Name: Sarah Johnson

Contact: 987-654-3210 | San Francisco, CA

Summary:

Skilled Data Engineer with 2+ years of experience designing, building, and maintaining scalable data pipelines and ETL processes. Proficient in cloud data platforms, big data technologies, and database management. Adept at collaborating with data scientists and analysts to enable data-driven decision-making.

Education:

- Master of Science, Data Science, University of California, Berkeley, 2023
- Bachelor of Science, Computer Engineering, San Jose State University, 2020

Skills:

- Programming: Python, SQL, Scala
- Data Platforms: Apache Spark, Hadoop, Kafka, Airflow
- Cloud: AWS (S3, Redshift, Glue), GCP (BigQuery, Dataflow)
- Tools: Apache NiFi, Docker, Git, Jenkins
- Databases: PostgreSQL, MySQL, MongoDB

Experience:

1. **Data Engineer, Innovate Analytics, San Francisco, CA** (Mar 2023 - Present)

- Developed and maintained scalable ETL pipelines ingesting data from multiple sources using Apache Spark and AWS Glue.
- Designed data models and optimized data warehouse queries, reducing report latency by 35%.
- Automated data quality checks and monitoring workflows using Apache Airflow, improving data reliability.
- Collaborated with data scientists to provision datasets for machine learning models and analytics.

Projects:

- **Real-Time User Analytics Pipeline:** Created an end-to-end streaming pipeline using Kafka, Spark Streaming, and AWS Redshift to provide real-time insights on user behavior.
- **Sales Data ETL Framework:** Developed reusable ETL components in Python and Airflow to standardize data ingestion across departments.

Certifications:

- AWS Certified Data Analytics – Specialty
- Google Cloud Professional Data Engineer