

Avinash Gupta

Bihar, India — +91-9939854595 — avinashgupta94@hotmail.com
LinkedIn — GitHub

Professional Summary

Certified **Data Engineer Associate** and **AI Engineer for Developers (Associate)** with 3+ years of experience in analytics and data engineering. Skilled in building scalable data pipelines using **AWS (Glue, RDS, S3, Redshift, Athena, EMR)** and **SQL/Python** for ETL, modeling, and automation. Experienced in integrating **GenAI workflows (RAG, embeddings, LLM automation)** to drive data intelligence and efficiency.

Technical Skills

- **Cloud & Data Engineering:** AWS (Glue, DataBrew, Redshift, Athena, EMR, RDS, S3, IAM, CloudWatch), ETL design, data lakes, Parquet, orchestration, cost monitoring
- **Programming & Databases:** Python (Pandas, NumPy, Boto3), Advanced SQL (CTEs, joins, optimization), automation scripting, performance tuning
- **Data Modeling & BI:** Star/snowflake schema, warehousing, Redshift Spectrum, schema-on-read, Power BI, Tableau, QuickSight
- **AI & Analytics:** Prompt engineering, LLM/GenAI integration, embeddings, RAG pipelines, semantic search, EDA, feature engineering

Certifications

- **Data Engineer Associate** – DataCamp, Oct 2025
- **AI Engineer for Developers (Associate)** – DataCamp, Oct 2025
- **AWS Fundamentals of Analytics, Glue, Redshift** – AWS, 2025
- **AWS Educate Badges – Databases & Generative AI** – AWS, 2025
- **Amazon EMR Getting Started** – AWS Training & Certification, Jul 2025
- **AI-Powered Data Analysis & Business Intelligence** – Be10x, 2025
- **Foundations of AI & Everyday Productivity** – Be10x, 2025

Projects & Portfolio Highlights

- **SQL Server RBAC (ECommerceDB):** Designed multi-role SQL system with secure permissions; managed access for 10K+ users. GitHub
- **AWS RDS – Sales & Customer Insights:** Ingested S3 CSVs into RDS; optimized schema and queries; reduced query latency by 20%. GitHub
- **AWS Glue DataBrew ETL:** Automated data transformation and exports to S3/Redshift; boosted prep efficiency 30%. GitHub
- **Amazon EMR Serverless – PySpark ETL:** Built PySpark pipeline to process 5M+ records daily; reduced runtime by 40%. GitHub

- **Power BI Dashboards – Amazon & Flipkart:** Developed KPI dashboards with DAX, visualizing sales and customer insights. GitHub

Professional Experience

Career Sabbatical – Professional Development & Portfolio Projects *July 2023 – Present*

- Completed end-to-end AWS data engineering projects covering ingestion, ETL, orchestration, and monitoring.
- Upskilled through AI Mastery Series (GenAI, SQL/Python/Power BI with AI) and AWS Analytics programs.

Catalog Associate, Amazon Development Centre India *February 2021 – June 2023* Bangalore, India

- Analyzed product catalog data to identify listing errors, inconsistencies, and duplication patterns; implemented corrective actions improving data accuracy by 15%.
- Monitored daily operational KPIs using Excel and SQL-based reports to track defect trends, publishing delays, and SLA adherence.
- Partnered with cross-country teams to automate defect tracking and vendor performance dashboards, enabling faster root-cause analysis and decision-making.

Graduate Apprentice Trainee – Engineering Operations *Jun 2019 – Jul 2020* Bangalore, India

- Developed inspection plans and bills of materials; ensured quality compliance and timely delivery.

Education Bachelor of Technology, Mechanical Engineering *Jul 2014 – Jun 2018* Dr. A.P.J. Abdul Kalam Technical University, Greater Noida, India

12th (Intermediate), S.P.V. Inter College – BSEB *Apr 2012 – May 2014*

10th (Matriculation) – St. Xavier's Jr/Sr School (CBSE) *Passed June 2011*

Languages

Hindi (Native), English (Advanced C1)