LOCATION: Remote, but must reside in one of the following areas:

AL - AZ - CA - CO - CT - GA - MD - MI - MO - NV - NJ - NC - TX - VA - WA - WI

Direct full time perm hire

Preference will be given to USA citizens

ONLY H1B candidates will be considered if fully qualified and 2-3 years left on visa.

Our client is a people search engine that both enlightens and empowers our customers. With over 12 billion records and 14 million visitors per month, we reconnect friends, reunite families, prevent fraud, and more.

We are looking for a Data Engineer, Analytics to join our team and help build our NextGen customer data management, analytics, and ML platform. This role will work as part of the larger data science and analytics team to build and maintain business-critical data pipelines and infrastructure.

Responsibilities:

Design and build maintainable and scalable infrastructure for data extraction, preparation, and loading of data from backend services and web browser clients to our data lake and other reporting databases

Build and manage tools to monitor pipeline performance and provide deeper insight into pipeline metrics

Monitor data pipelines and services for issues and ensure identified bugs are routed and resolved

Work with large, complex SQL and NoSQL databases

Create and maintain technical documentation and write well-abstracted, reusable, and efficient code

Essential Requirements:

2+ years of experience in data engineering or software development

Strong competency in SQL and Python programming

Experience with AWS eco-system and working with services like EMR, RDS, Redshift, Kinesis, etc.

Experience working with large data sets in a Hadoop or Spark environment

Strong organizational skills and detail-oriented mindset

BS in Computer Science, Information Systems, or equivalent skills and experience

Extra Points:

5+ years of experience in backend software development

Strong Ruby and/or Java programming skills

Experience with and understanding of ETL tools

Understanding of event data collection tools like Segment Tealium, mParticle, etc.

Tech Stack:

SQL &amp NoSQL databases

Python programming

AWS - (Kinesis, EMR, Redshift etc…)

EMR, RDS, Redshift, Kinesis

Hadoop or spark environment

Ruby or Java programming skills

ETL tools

Segment, Tealium, mParticle (data collection tools)