Sai Sampath Kumar Baisetti

Big Data Engineer

Singapore

- Email me on Indeed: http://www.indeed.com/r/Sai-Sampath-Kumar-Baisetti/22929e156fd21d10

Big Data Developer

Work Experience

Big Data Engineer

Tata Consultancy Services

August 2016 to Present

• 3+ years of Experience with Big Data Ecosystems and currently working on creatng Ingeston and Computaton frameworks by leveraging Spark with Job Server, Hive, Sqoop, Impala and Apache NIFI at Tata Consultancy Services (TCS) for client DBS.
• Good Understanding and working experience with HDFS Architecture and Yarn Scheduler. Expertse with Hadoop Ecosystem tools like Kafaa, HBasea, Impaalaa, Cassandraa, Hivea, Sqoopaa, Ooziea, Piga, Apaache Spaark.
• Good knowledge on SQL and can write SQL queries to retrieve required results. Experience in working with No SQL databases like Cassandra and MongoDB and Relatonal Databases like Oraclea, MariaDB.
• IBM certified Docker and Blockchain developaer.
• Have good experience with Apaache Spaark and can write efcient code that computes on huge amount of data with lesser processing tme and handle diferent fle formats like Avroa, paarquet and also developed custom hive serdes to facilitate custom writng style with diferent comparession algorithms.
• Expertse in Developing Ingeston frameworks to ingest diferent types of Files (Delimiteda, Positiona, CSVa, XMLa, JSON) and RDBMS tables into HDFS using Apaache NiFi and Apaache Spaark.
• Demonstrated experience to work and develop with Spaark job server, A rest API service that will allow usage of shared spark contexts to reduce job run tme.
• Experience in working with Devopas tools and have built continuous Integration and Continuous Depaloyment paipaelines using Jenkins and Bitbucket.
• Experience in usage of tools like Google Code CQ Engine and Drools to apply business rules during data processing stage.
• Have the ability to code in Java, Python and Scala and also have the ability to write Shell Scriptng.
• Developing on spark using Spaline, an open source tool for tracking lineage of data processed through Spark for business purpose.
• Experience in analysing huge datasets and fnding insights and paterns in both structured and unstructured data.
• Containerisation of Apache NiFi using docker to facilitate ingeston onto diferent HDFS servers to avoid load on gateway nodes and edge nodes when NiFi runs as a single Independent Instance using metadata stored on MariaDB.
• Have fundamentals in AWS and have pushed data into S3. Good understanding of

diferent AWS tools.
• Have good understanding of diferent machine learning models by plotng using pandas and Seaborn and using Keras for TensorFlow to develop diferent kinds of models.
• Can relate to diferent terms used in Finance and Risk domains of the bank.

Education

Bachelor's

Certifications/Licenses

IBM Certified Block Chain Developer
August 2019 to Present

IBM Certified Docker Essentials
September 2019 to Present