

RITIK DAGAR

Senior Data Engineer | AWS | Python | SQL | Spark | Databricks | Cybersecurity Analytics | Team Lead
ritikdagar0203@gmail.com | +91-9958152151 | linkedin.com/in/ritikdagar | github.com/ritikdagar

PROFESSIONAL SUMMARY

Senior Data Engineer at Barclays with 5+ years delivering production data platforms on-premise and on AWS. Promoted twice in 4 years. Lead a 12-engineer team automating 500+ KRI/KCI metrics and building cybersecurity analytics for CISO, Intel, Hunt, and Detection teams. Experienced in designing cloud-native architectures on AWS using Glue, S3, Redshift, and Databricks — including medallion lakehouse patterns, partition strategy, incremental ingestion, and cost optimisation. Recognised 5 times by the Barclays CISO organisation for automation excellence.

CORE SKILLS

Cloud and Big Data: AWS S3, AWS Glue (PySpark, Visual ETL, Bookmarks), AWS Lambda, AWS Step Functions, Amazon Redshift (Distribution Keys, Sort Keys, Workload Management), Amazon CloudWatch, AWS IAM, AWS EventBridge, Databricks (Medallion Architecture, Unity Catalog, Delta Lake, Auto Loader)

Data Engineering: Python (Pandas, Boto3, PySpark — partitioning, caching, broadcast joins, shuffle optimisation), Advanced SQL (Stored Procs, Window Functions, Query Optimisation, Execution Plans), SSIS, dbt (incremental models, snapshots)

Pipeline Design Patterns: Medallion Architecture (Bronze/Silver/Gold), Incremental Load with Watermarks, Partition Pruning, Batch vs Streaming Decision Framework, Deduplication via Hash Keys, Schema Evolution Handling, SCD Type 1 and 2

Data Quality and Observability: Source-to-Target Reconciliation, Row Count and Null Checks, Duplicate Detection, Business Rule Validation, CloudWatch Alerting, Great Expectations

Scheduling and DevOps: AWS Step Functions, AWS EventBridge, Autosys, GitLab CI/CD, GitHub Actions, AWS CodePipeline

Reporting and Visualisation: Tableau, Amazon QuickSight, SQL Server, REST APIs, SharePoint, Splunk

Security Domain: KRI/KCI Metrics, MITRE ATT&CK, Cyber Ops Reporting, Incident Ageing, Proactive/Reactive Detection, Audit Reporting

Leadership: Team Lead (12 Engineers), Graduate Training Programme, Senior Stakeholder Management, Cloud Solution Design, Cost Optimisation, Agile/Scrum

