

Anshita Srivastava

Software Engineer III — Data Engineering — Walmart Global Tech India
+91-6306908563 — anshita22srivastava@gmail.com — linkedin.com/in/anshita-srivastava

Professional Summary

Data engineer with 4 years of experience designing data pipelines and dimensional data models, and building ETL systems that turn raw data into trustworthy, business-ready datasets. Partners closely with business stakeholders and cross-functional teams to gather requirements and translate them into scalable data architecture, with a strong focus on data reliability, integrity, and governance — built automated validation frameworks and resolved metric and aggregation inconsistencies across large-scale analytics pipelines. Hands-on with Apache Spark, Airflow, and BigQuery for large-scale distributed data processing, and writes technical design and requirements documentation for existing and future data systems.

Work Experience

Software Engineer III — Walmart Global Tech

Sep 2024 – Present

- Designed, built, and optimized data architecture and ETL pipelines on Spark, supporting customer analytics across multiple international regions and enabling production rollout for 24 large-scale datasets.
- Built and maintained data platforms enabling data reliability, integrity, and governance by identifying and resolving duplicate mappings, aggregation inconsistencies, and metric inaccuracies across large-scale analytics pipelines.
- Designed dimensional and hierarchical data models that expanded analytics coverage and simplified downstream data consumption for business stakeholders.
- Led migration of Spark workloads to serverless infrastructure across multiple production environments, reducing infrastructure costs by 53% while improving scalability and performance.
- Improved Spark application performance by 56% through executor tuning, adaptive resource allocation, partition optimization, and caching strategies.
- Consolidated 8 fragmented pipeline modules into a unified, reusable framework, reducing maintenance overhead by ~40% and establishing data standards for independent, modular deployments.
- Partnered with product managers, business stakeholders, and cross-functional engineering teams to gather requirements and translate business needs into scalable, well-architected data solutions.
- Investigated and resolved 15+ high-priority production incidents annually, ensuring platform reliability across multiple international deployments.

Software Engineer II — Walmart Global Tech

Jun 2022 – Aug 2024

- Built distributed ETL systems processing over 1.5 TB of clickstream data daily using Scala, Apache Spark, Airflow, Hive, BigQuery, Dataproc, and Presto to produce useful, business-ready datasets.
- Increased data reliability through automated validation frameworks, monitoring, and quality enforcement across production analytics pipelines.
- Designed scalable batch data pipelines and established best practices for datasets and reporting powering customer analytics across multiple business domains.
- Improved engineering productivity by automating deployment pipelines, production validation, and end-to-end testing using CI/CD; maintained SonarQube coverage above 80%.
- Migrated data pipelines and Airflow DAGs between GCP projects with zero data loss and no business disruption.
- Led platform upgrades including Spark, Dataproc, Airflow, and BigQuery connector migrations with zero production downtime.

Technical Skills

- Languages: SQL, Python, Scala, Java, C++
- Data Pipelines & Modeling: Apache Spark, Apache Airflow, Hadoop, Hive, BigQuery, Dataproc, Dataproc Serverless, Presto — dimensional & hierarchical data modeling, ETL/ELT architecture
- Data Reliability & Governance: automated validation frameworks, data quality enforcement, metric standardization, monitoring & incident response
- Cloud: Google Cloud Platform (BigQuery, Dataproc, GCS)
- Engineering Practices: CI/CD, SonarQube, Unit Testing, Git, GitHub, Looker

Academic Qualifications

- M.Tech, Geoinformatics — IITK (Indian Institute of Technology, Kanpur) — CPI: 9.41/10.0 — 2020–2022
- B.Tech, Computer Science and Engineering — Institute of Engineering and Rural Technology, Prayagraj — 82.61% — 2016–2020

Certifications and Achievements

- Bravo Award – Walmart Global Tech, for outstanding technical contributions and high-impact business outcomes
- Google Cloud Certified – Generative AI Leader
- HackerRank SQL (Advanced) — GCP Essential Training for Developers — Data Engineering Foundations
- Academic Excellence Award – IIT Kanpur, Top 10% of department (AIR 1159/97,481 – GATE 2020, CS & IT)