Emily Johnson

206-555-0198 | emily.johnson@email.com | Seattle, WA
LinkedIn | GitHub | Portfolio

Professional Summary

Senior Data Engineer with 7+ years of experience designing and optimizing scalable data pipelines. Specialized in distributed data processing using Python and Spark, with extensive cloud expertise in AWS and Azure. Proven track record of leading teams to implement enterprise-grade data solutions handling 10TB+ daily workloads.

Professional Experience

Senior Data Engineer at TechData Corp

Seattle, WA | 03/2021 - Present
- Designed real-time analytics pipeline processing 15M events/day using Spark Structured Streaming
- Reduced ETL latency by 40% through PySpark optimizations and dynamic resource allocation
- Led migration of legacy Hadoop system to cloud-native architecture on AWS (Glue, Redshift, S3)
- Implemented CI/CD pipelines for data workflows using Airflow and GitHub Actions
Technologies: Python, Spark, AWS, Airflow, Kafka, Parquet, Docker

Data Engineer at CloudAnalytics Solutions

San Francisco, CA | 06/2017 - 02/2021
- Built scalable data ingestion framework handling 10TB+ daily from 50+ data sources
- Developed automated data quality monitoring system reducing data errors by 65%
- Optimized Spark clusters achieving 30% cost reduction through performance tuning
- Integrated Azure Data Factory pipelines with Databricks for ML feature engineering
Technologies: Python, Spark, Azure, SQL, Hadoop, Hive, Power BI

Education

Bachelor of Science in Computer Science

University of Washington - Seattle, WA | 2016
GPA: 3.7
Relevant Coursework: Data Structures & Algorithms, Database Systems, Distributed Systems, Cloud Computing, Big Data Analytics

Skills

* Python (Advanced)
* Apache Spark

* AWS (Glue, Redshift, S3)

* Azure (Data Factory, Databricks)

* SQL & NoSQL Databases

* Data Modeling

* ETL/ELT Architecture

* Distributed Systems

* Stream Processing

* Terraform

* Kubernetes

* Data Governance

Certifications

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