliance.org**

Volunteer Data Engineer *Jan 2022–Present*

- Maintain data pipelines for a public transit open-data portal
- Collaborate with analysts to publish monthly ridership and accessibility datasets
- Mentor volunteers on SQL, Python, and data modeling best practices