PROFESSIONAL EXPERIENCE

Barclays	2021 - Present
Senior Data Engineer Noida, India	2024 - Present
- Lead 12 engineers delivering cloud-native data pipelines and analytics across risk, compliance, and cybersecurity domains on AWS — owning architecture, cost governance, and team development. - Architected AWS medallion lakehouse for KRI/KCI metrics: Auto Loader ingests raw files from S3 Bronze layer, Glue PySpark jobs apply deduplication, narrow-to-wide transformations, and SCD Type 2 logic into Silver Delta tables, Gold layer aggregates into Redshift for Tableau — handles mixed-volume daily loads ranging from thousands to millions of records per metric. - Resolved Glue job performance degradation on high-volume pipeline: profiled Spark execution DAG, identified data skew caused by uneven metric category distribution, applied salting on partition key, switched from sort-merge join to broadcast join for small dimension tables, and enabled Glue bookmark for incremental processing — job runtime dropped from 3.5 hours to 28 minutes. - Eliminated duplicate records across 20+ source systems: implemented SHA-256 hash deduplication on business keys at Bronze ingestion, added MERGE INTO Delta Lake statements at Silver layer, and enforced source-to-target row count reconciliation via Great Expectations — duplicate incidents reduced to zero. - Standardised batch vs streaming processing decisions: batch for regulatory KRI/KCI metrics (low frequency, high accuracy), structured streaming via Kinesis for cyber operations dashboards where 15-minute staleness triggers operational risk — decision framework documented and adopted by all 12 engineers. - Cost optimised AWS data platform by 35%: right-sized Glue DPUs using job profiling, enabled S3 Intelligent-Tiering for cold metric history (saving 60% on storage), implemented Redshift WLM queues to prevent heavy analytical queries from blocking dashboard loads, and scheduled non-urgent Glue jobs during off-peak hours. - Recognised 5 times by Barclays CISO organisation for automation delivery — only data engineering team to receive this distinction repeatedly. - Introduced AI-assisted development (Copilot, Claude) for Glue job generation, PySpark optimisation suggestions, and CloudFormation templates — cut team delivery effort by 40%. - Built graduate onboarding programme covering Python, PySpark fundamentals, AWS architecture patterns, and Barclays data standards; 4 mentees promoted within 18 months.	
Data Engineer Pune, India	2022 - 2024
- Led initial AWS migration: moved SSIS batch pipelines to AWS Glue, designed S3 Bronze/Silver partition structure with date and source prefixes, established Glue bookmark-based incremental load pattern — became the standard adopted by the wider team. - Engineered KRI/KCI pipeline framework — SSIS/Glue hybrid processing 20+ sources, SQL Server and Redshift transformation layers, Tableau dashboards replacing manual Excel reporting.	

- Cut pipeline execution time by 35% through Spark partition tuning, eliminating wide shuffles by restructuring joins, and pushing filter predicates upstream before transformations.

Analyst | Pune, India

2021 - 2022

- Delivered SQL reporting pipelines and Tableau dashboards eliminating 10+ hours/week of manual stakeholder reporting.
- Independently caught and resolved a data duplication bug silently inflating live financial metrics — flagged via custom reconciliation check.

KEY PROJECTS

KRI/KCI Metrics — AWS Medallion Lakehouse | AWS Glue | S3 | Delta Lake | Redshift | Step Functions | Databricks | Tableau

- Designed Bronze/Silver/Gold medallion architecture on AWS: S3 as raw landing zone with Auto Loader ingestion, Glue PySpark Silver jobs handling deduplication, narrow-to-wide transformations, and SCD Type 2 history, Gold layer in Redshift with columnstore-style sort keys for fast Tableau aggregations across 500+ metrics.
- Single parameterised Glue job template handles all source types dynamically — partitioned by source and load_date, Glue bookmarks ensure only new data is processed, Unity Catalog on Databricks governs table access across teams.

Cyber Threat Visibility Platform — AWS | Python | Splunk | Lambda | S3 | Glue | Redshift | Kinesis | Tableau

- Unified CISO, Intel, Hunt, and Detection reporting on AWS — Lambda-triggered Splunk extraction lands in S3 Bronze, Glue transforms to Redshift Silver, Kinesis Data Streams feeds near-real-time ops dashboards with 15-minute latency SLA. Batch and streaming layers coexist cleanly.
- Redshift WLM configuration separates interactive dashboard queries from batch ETL loads — no query queue contention, dashboard P95 latency under 2 seconds for 500+ concurrent users.

Security Compliance Tracker — AWS Serverless | Lambda | S3 | Glue | Redshift | EventBridge | Great Expectations | Tableau

- Serverless compliance pipeline tracking 50,000+ employees: EventBridge-scheduled Lambda ingests SharePoint API, lands in S3 with date partitioning, Glue job deduplicates using SHA-256 hash on employee ID and training code, Great Expectations validates completeness before Redshift load, Tableau dashboard used as regulatory audit evidence.
- Zero-touch delivery with full observability — CloudWatch alarms on Glue job failure, Lambda timeout, and Great Expectations validation breach; eliminated 4 hours/week of manual work.

EDUCATION

SRM Institute of Technology

2017 - 2021

B.Tech (Computer Science) | GPA: 89.25% — First Class with Distinction

Greenfield Public School

2017

Class XII — PCM | 85.4%