Job Description – Senior Data Engineer

Company: Acme Systems
Role: Senior Data Engineer (Streaming & Analytics)

About the team:
We build real-time data products on AWS using Kafka, Spark Structured Streaming, and Flink. We care about
low-latency ETL, schema governance, data quality SLAs, and cost optimization.

Responsibilities:
- Design, build, and operate streaming pipelines (Kafka/Flink) and batch workflows (Spark/Airflow).
- Model data for analytics in Parquet, Delta/Iceberg; partitioning, Z-ordering, compaction.
- Develop CI/CD for data (GitLab/GitHub), testing, and observability (Prometheus/Grafana/CloudWatch).
- Partner with product/ML teams to ship features and A/B tests; own data SLIs/SLOs.
- Mentor junior engineers; lead design reviews.

Must-have skills:
- 5+ years with one of: Java/Scala/Python. Strong SQL.
- Kafka (producers/consumers), schema registry/Avro, exactly-once semantics basics.
- Spark (core + SQL), or Flink (DataStream/SQL); checkpointing, state backends.
- Data modeling for columnar formats (Parquet), partitioning strategies; Bloom/minmax indexes.
- AWS: S3, Glue/Athena/Lambda; IAM fundamentals.
- Strong debugging and production ownership.

Nice-to-have:
- Airflow or Argo; Lakehouse (Delta/Iceberg/Hudi)
- Vector DB / embeddings basics for RAG
- Cost optimization ("FinOps") stories.
- BI exposure (Tableau/Power BI/Looker).

Screening focus areas (examples):
- Kafka partitioning and consumer group rebalancing trade-offs
- Spark vs Flink for exactly-once and low-latency
- Parquet layout (row group size, dictionary encoding)
- Backfills and late-arriving data handling