John Willis
johnwillis@example.com | 001-680-912-1569x4525 | linkedin.com/in/johnwillis | github.com/john4008

SUMMARY

Highly motivated Mid-Level Software Engineer with 3 years of experience in Big Data, specializing in optimized Spark jobs for large datasets, and a proven track record of delivering high-performance data processing solutions.

EXPERIENCE

Senior Data Engineer
VantaLabs, 2022–Present
- Designed and implemented a scalable data warehouse using Apache Hive and HBase, resulting in a 30% reduction in data processing time
- Developed and optimized Spark jobs for large-scale data processing, achieving a 25% increase in throughput
- Collaborated with the data science team to integrate machine learning models into the data pipeline, using TensorFlow and Scikit-learn
- Maintained and improved the performance of existing data pipelines, resolving issues and implementing new features as needed

Data Engineer
HealthNet AI, 2020–2022
- Built and deployed a real-time data streaming platform using Apache Kafka and Apache Storm, handling 10,000+ events per second
- Worked with the data analytics team to design and implement data visualizations using Tableau and Power BI
- Implemented data quality checks and validation using Apache Beam and Python
- Participated in the development of a cloud-based data lake using AWS S3 and Apache Spark

---

PROJECTS

Optimized Spark Jobs for Large Datasets
Personal Project, 2022
- Researched and implemented optimization techniques for Spark jobs, resulting in a 40% reduction in execution time
- Developed a toolkit for automating Spark job optimization, using Python and Apache Spark
- Presented the project at a local meetups and received positive feedback from the community

Real-Time Data Streaming Platform
Hackathon Project, 2020
- Designed and implemented a real-time data streaming platform using Apache Kafka and Apache Storm, handling 1,000+ events per second
- Collaborated with a team of developers to build a working prototype within 24 hours
- Won the "Best Use of Data" award at the hackathon

TECHNICAL SKILLS

Languages: Java, Python, Scala
Frameworks: Apache Spark, Apache HBase, Apache Hive
Cloud: AWS, GCP
Tools: Git, Docker, Jenkins
Databases: MySQL, MongoDB, Apache Cassandra
OS: Linux, Windows, macOS

EDUCATION

Master of Science in Computer Science, Stanford University, 2020
GPA: 3.9/4.0
Relevant Courses: Distributed Systems, Data Mining, Machine Learning
Thesis: "Optimizing Spark Jobs for Large-Scale Data Processing"
Mentored undergraduate students in computer science and participated in research projects.