Job Description

We are looking for a Big Data Engineer that will work on the collecting, storing, processing, and analyzing of huge sets of data. The primary focus will be on choosing optimal solutions to use for these purposes, then maintaining, implementing, and monitoring them. You will also be responsible for integrating them with the architecture used across the company.

Responsibilities

- Selecting and integrating any Big Data tools and frameworks required to provide requested capabilities
- Implementing ETL process
- Monitoring performance and advising any necessary infrastructure changes
- Defining data retention policies

Skills and Qualifications

- Proficient understanding of distributed computing principles
- Ability to solve any ongoing issues with operating the cluster
- Proficiency with Hadoop v2, MapReduce, HDFS
- Experience with building stream-processing systems, using solutions such as Storm or Spark-Streaming
- Good knowledge of Big Data querying tools, such as Pig, Hive, and Impala
- Experience with Spark
- Experience with integration of data from multiple data sources
- Experience with NoSQL databases, such as HBase, Cassandra, MongoDB
- Knowledge of various ETL techniques and frameworks, such as Flume
- Experience with various messaging systems, such as Kafka or RabbitMQ
- Experience with Big Data ML toolkits, such as Mahout, SparkML, or H2O
- Good understanding of Lambda Architecture, along with its advantages and drawbacks
- Experience with Cloudera