

Data Engineers

Location: San Francisco CA

Duration: 12+ Months

Working with engineers and analysts, you'll have deep support and direct access to tools and resources to consume an ocean of historical data to produce actionable predictions of the future. Working directly with analytics, you'll help to educate business and executive leadership, contributing to goals, scenario analysis, and messaging. Data Engineering team is a highly-driven group of data, back-end and full-stack engineers who use agile methodologies to build and maintain stable, performant, maintainable and testable data storage.

- Experience generating seasonal/periodic forecast models
- Experience generating linear/logistic regression models
- Experience coding in Python3
- Extensive experience querying in SQL
- Experience in automating data processing
- Experience working with GitHub or other source control
- Software development lifecycle in a professional engineering setting: environments, testing, deployment
- Experience generating more advanced predictive models: decision trees, random forest, gradient boosting
- Experience at a similar scale of data processing (Multi-TB/billions of rows) in large-scale distributed data warehouses
- Create and maintain optimal data pipeline architecture,
- Assemble large, complex data sets that meet functional / non-functional business requirements.
- Identify, design, and implement internal process improvements: automating manual processes, optimizing data delivery, re-designing infrastructure for greater scalability, etc.
- Build the infrastructure required for optimal extraction, transformation, and loading of data from a wide variety of data sources using SQL and AWS 'big data' technologies.
- Build analytics tools that utilize the data pipeline to provide actionable insights into customer acquisition, operational efficiency and other key business performance metrics.
- Work with stakeholders including the Executive, Product, Data and Design teams to assist with data-related technical issues and support their data infrastructure needs.
- Keep our data separated and secure across national boundaries through multiple data centers and AWS regions.
- Create data tools for analytics and data scientist team members that assist them in building and optimizing our product into an innovative industry leader.
- Work with data and analytics experts to strive for greater functionality in our data systems.

Qualifications for Data Engineer

- Bachelors/Masters Degree in a related business or technical discipline, or the equivalent
- Advanced working SQL knowledge and experience working with relational databases, query authoring (SQL) as well as working familiarity with a variety of databases.

- Experience building and optimizing 'big data' data pipelines, architectures and data sets.
- Experience performing root cause analysis on internal and external data and processes to answer specific business questions and identify opportunities for improvement.
- Strong analytic skills related to working with unstructured datasets.
- Build processes supporting data transformation, data structures, metadata, dependency and workload management.
- A successful history of manipulating, processing and extracting value from large disconnected datasets.
- Working knowledge of message queuing, stream processing, and highly scalable 'big data' data stores.
- Strong project management and organizational skills.
- Experience supporting and working with cross-functional teams in a dynamic environment.
- We are looking for a candidate with 5+ years of experience in a Data Engineer role, who has attained a Graduate degree in Computer Science, Statistics, Informatics, Information Systems or another quantitative field. They should also have experience using the following software/tools:
 - Experience with big data tools: Hadoop, Spark, Kafka, etc.
 - Experience with relational SQL and NoSQL databases, including Postgres and Cassandra.
 - Experience with data pipeline and workflow management tools: Azkaban, Luigi, Airflow, etc.
 - Experience with AWS cloud services: EC2, EMR, RDS, Redshift
 - Experience with stream-processing systems: Storm, Spark-Streaming, etc.
 - Experience with object-oriented/object function scripting languages: Python, Java, C++, Scala,