

SUMMARY

Highly experienced Senior Software Engineer with 8 years of experience in Big Data, specializing in optimizing Spark jobs for large datasets, and delivering scalable data processing solutions.

EXPERIENCE

Senior Software Engineer

DataLab Inc, 2018–Present

- Led the development of a real-time data analytics platform using Apache Kafka, Spark, and HBase, resulting in a 30% increase in data processing efficiency
- Designed and implemented a data warehousing solution using Amazon Redshift, reducing query times by 50%
- Mentored junior engineers in optimizing Spark jobs for large datasets and improving code quality

Software Engineer

BigDataCorp, 2015–2018

- Built a scalable data processing pipeline using Apache Beam and Google Cloud Dataflow, handling 100K+ events per second
 - Collaborated with the data science team to develop predictive models using Python and scikit-learn, resulting in a 25% increase in sales forecasting accuracy
 - Created technical documentation for data processing workflows using Apache Airflow and Markdown
-

PROJECTS

Spark Optimization Toolkit

Personal Project, 2020

- Developed a set of tools and techniques for optimizing Spark jobs on large datasets, resulting in a 40% reduction in processing time
- Published the project on GitHub and presented it at a local meetup, receiving positive feedback from the community

Real-time Analytics Platform

DataLab Inc, 2019

- Contributed to the development of a real-time analytics platform using Apache Kafka, Spark, and Cassandra, handling 10K+ concurrent users
 - Implemented a data visualization dashboard using Tableau and Python, providing insights into user behavior and engagement
-

TECHNICAL SKILLS

TECHNICAL SKILLS

Languages: Java, Python, Scala

Frameworks: Apache Spark, Apache Beam, Apache Kafka

Cloud: AWS, GCP, Azure

Tools: Git, Docker, Jenkins, Apache Airflow

Databases: HBase, Cassandra, Amazon Redshift

OS: Linux, macOS, Windows

EDUCATION

B.S. in Computer Science, Stanford University, 2015

GPA: 3.5/4.0

Relevant Courses: Data Structures, Algorithms, Computer Systems, Big Data Processing

Thesis: "Optimizing Spark Jobs for Large Datasets"