

Joseph Russell

Python Developer

Vancouver, British Columbia

email@email.com

Summary

Experienced PySpark developer with 4+ years of experience in building scalable and distributed data processing solutions. Proficient in developing complex ETL pipelines using PySpark, SQL, and NoSQL databases. Committed to delivering high-quality work and continuously improving skills to stay current with new technologies.

Work experience

Company Name , City, State

PySpark Developer, Employment Dates

- Designed and developed ETL pipelines using PySpark and Hive to process large datasets in a distributed environment.
- Collaborated with data scientists and business analysts to understand requirements and develop data models that support data analysis and business intelligence.
- Worked with NoSQL databases such as HBase, Cassandra, and MongoDB to store and manage large volumes of data.
- Developed and maintained data processing and cleansing scripts using Python and PySpark.
- Optimized Spark jobs by tuning performance parameters and implementing best practices for partitioning and caching data.
- Created and maintained data pipelines using Apache Airflow to schedule, monitor, and manage workflows.
- Developed and deployed machine learning models using PySpark and libraries such as MLLib and TensorFlow to perform predictive analysis on large datasets.
- Conducted code reviews and mentored junior developers on PySpark best practices, leading to improved code quality and faster delivery times.

Company Name , City, State

PySpark Developer, Employment Dates

- Developed and maintained PySpark-based data processing pipelines for batch and streaming data using Kafka, Spark Streaming, and Spark SQL.
- Designed and implemented a data lake architecture using AWS S3, Glue, and Athena, allowing for flexible and scalable data storage and querying.
- Built and deployed PySpark-based machine learning models for predictive maintenance and anomaly detection on industrial equipment.
- Worked with DevOps teams to deploy and maintain PySpark jobs on Apache Mesos and Kubernetes clusters.
- Conducted technical interviews for hiring candidates and provided feedback on PySpark skills and technical knowledge.

Skills

Programming: Apache Airflow



Programming: Caching



Programming: Git



Big Data: Apache Kafka



Big Data: Cassandra

