**Job Description**

Data Engineer (Experience: 3-5 years)

**Our ideal candidate has experience with**

* One or more programming languages (Python, Scala, Java, etc.)
* AWS services such as EMR, S3, Athena, Glue, IAM, RDS, etc.
* A diverse set of databases like MongoDB, Redshift, Postgres etc.
* Git and Github
* CI/CD Pipelines

**You might be an especially great fit if you**

* Enjoy wrangling huge amounts of data and exploring new data sets
* Value code simplicity and performance
* Obsess over data: everything needs to be accounted for and be thoroughly tested
* Plan effective data storage, security, sharing and publishing within an organization
* Constantly thinking of ways to squeeze better performance out of data pipelines

**Nice to haves**

* You’ve built an end-to-end production-grade data solution that runs on AWS
* You understand the tradeoffs between different schema designs like normalization vs. denormalization
* In addition to data pipelines, you’re also good with Kubernetes and Terraform

**Bonus**

* You have experience building machine learning pipelines using tools like SparkML, Tensorflow, Scikit-Learn, etc.
* You are familiar with data processing frameworks like Spark
* You are familiar with orchestration tools such as Airflow, Luiji, Azkaban, Cask, etc.
* You understand the differences between different storage formats like Parquet, Avro, Arrow and JSON

**Responsibilities**

As a Data Engineer, you will

* Build large-scale batch and real-time data pipelines from scratch (preferably on AWS)
* Drive best practices in continuous integration and delivery
* Drive optimization, testing, and tooling to improve data quality
* Collaborate with other software engineers, machine learning experts, and stakeholders, taking learning and leadership opportunities that will arise every single day