Bollineni Chandra Sekhar

Machine Learning Data Engineer

chandrasekhar090591@gmail.com   3039156458   Denver, Colorado   www.linkedin.com/in/chandrasekharbollineni

Results-driven IT professional with nearly 10 years of experience in software development, particularly in the Big Data ecosystem and machine learning. I specialize in leveraging Hadoop components like HDFS, Spark, and Hive to create scalable, efficient computing solutions. My skills include building robust ETL pipelines using Python, MySQL, and Apache Spark, as well as developing machine learning models with Python and scikit-learn. I have hands-on experience with cloud platforms such as AWS and Azure, and I excel in SQL development, performance tuning in Hive, and creating data visualizations with Tableau. Additionally, I'm proficient in containerization and orchestration technologies like Docker and Kubernetes, which I leverage to enhance CI/CD processes.

I also bring a strong background in management and operations, focusing on optimizing team performance and driving projects to successful completion. My commitment to collaboration and adaptability ensures I meet dynamic business needs effectively.

WORK EXPERIENCE

Deloitte - Deputy Manager

Dec 2023 - Jul 2024   India   India

- Analyzed business requirements and created 15 detailed specifications, achieving full compliance with project development guidelines.
- Designed ETL pipelines with Spark and Hive for data extraction and ingestion into S3, resulting in a 30% increase in data availability.
- Enhanced PySpark data loading procedures, effectively managing over 1TB of daily data and improving pipeline reliability by 25%.
- Developed solutions and workarounds during issue analysis, reducing resolution times by 20% for end users and clients.
- Contributed to Agile practices through daily scrum meetings and sprint planning, leading to a 15% improvement in project timelines and team collaboration.

Tech Mahindra - Technical Project Manager

May 2021 - Sep 2023   India

- Managed the complete Big Data project lifecycle from scoping to support, achieving a 95% on-time delivery rate.
- Developed efficient data curation frameworks with Spark and Hive, reducing processing times by 50%, which significantly advanced project timelines.
- Facilitated client meetings for requirements gathering, resulting in better alignment with client expectations and a 30% decrease in rework.
- Successfully delivered key milestones on time by overseeing the entire project lifecycle, boosting team efficiency by 30%.

Netapp - Data Engineer

Apr 2020 - May 2021   India

- Optimized log data management using Kafka to process 5 million records daily, which improved data retrieval speeds by 25%.
- Engineered an ingestion framework with DynamoDB and Cassandra using Python and Big Data technologies, enhancing retrieval speeds by 40%.
- Leveraged Spark for creating RDDs and DataFrames, achieving a 30% improvement in data transformation performance.
- Enhanced data accessibility by 20% through the extraction and ingestion of data into Azure Data Lake Store via ETL pipelines with Spark and Hive.

Nokia - Data Engineer

Jun 2017 - Apr 2020   India

- Increased data accessibility by 20% by implementing ETL pipelines with Spark and Hive for Azure Data Lake Store.
- Utilized PySpark to apply data cleansing techniques, generating DataFrames and RDDs and reducing preprocessing time by 35%.
- Contributed to the development of 50 test scripts and cases, increasing testing coverage by 25% for priority user stories.
- Processed Big Data on HDInsight clusters within Microsoft Azure, achieving a 50% increase in processing capacity.
- Streamlined collaboration via Bitbucket, resulting in a 25% reduction in code integration errors.
- Enhanced system reliability by optimizing data transfer processes between cloud platforms, yielding a 40% improvement in performance.

Realtics - Python Engineer

Nov 2014 - May 2017   India

- Assessed business requirements and formulated detailed specifications, which enhanced program development timelines by 20%.
- Employed Sqoop for data extraction from multiple databases and sources.
- Enhanced efficiency in HDFS by 40% through strategic optimization of MapReduce processes using advanced compression methods.
- Spearheaded data extraction and ingestion initiatives into a Hadoop Data Lake, creating ETL pipelines with Spark and Hive that improved data flow efficiency by 40%.

EDUCATION

Regis University - Masters, Data Science

Aug, 2024 - current, 4.0/4.0   Denver, Colorado

SKILLS

Machine Learning: TensorFlow, Keras, Scikit-Learn, NumPy, Pandas, Seaborn, Matplotlib, ML Algorithms, Statistics;
Big Data Technologies: HDFS, MapReduce, Sqoop, Pig, Hive, Kafka, Spark, NiFi, ZooKeeper, Elasticsearch, AWS (EMR, S3, Glue, Athena, Redshift, Lambda, Kinesis, SageMaker), Azure (Data Factory, Databricks, Cosmos DB, Event Hubs, Functions, Kubernetes Service);
Databases: SQL (Oracle, MySQL, PostgreSQL), NoSQL (MongoDB, Cassandra).
Programming Languages: C, SQL, Python, C++, Shell Scripting, R;
Web Services: XML, SOAP, REST APIs;
Operating Systems: Windows, Unix, Mac;
Front End Technologies: HTML, CSS, JavaScript.