

Data Engineer

We're looking for a smart, self-motivated Data Engineer to join our team. This role will be responsible for data collection, extraction, maintaining pipelines and actively researching and verifying new data sources.

About you:

- You care about getting the best possible outcome, you have a passion for what you do which you can clearly convey by your actions.
- You have an eye for detail and order, being able to spot problems in code or data which others might miss or take longer to find.
- You have a desire to explore and test concepts, ideas and theories.
- You have a strong sense of responsibility, and the ability to breakdown, estimate and manage your workflow with stakeholders.
- You have a keen interest in cloud computing (we work with AWS) and have some knowledge on cloud and data security, networking and running complex data pipelines in the cloud.

About Us:

We're small (20+ people and growing fast), innovative and varied group, solving big problems in real estate data and analytics. We are seeking enthusiastic, creative, intelligent and fun individuals to join us in helping build the best platform on the market. In return we can offer you a fun and hard-working environment where you can clearly see your contribution to the company's success.

What you'll be doing:

The hire will be working with our team who are responsible for optimising data flow and collection, researching and investigating new data sources, and expanding and maintaining our existing data pipelines. The ideal candidate is an experienced data pipeline builder and data wrangler who enjoys working in a fast paced and interesting environment. The Data Engineer will support our software developers and data scientists on data initiatives and will ensure optimal data delivery architecture is consistent throughout ongoing projects. They must be self-directed and comfortable supporting the data needs of multiple product verticals. The right candidate will be excited by the prospect of supporting our next generation of products and data initiatives.

- Research and assemble large, complex data sets that meet functional / non-functional business requirements.
- Work with stakeholders in the wider Product and Sales teams to assist with data-related technical questions and support their data infrastructure needs.
- Continually improve the infrastructure required for optimal collection, extraction and initial transformation of data from a wide variety of data sources using AWS big data processing technologies such as Glue.
- Work with existing data and analytics experts to strive for greater functionality in our data systems.

Our Stack

The current data collection and extraction processes are written in Go and Python, and our processing and pipelines are written in PSQL, Spark (Python and Scala) using AWS Glue to orchestrate. We work mostly with batch data rather than streams.

Excellent SQL skills are required, and knowledge of window functions, optimising queries, geospatial queries, GIS, and general data wrangling are a must.

We are open to people from a diverse range of backgrounds and would have a preference for people with an understanding of most of the following:

- Python/Go for data collection
- Postgres and PostGIS
- AWS Glue/S3/RDS/EC2
- Spark/Scala is nice to have but not required initially

We are looking for someone with proven experience of data collection and data pipelines, and who is comfortable working with stakeholders and directly with the CTO and Product Owner. We are a relatively small team, so you must be comfortable with conversing across all aspects of development and reacting quickly to new information.