

Anand Raj - Data Engineer | 3+ Years | AWS Certified

[✉ geurecruitanandraj@gmail.com](mailto:geurecruitanandraj@gmail.com) | ☎ +91-8825378208 | 🗺 Gurugram, Haryana

[GitHub](#) | [\[LinkedIn\]](#) | **Notice: 1 Jan 2026**

PROFESSIONAL SUMMARY

Results-driven Data Engineer with 3+ years designing ETL pipelines and data solutions at scale. AWS Solutions Architect – Associate certified. Expert in Apache Airflow, PySpark, and hybrid cloud (AWS/Azure/GCP). Delivered 65% improvement in regulatory compliance through data validation frameworks. Skilled in Medallion Architecture, data quality, and pipeline orchestration processing 100M+ records daily.

TECHNICAL SKILLS

Languages: Python, PySpark, SQL, C++ | **ETL & Data Processing:** Apache Airflow, AWS Glue, Azure Data Factory, PySpark

Cloud Platforms: AWS (S3, Athena, Lambda, Step Functions, Glue, EMR, Data Catalog) | Azure (Data Factory, Synapse, Databricks, ADLS) | GCP (DataProc)

Databases: MySQL, PostgreSQL, Sybase | **Big Data:** Hadoop HDFS, Hive, Kafka

Data Engineering: ETL Pipelines, Data Pipeline Orchestration, Medallion Architecture, Data Quality, Schema Evolution, CDC, Incremental Loading, Data Governance

DevOps: Docker, Git, Jupyter | **Certifications:** AWS Solutions Architect – Associate

EXPERIENCE

Cognizant Technology Solutions | Associate Data Engineer, Moody's Ratings | Sept 2022 – Present

- **Data Quality & Validation:** Built Python validation tool (Pandas, NumPy) reducing regulatory submission errors by **65%**, ensuring 99.9% data accuracy and European compliance
- **AWS Cloud Migration:** Migrated legacy Sybase/Informatica infrastructure to serverless AWS (S3, Glue, Lambda, Step Functions), achieving **40% cost reduction** and **30% operational efficiency gain**
- **Error Handling:** Analyzed root causes of rating failures, implemented exponential backoff and dead-letter queues, reducing submission errors and downtime
- **Tech Stack:** PySpark, Python, AWS (S3, Athena, Lambda, Step Functions, Glue), PostgreSQL, XML

PROJECTS

1. Azure E-Commerce ETL Pipeline | Azure Data Factory, Databricks, Synapse, PySpark, MongoDB, GCP DataProc | [GitHub](#)

- Implemented Medallion Architecture (Bronze → Silver → Gold) processing **7 distinct tables** ingested from HTTP sources to ADLS Gen2
- Developed parameterized Azure Data Factory pipelines with data quality checks, incremental refresh using watermarking, and MongoDB enrichment
- Replicated architecture on GCP DataProc demonstrating multi-cloud proficiency; achieved **sub-minute query latency** optimizing Synapse views and CETAS materialization
- Scalable design supports **10x data volume** increase with minimal architectural changes

2. Apache Airflow NASA APOD Pipeline | Airflow (TaskFlow API), Docker, PostgreSQL, Python, REST APIs | [GitHub](#)

- Containerized Airflow environment using Docker Compose with isolated PostgreSQL metadata store for automated daily APOD ingestion
- Refactored legacy XCom-based DAGs to modern **TaskFlow API** improving maintainability, type hints, and scalability; implemented robust error handling with retries and alerting
- Built multi-stage data validation (schema, nulls, types) and idempotent ETL logic with insert-or-update database operations

3. Data Quality & Regulatory Compliance Framework | Python, Pandas, NumPy, SQL, XML

- Implemented 50+ validation rules: data completeness, accuracy, consistency, regulatory format compliance, cross-table integrity
- Result: **65% error reduction, 99.9% submission accuracy, 20 hours/week QA time savings**, zero compliance violations in audits

EDUCATION

B.Tech. Computer Science & Engineering | Graphic Era Deemed University, Dehradun | May 2022 | CGPA: 9.36/10

KEY ACHIEVEMENTS

- ✓ 65% reduction in regulatory submission errors | ✓ 40% cost savings through cloud modernization |
- ✓ 99.9% data accuracy compliance | ✓ Multi-cloud architecture (AWS/Azure/GCP) | ✓ Modern Airflow practices (TaskFlow API) | ✓ Medallion Architecture implementation | ✓ Scalable pipelines (100M+ records daily)