Data Engineer

Salary: 48-60K
Paid Visa & flight
Paid Apartment for three months

Who we are:
- the second largest furniture company after IKEA
- has an unique logistics-system in Germany
- exist over 50 years in fourth generation
- 19 centers, 20.000 employees, 6 warehouses
- located in Berlin Schönefeld, very modern office

Role
Help us with becoming a more data-driven company. Contribute across the full analytics and data-driven technology stack, from data acquisition to managing databases, the ETL- and the deployment process. As a Data Engineer you will work in an agile cross-functional team with Product Owners, Developers, Business Analysts and Software Testers to drive our online shops and other digital products forward. You will help us getting the right data and ensure we are making use of it the correct way.

Common Tasks
You will support and work closely with our Data Scientists. Together you will be responsible for creating and maintaining end-to-end data products. Your tasks will include:
• Manage and maintain multiple data-pipelines.
• Typical questions: How do we get data from a to b? How frequently do we have to push/ pull the data? In what structure/ format do we save the data and how do we make it accessible?
• Making sure all production tasks are working properly in terms of execution and scheduling
• Typical questions: Do we have dependencies on some tasks? What to do when a task fails? How often should we schedule and execute certain tasks?
• Support the data scientists in making their code more maintainable and scalable
• Typical questions: What environments to use? Where to place and how to design the models?
• Identify repeatable tasks/ routines and implement abstraction layers that make our lifes easier
• Typical questions: How can we make tools and applications easier to maintain and more flexible? How do we adapt to changes?
• Together with the help of the devops-team: define and design a sustainable and secure cloud datainfrastructure
• Typical questions: What cloud instances to use? What databases to use? How do multiple services and instances communicate with each other?
Profile
• Very good Python programming skills
• Very good understanding of database and storage technologies (eg. MySQL/ MariaDB, PostgreSQL, Redshift, S3)
• Good understanding of Apache big data solutions (such as Hadoop, Airflow and Spark)
• Working experience with AWS cloud services (eg. Kinesis, S3, Redshift, EMR)
• Understanding of application and cloud security
• English required

Bonus
• Working experience with SnowPlow Event Tracking
• Experienced with Kubernetes and Docker
• German language skills, in addition to English, are a plus