**Data Engineer**

"The front page of the internet,” Reddit brings over 430 million users together each month through their common interests, inviting them to share, vote, comment, and create across thousands of communities. Come for the cats, stay for the empathy.

As a data engineer, you will build and maintain the data infrastructure tools used by the entire company to generate, ingest, and access petabytes of raw data. A focus on performance and optimization will enable you to write scalable / fault tolerant code while collaborating with a team of top engineers, all while learning about and contributing to one of the most powerful streaming event pipelines in the world.

Not only will your work directly impact hundreds of millions of users around the world, but your output will also shape the data culture across all of Reddit!

**How you will contribute:**

* Refine and maintain our data infrastructure technologies to support real-time analysis of hundreds of millions of users.
* Consistently evolve data model & data schema based on business and engineering requirements.
* Own the data pipeline that surfaces 40B+ daily events to all teams, and the tools we use to improve data quality.
* Support warehousing and analytics customers that rely on our data pipeline for analysis, modeling, and reporting.

**Qualifications:**

* 2+ years of experience writing clean, maintainable, and well-tested code.
* Experience with Python and/or Scala.
* Familiarity with large scale distributed real-time tools such as Kafka, Flink, or Spark.
* Familiarity with ETL design (both implementation and maintenance).
* Bonus points for experience with (or desire to learn) Kubernetes.
* Excellent communication skills to collaborate with stakeholders in engineering, data science, and product.