Jane Doe

Data Engineer • AWS Certified • Cloud Data Solutions

jane.doe.email@email.com | (555) 123-4567 | Anytown, USA

linkedin.com/in/janedoe-example | github.com/janedoe-example

Profile Summary

Results-oriented Data Engineer with 5+ years of experience designing, building, and maintaining scalable data pipelines and infrastructure, primarily within the AWS ecosystem. Proven ability to leverage services like S3, Redshift, Glue, EMR, and Lambda to process large datasets, optimize ETL workflows, and deliver actionable insights. Seeking to apply expertise in cloud data solutions and big data technologies to drive data-driven decision-making.

Technical Skills

Cloud Platforms: AWS (S3, Redshift, Glue, EMR, Lambda, Kinesis, EC2, RDS, IAM, CloudWatch)
Data Warehousing: Amazon Redshift, Snowflake (Basic)
ETL/Data Integration: AWS Glue, Apache Airflow, Python (Boto3, Pandas), SQL
Big Data Technologies: Apache Spark (PySpark), Hadoop Ecosystem (HDFS, Hive)
Databases: PostgreSQL, MySQL, NoSQL (DynamoDB - Basic)
Programming Languages: Python, SQL, Scala (Basic), Bash Scripting
BI & Visualization: Tableau (Basic), AWS QuickSight (Basic)
DevOps & CI/CD: Docker, Git, GitHub Actions (Basic), Terraform (Basic)
Operating Systems: Linux, macOS, Windows

Professional Experience

Senior Data Engineer | **CloudData Corp**

Anytown, USA 2020 – Present

- Designed and implemented scalable ETL pipelines using AWS Glue and Python (PySpark) to process terabytes of clickstream data daily, reducing processing time by 30% and enabling faster reporting.
- Managed and optimized a large-scale Amazon Redshift data warehouse, implementing workload management (WLM) and performance tuning techniques that improved query performance by 25%.
- Developed and deployed serverless data processing workflows using AWS Lambda and Kinesis for real-time data ingestion and analysis.
- Built and maintained data orchestration jobs using Apache Airflow, ensuring reliable execution of complex dependencies across multiple AWS services.
- Collaborated with data scientists and analysts to understand data requirements and build data models optimized for analytical querying in Redshift.
- Utilized Terraform for provisioning and managing AWS data infrastructure, promoting Infrastructure as Code (IaC) best practices.
- Mentored junior data engineers on AWS best practices, Python scripting, and SQL optimization.

Data Engineer | **Tech Solutions Inc.**

Somewhere, USA 2018 – 2020

- Developed Python scripts and SQL queries to extract, transform, and load data from various relational databases (PostgreSQL, MySQL) into a central data repository.
- Assisted in the migration of on-premises data pipelines to AWS, primarily using S3 and EC2-based processing.
- Created automated data quality checks using Python and SQL to ensure data integrity.
- Supported ad-hoc data requests from business stakeholders using SQL and basic data visualization tools.

Education

M.S. in Computer Science | State University Anytown, USA | 2018 - Focus: Data Science & Cloud Computing - Thesis: Performance Analysis of Distributed Data Processing Frameworks

B.S. in Information Systems | Tech Institute Somewhere, USA | 2016 - Minor: Mathematics

Certifications

- **AWS Certified Data Analytics – Specialty** (2021)
- **AWS Certified Solutions Architect – Associate** (2019)

Projects

Real-time Analytics Dashboard **2019**

- Developed a proof-of-concept dashboard using AWS Kinesis, Lambda, and QuickSight to visualize streaming sensor data.

Data Lake Implementation **2018**

- Contributed to building a data lake on AWS S3, establishing partitioning strategies and metadata management using AWS Glue Data Catalog.

Keywords

(This section will be auto-populated with unused keywords)