

Vibhor Bansal

Bangalore, India — +91-8791774416 — vibhorbansal1312@gmail.com — [LinkedIn](#) — [GitHub](#)

SUMMARY

Results-driven **Data Engineer** with **5+** years of experience designing and optimizing end-to-end data pipelines in **Azure and AWS**. Skilled in ETL/ELT, data modeling, CDC/SCD, and orchestration using Airflow. Proven track record in building scalable, cost-efficient and secure data workflows, integrating batch and streaming sources like Kafka and relational databases into **Snowflake** and Delta Lake for analytics.

CORE SKILLS

- Cloud Platforms: Azure (ADLS Gen2, ADF, Databricks, Synapse), AWS (S3, Lambda, Glue, CloudWatch)
- Data Technologies: Snowflake, Delta Lake, SparkSQL, Medallion Architecture, Kafka, Change Data Capture
- Processing Frameworks: Apache **Spark**, PySpark, Airbyte
- Orchestration & Automation: Airflow, Azure Data Factory
- Programming & Scripting: **Python**, Java, Shell
- Data Operations: Data Modeling, Data Warehousing, SQL Optimization, Data Quality & Governance
- Monitoring Tools: Grafana, Prometheus, Splunk, Dynatrace, SonarQube
- CI/CD: Jenkins, **Argo CD**, Docker, Helm, Kubernetes

PROFESSIONAL EXPERIENCE

WPP Media – Data Engineer Bangalore Dec 2025 – Present

- Working on a large-scale **Media Data Lake** project, ingesting and processing media data using **PySpark** into a Medallion Architecture (Bronze → Silver → Gold).
- Building and optimizing PySpark-based ingestion pipelines to handle high-volume media datasets across multiple domains with reliability and performance.
- Applying data quality checks and transformations at each layer of the Medallion architecture to ensure clean, analytics-ready data in the Gold layer.

Optum – Senior Software Engineer Noida Mar 2022 – Dec 2025

- Architected an end-to-end CDC pipeline processing **200GB data daily** from PostgreSQL transactional DB to Snowflake Data Warehouse using Airbyte for ingestion and Airflow for orchestration. Enabling real-time analytics for 50+ business users.
- Implemented Postgres triggers to capture DDL and DML changes automatically in the audit table.
- Built Snowflake **procedures** to process data from staging layer into raw and foundation layers (latest unique records), applying SCD logic using datetime column and record status flag.
- Designed a meta-driven transformation framework, enabling dynamic mapping from audit to target tables.
- Integrated **Airflow DAGs** to trigger Airbyte sync jobs and Snowflake procedures in the downstream for seamless orchestration. Developed a reconciliation process to validate daily loads, truncating previous day's records from audit table post successful dag run.
- Developed and maintained a Java Spring Boot microservice handling **100K+ Kafka** messages/hour and process high-volume data from Kafka topics into Snowflake, achieving 99.9% uptime and reducing data processing costs by **30%** through optimized batch sizing and utilizing Snowflake Stream to capture and push the latest data into the foundation layer.
- Managed **10+ domains**, achieving 90%+ code coverage, SonarQube gate 9 and zero critical vulnerabilities.
- Led migration from Jenkins to Argo CD, **saving 15 hours** per month in production deployment.
- Built ADF pipelines to ingest CSV files into Landing Zones, convert to Parquet, and transform in **Databricks** (Bronze → Silver → Gold), creating Delta tables and implementing **event-based triggers** for healthcare data, applying transformations and aggregations; stored final datasets as Delta tables for efficient querying.
- Implemented SCD logic for low-change datasets and event-based triggers for high-frequency domains, ensuring near **real-time** Gold layer updates.
- Automated pipelines using **Azure Data Factory** for data ingestion and orchestration in Azure environments.

Bank of America – Apprentice Trainee Chennai Jul 2021 – Mar 2022

- Developed an AWS ETL pipeline for trading data using **Lambda, S3, Glue, and Athena**.
- Automated daily extraction with CloudWatch triggers, reducing manual work by 90%.
- Data cleansing and transformation are performed before loading into analytics layers, enabling near-real-time insights for business analysts.

EDUCATION

B.Tech in Software Engineering – SRM Institute of Science & Technology (84.5%)

2021