cerberus

Data Engineer

Our Data Engineers are heart of CTS! We create cutting-edge solutions to manage the unyielding flow of information and support hundreds of data analysts, data scientists and product experts around the world in bringing transparency and meaning to the data that makes our businesses successful. We pride ourselves as technology practitioners strongly believing that change & innovation is the only constant within our industry. Our team relies on the careful curation of underlying data to answer questions that are often not simple or clear. You will be creating solutions that impacting consumers across geographies & industries (healthcare, retail/ consumer, telecom, industrial) driving technology transformation. At the core of our team, we believe data is best used to create transparency & generate communal value.

What You'll Do:

» Build & deploy large-scale ETL and stream processing pipelines in our serverless microservice infrastructure built on top of industry standard technology like Kubernetes & Kafka.
» Manage workflows in support of both product and our AI/Data Science pipeline, you'll be introduced to our unique ingest and processing pipelines turning proprietary data assets into groundbreaking solutions
» Build stream ingestion processes to efficiently send, process, analyze & publish data
» Perform analyses of large structured and unstructured data to solve multiple & complex business problems
» Investigate and prototype different task dependency frameworks to understand the most appropriate design for a given use case
» Work hand-in-hand with the data science team to understand various user or content trends that influence product changes and customer acquisition strategies

Who You Are:

» An Engineer interested in working in both streaming and batch processing environments [Spark, Kafka streaming, Kinesis)
» A tech-enthusiast excited to work with Cloud Based Technologies (GCP, Azure, AWS)
» A doer who loves to produce meaningful analytic insights for an innovative, data-intensive products,
» Always curious about analytics frameworks and you are well-versed in the advantages and limitations of various big data architectures and technologies.
» Believer in transparency & communication
» Coding skills for analytics and data engineering/manipulation (Scala, Java and Python)
» Experience with SQL & NoSQL database systems, S3 & distributed big data technologies including Hadoop and Spark. Knowledge /awareness of orchestration systems like Airflow, NiFi, Pentaho a plus but not required.

The Tools We Use:

Tools can be learned, so please don't shy away from applying if you're a strong engineer! To give you a flavor of our current tools:

» Language: Scala/Java/Python
» Streaming: Spark Streaming, Kafka
» Cloud Technologies: GCP (BigQuery, Cloudproc, Cloudflow, BigQuery, Compute Engine), Azure (HDInsight, data factory)

Other Requirements:

U.S work authorization- We are unable to provide sponsorship