Resume – Alex Doe

Name: Alex Doe
Email: alex.doe@example.com | Location: Hyderabad, IN

Summary:
Senior Data Engineer (6+ yrs) specialized in real-time streaming on AWS. Built Kafka + Flink pipelines (p95 < 800ms),
Spark batch ETL (1.5 TB/day), and cost-optimized S3/Parquet storage (35% savings via file-size tuning + partition pruning).
Comfortable with SQL, Java, and Python; dashboards in Power BI/QuickSight.

Experience:
1) Data Engineer, StarStream Labs (2022–Present)
   - Designed Flink DataStream jobs for clickstream + CDC; exactly-once with Kafka + RocksDB state.
   - Implemented S3 Parquet compaction; reduced small-files by 92%, Athena costs by 48%.
   - Added schema evolution with Confluent SR (Avro) and compatibility tests in CI.
   - Observability: Prometheus + Grafana; SLOs on end-to-end latency and error budgets.
   Tech: Flink, Kafka, Java 17, Avro, Parquet, AWS (S3, Glue, Athena, Lambda), GitLab CI.

2) Data Engineer, BlueCloud (2019–2022)
   - Spark ETL on EMR + Airflow, 1.2TB/day; optimized skewed joins with salting & AQE.
   - Built A/B metrics model in Parquet; Power BI dashboards for PMs.
   Tech: Spark 3.x, Python, SQL, EMR, S3, Airflow, Power BI.

Projects:
- Real-time fraud features: Kafka→Flink→S3 Parquet; p95 700ms; checkpointing & backfill strategy.
- Cost guardrails: automatic compaction + retention rules → 35% S3 + 40% Athena savings.
- RAG prototype for internal KB: Chroma + MiniLM embeddings; ticket search + summarization.

Skills:
- Languages: Java, Python, SQL
- Streaming: Kafka, Flink; Batch: Spark, Airflow
- Storage: S3, Parquet, Delta/Iceberg (basics)
- Cloud/DevOps: AWS, Docker, GitLab CI, Terraform (basic)
- BI: Power BI, QuickSight

Education:
B.Tech CSE, 2018