

Anchit Gupta

Senior Data Engineer — GenAI Cloud-Native Analytics
India (Open to Remote)
anchitgupt2012@gmail.com
linkedin.com/in/anchitgupt

Summary

Senior Data Engineer with 4+ years of experience building scalable, cloud-native data solutions and GenAI-powered platforms across gaming and enterprise domains. Proven track record in leading data platform modernization, real-time analytics pipelines, and observability frameworks that reduce operational overhead and accelerate business decisions. Adept in Python, PySpark, Databricks, AWS, and LangChain with strong cross-functional leadership across data, ML, and product teams.

Skills

- **Programming & Scripting:** Python (advanced), SQL, Bash, C/C++, JavaScript
- **Data Platforms:** Apache Spark (PySpark), Delta Lake, Apache Airflow, Databricks, dbt
- **Cloud & Infra:** AWS (S3, Redshift, Lambda), Snowflake, MongoDB, Docker, GitHub Actions
- **GenAI & ML:** LangChain, HuggingFace, MLflow, LLaMA, NLP (spaCy, fuzzy matching)
- **ETL & MDM:** Talend, Canonical Models, MDM Pipelines, Golden Record Logic
- **Observability:** Alerting Frameworks, Anomaly Detection, SLA Monitoring
- **BI & Reporting:** Tableau, Power BI, Business KPI Dashboards

Experience

Senior Data Engineer, Sevaro Health –

Oct 2025 – Present

- Architected a migration strategy to a Medallion architecture, projected to reduce technical debt by 20%–30% and optimize operational workflows.
- Engineered CI/CD pipelines using GitHub Actions to automate data pipeline deployments and ensure code quality.
- Implemented granular Role-Based Access Control (RBAC) policies in Databricks to strengthen data security and governance.
- Established a data observability framework and built monitoring dashboards to improve pipeline reliability and incident response.
- Spearheaded cloud cost management audits to track resource utilization and minimize infrastructure expenses.
- Accelerated the delivery of high-priority business requirements by proactively resolving cross-functional dependencies.

Data Engineer III, Jungle Games –

Jul 2024 – Oct 2025

- Built a LangChain-based GenAI summarization agent used daily across 5 games, saving 3.5 hours per analyst/day.
- Migrated Spark jobs from EMR to Databricks, improving ETL performance by 40% and reducing compute costs by 30%.
- Developed a LLaMA-based Rummy bot for the Learn/Practice section, enhancing user onboarding.
- Built MLOps pipelines for A/B testing and dynamic recommendation systems.

Data Engineer II, Jungle Games –

Jul 2023 – Jul 2024

- Led 3 engineers to develop a centralized alerting and observability framework.
- Designed a reusable alerting system with email support, reducing alert development effort by 70%.
- Created freshness and anomaly detection modules, reducing data quality incidents by 50%.

Data Engineer I, Jungle Games –

Oct 2022 – Jul 2023

- Built self-serve analytics layers for cross-functional teams.
- Enhanced GST/TDS pipelines, reducing turnaround time by 20%.
- Improved KPI dashboards, cutting SLA delays by 80%.

Associate, Axtria –

Apr 2022 – Sep 2022

- Developed MDM pipelines to unify 50M+ records using NLP, achieving 87% match accuracy.
- Delivered ETL tools and Golden Record logic for pharma client data.

Data Analyst, Axtria –

Jul 2021 – Apr 2022

- Generated HCP/HCO hierarchies and automated stakeholder notifications.
- Reduced matching algorithm time by 40% through memory optimization.

Education**IIIT Delhi** — M.Tech, Computer Science

Jul 2019 – Jul 2021

Moradabad Institute of Technology (AKTU) — B.Tech, Computer Science

2014 – 2018

Certifications

- Databricks Fundamentals Accreditation
- LangChain for LLM Application Development
- Astronomer: Apache Airflow Fundamentals
- Azure Databricks & Spark Core
- MTA Database Fundamentals, J2SE Java, Android App Development