Ankit Bhardwaj

Professional Summary

Data Engineer with expertise in big data processing, real-time pipelines, and cloud-based analytics. Completed 2-month Big Data training at Capgemini, delivering a scalable Netflix Data Analysis pipeline using PySpark, Scala, and Hive QL. Skilled in Python, SQL, Spark, Hive, Kafka, and AWS, with side projects in data visualization, streaming, and ETL automation.

Skills Profile

Technical

Operating System	Windows 10/11, Amazon Linux (EC2)
Environment	Client/Server, Cloud-based Data Processing (AWS), Distributed Computing (Apache Spark, Apache Kafka), Big Data Analytics, Business Intelligence (Amazon QuickSight)
Database	Hive, Amazon Athena, AWS Glue Data Catalog, MySQL , Amazon S3 (as data lake storage)
Cloud	AWS Management Console, Jupyter Notebook
Languages	Python, Scala, SQL (HiveQL, Athena SQL)
Other	Git/GitHub, Data Cleaning & Transformation (Pandas, PySpark), ETL Development, Data Visualization (QuickSight, Matplotlib, Seaborn), Data Warehousing (Partitioning, Bucketing), Real-time Data Streaming (Kafka), Cloud Services (AWS EC2, S3, Glue, Athena, IAM), Workflow Orchestration (Apache Airflow), Big Data Processing (Spark Core, Spark SQL), Performance Optimization, Agile/Scrum Methodologies

Functional

Management and Leadership Skills	• Experienced in **planning and task allocation** to maintain balanced workloads across teams. • Adept at **project management**, ensuring all deliverables meet time-specific goals and deadlines. • Strong focus on **clear communication, accountability, and timely progress updates** for effective coordination.
Training & Development	• Skilled at **knowledge sharing** to enhance team learning and technical competence. • Proficient in **explaining complex technical concepts** in a clear and accessible manner. • Provided **peer assistance** during hands-on training in **Big Data** and **Cloud technologies**. • Collaborative approach to **troubleshooting and resolving technical issues** within the team.

Professional Experience

Netflix Data Analysis with Spark and Hive **May 2025 – July 2025**

Role: Data Engineer

Built a scalable big data pipeline using **PySpark**, **Scala**, *and* **Hive QL** *for analyzing Netflix content.*

- Preprocessed data with **Pandas**, cleaned and transformed using Spark.
- Optimized queries via **partitioning** on type & release_year and **bucketing** on show_id.
- Derived insights on top genres, high-content countries, and category distribution.

Environment: PySpark, Scala, Hive QL, Apache Spark

Real-Time Stock Market Data Pipeline **May 2025 – July 2025**

Role: Data Engineer

Architected an end-to-end real-time data ingestion and analysis pipeline.

- *Simulated live stock data in* ***Python*** *and streamed via* ***Apache Kafka*** *on* ***AWS EC2****.*
- *Persisted JSON data to* ***Amazon S3****, automated schema discovery with* ***AWS Glue****, and enabled ad-hoc SQL queries with* ***Amazon Athena****.*
- *Delivered a scalable framework for real-time financial analytics.*

Environment: Python, Apache Kafka, Amazon Web Services (S3, Athena, Glue, EC2), SQL

Netflix Data Visualization – Amazon QuickSight **May 2025 – July 2025**

Role: Data Analyst

Engineered an interactive BI dashboard in **Amazon QuickSight** *analyzing Netflix's catalog (6K+ titles) with bar charts, donut charts, and pivot tables.*

- Connected **Amazon S3** dataset via manifest file for structured ingestion.
- Implemented advanced filters for deep dives into content by release year, genre, and format.
- Published a shareable PDF dashboard for stakeholder consumption.

Environment: Amazon QuickSight, Amazon S3

Twitter ETL Pipeline – Apache Airflow **May 2025 – July 2025**

Role: Data Engineer

Designed and deployed an automated ETL pipeline for Twitter data.

- Extracted tweets with **Tweepy**, transformed via **Pandas**, and stored structured data in **Amazon S3**.
- Orchestrated daily runs using **Apache Airflow** DAGs on **AWS EC2** with secure IAM roles.
- Created a reliable data lake for downstream analytics.

Environment: Python, Apache Airflow, AWS (EC2, S3, IAM)

Education

Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India, Bachelor of Technology in Computer Science, 2025