

# John Doe

johndoe@email.com | (123) 456-7890 | linkedin.com/in/johndoe | github.com/johndoe

Seattle, WA

## Professional Summary

Data Engineer with 5+ years of experience designing, building, and optimizing large-scale data pipelines and architectures. Proficient in Python, SQL, and cloud platforms (AWS, GCP). Adept at transforming raw data into actionable insights, improving data processing efficiency by up to 40%. Strong collaborator with a proven track record in delivering scalable solutions for business-critical applications.

## Technical Skills

- **Languages:** Python, SQL, Scala, Java
- **Databases:** PostgreSQL, MySQL, Snowflake, Redshift
- **Big Data:** Apache Spark, Hadoop, Kafka, Airflow
- **Cloud:** AWS (S3, Lambda, Glue, RDS), GCP (BigQuery, Dataflow), Azure
- **Tools:** Docker, Kubernetes, Terraform, Git, Jenkins
- **Data Modeling:** ETL/ELT, Dimensional Modeling, Data Warehousing

## Professional Experience

**Senior Data Engineer** *TechCorp, Seattle, WA* *Jan 2022 – Present*

- Designed and implemented scalable ETL pipelines using Apache Spark and AWS Glue, reducing data processing time by 35% for a customer analytics platform serving 10M+ users.
- Built real-time data streaming solutions with Kafka and AWS Kinesis, enabling near-real-time dashboards for business stakeholders.
- Optimized Snowflake data warehouse queries, cutting costs by 20% through efficient indexing and partitioning strategies.
- Automated data workflows using Apache Airflow, improving pipeline reliability and reducing manual intervention by 50%.
- Collaborated with data scientists to deploy machine learning models into production using Docker and Kubernetes.

**Data Engineer** *DataSolutions Inc., San Francisco, CA* *Jul 2019 – Dec 2021*

- Developed and maintained data pipelines in Python and SQL, processing 500TB+ of data monthly for financial reporting systems.
- Migrated on-premises Hadoop cluster to GCP BigQuery, reducing infrastructure costs by 30% and improving query performance.

- Implemented CI/CD pipelines with Jenkins and Terraform for infrastructure-as-code deployments, accelerating release cycles by 25%.
- Designed star-schema data models for a customer-facing analytics platform, enhancing query efficiency for ad-hoc reporting.

#### **Junior Data Engineer**

*StartupX, Austin, TX*

*May 2017 – Jun 2019*

- Built ETL processes using Python and PostgreSQL to support marketing analytics, processing 10M+ daily user events.
- Assisted in developing data quality checks, reducing data discrepancies by 15% across reporting systems.
- Supported data migration to AWS Redshift, enabling scalable storage for growing datasets.

## **Education**

**Bachelor of Science in Computer Science** *University of Texas, Austin, TX 2013 – 2017*

- Relevant Coursework: Database Systems, Distributed Systems, Machine Learning

## **Certifications**

- AWS Certified Solutions Architect – Associate (2023)
- Google Cloud Professional Data Engineer (2022)
- Databricks Certified Associate Developer for Apache Spark (2021)

## **Projects**

### **Real-Time Fraud Detection System**

*Personal Project, GitHub*

- Developed a real-time fraud detection pipeline using Kafka, Spark Streaming, and PostgreSQL, processing 1M+ transactions daily.
- Deployed on AWS using Docker containers, achieving 99.9% uptime.