Lillie Stufflebean

Big Data Engineer

I am a Big Data Engineer with over three years of experience in configuring, deploying, and managing big data solutions. I have extensive knowledge of big data technologies such as Hadoop, Spark, and Kafka, and am highly skilled in writing and optimizing complex SQL queries. I am also familiar with a variety of data visualization tools and have experience in building data pipelines for ETL operations. I have a proven track record of successfully designing and deploying big data solutions and have been commended for my ability to quickly understand customer requirements and develop solutions that meet their needs. I am an excellent communicator, able to effectively collaborate with project teams and stakeholders.

lillie.stufflebean@gmail.com

(229) 621-7056

Fargo, ND

Education

Bachelor of Science in Big Data Engineering at North Dakota State University, ND

Aug 2016 - May 2020

Relevant Coursework: Advanced Data Structures and Algorithms, Database Systems and Applications, Big Data Storage and Analysis, Machine Learning, and Data Mining.

Links

linkedin.com/in/lilliestufflebean

Skills

- Hadoop
- Apache Spark
- Data Warehousing
- ETL/ELT Pipelines
- Machine Learning
- NoSQL Databases
- Cloud Computing

Languages

Employment History

Lead Big Data Engineer at Microsoft, ND

Oct 2022 - Present

- Developed a big data platform using Azure Data Lake that processed over 10 petabytes of customer information, resulting in an increase in accuracy and efficiency by 30%.
- Created automated ETL pipelines to facilitate the ingestion of unstructured datasets from various sources into Hadoop clusters with 95% success rate.
- Led a team of 5 engineers responsible for designing distributed computing applications on Apache Spark which improved performance metrics by 40%.
- Developed real-time streaming analytics solutions utilizing Microsoft Stream Analytics Services to process millions records per day, leading to significant cost savings for the company.

Senior Big Data Engineer at Amazon Web Services, ND

Jul 2020 - Sep 2022

- Developed an end-to-end big data pipeline for Amazon Web Services, ND that improved performance by 30% and decreased latency time from 6 hours to 3 minutes.
- Developed a Lambda architecture on AWS EMR cluster of over 100 nodes with HiveQL & SparkSQL as the query languages which enabled real-time analytics processing at scale.
- Implemented advanced machine learning algorithms such as Random Forest and Neural Network Models leveraging Apache Mahout libraries resulting in 20x increase in accuracy rate while reducing cost by 50%.
- Designed and developed automated ETL jobs using Sqoop, Kafka Streams API's running on EC2 instances enabling easy transfer of large datasets between different storage systems within 4 hrs compared to 8 hrs before implementation.

Certificates

Cloudera Certified Professional: Data Engineer

Apr 2021

Hortonworks HDP Certified Apache Hadoop Developer

Jan 2020

Memberships