

# Senior Data Engineer

## U.S Based Remote

As the leading independent modern media company, Vox Media ignites conversations and influences culture. Across digital, podcasts, TV, streaming, live events, and print, we tell stories that affect our audience's daily lives and entertain as much as they inform.

Our portfolio features influential and respected editorial properties including Vox, New York Magazine, The Verge, The Cut, Eater, Vulture, The Strategist, Polygon, SB Nation, Intelligencer, Curbed, Grub Street, and Recode. Off-platform, the Vox Media Podcast Network offers one of the largest collections of popular podcasts, and Vox Media Studios produces and distributes award-winning nonfiction shows. Powered by innovative technology that scales quality, the Chorus publishing platform and Concert advertising marketplace answer the always-changing needs of modern audiences, creators, and marketers.

Vox Media has been named one of Fast Company's "Most Innovative Companies in Media," an Inc. "Company of the Year," Digiday's "Best Company for Parents," and one of the Best Places to Work for LGBTQ Equality by the Human Rights Campaign.

Vox Media is also a remote-first company. Our teams have been and always will be distributed across many regions and time zones. We work and laugh together over Zoom, Slack, and collaborative documentation.

## About Vox Media's Data Team:

Our data team is a small, but a growing, knowledgeable group of data enthusiasts. The team exists to provide Vox Media stakeholders with democratized and integrated data solutions to gain clarity, insights, and scale on first-party data. We work full-stack, playing with infrastructure, data stores, APIs and so much more.

Our goal is to become a world-class data engineering team leading the way in developing production-ready and state-of-the-art data science solutions, and we're looking for a passionate senior data machine learning engineer to join our team. This role is part of the Revenue Product group and will report to the Revenue Data Engineering Manager. We are excited for you to join us!

## What you'll do:

- Evaluate, revamp, and scale data processes and needs for team stakeholders, downstream internal customers, and ML models.
- Run statistical analysis and generate summary reports and distributions of relevant data features to aid in research, reporting, and machine learning.
- Work with data to empower others to access and understand answers to their own data questions.
- Develop scripts, parsers, and applications to process, load, analyze data to solve business problems, build and maintain the infrastructure to bring data capabilities to the Revenue Data team.
- Monitor data errors and quality control on ingested data, performing the necessary steps needed to ensure delivery of timely, accurate, and available data.

- Mentor and guide future data engineers on the team to manage additional data engineering tasks, responsibilities, and projects.
- Contribute to the team's knowledge base in the form of documentation.
- Actively track and communicate task status in Jira and Slack.
- Be a part of a thoughtful and energetic engineering team at Vox.
- Have fun tackling challenges and learning within a diverse team of Engineers, Data Analysts, Product Managers, and Project Managers to deliver integrated, democratized, and impactful data solutions.

## What you'll bring

- **Expertise in the following technologies:**
  - Python (Tensorflow, PySpark, Pandas, Sci-kit Learn, MLlib), Scala, SQL, Spark.
- Experience building or maintaining ETL processes.
- Expertise in developing solutions to analyze large data sets efficiently.
- Expertise in data processing and data wrangling.
- Knowledge in Data Warehousing or Data Lake design and implementation.
- Willingness to work with remote-first teams; comfort communicating remotely through Slack, docs, video calls, and diagrams.
- Ability to work independently as well as collaboratively with multi-disciplinary teams.
- Ideas, opinions, and the ability to share them through respectful code and design reviews, proposals, calls, diagrams, and team-wide discussions.
- Passion for staying curious, learning from, and sharing knowledge.

## Bonus points for:

- Experience developing large-scale distributed applications.
- An understanding of Machine Learning techniques including deep learning (neural networks) and unsupervised machine learning (clustering, feature selection, etc).
- Proficiency in the model validation process and optimizations.
- Experience exposing applications as web services or API endpoints.
- Experience with Big Data.
- Having used any of the following technologies:
  - Apache Kafka, Airflow, NiFi, and Flume
  - Google Cloud Platform (GCP): Dataproc, GKE, Spanner, etc
  - Node.js
  - Elasticsearch, Logstash, Kibana
  - BI Tools like Tableau, Looker, or PowerBI
- Front-end development expertise.
- Experience in debugging failed production jobs.
- Awareness of data privacy practices (GDPR and CCPA)

For more on the things we're building and problems we're solving, and what it's like to work with the Product, Design & Technology teams, see our [Product Team blog](#). In an effort to provide an inclusive and respectful working environment, all team members adhere to our [code of conduct](#).

## About working at Vox Media:

This is a permanent, full-time position with excellent benefits—including flexible hours and generous parental leave. Vox Media strives to provide comprehensive healthcare options for our employees and to ensure that our healthcare and other benefits are LGBTQ-inclusive. You'll be joining a group of focused, hard-working, creative people who are passionate about doing work that's challenging and fun—and who strive to maintain a healthy work/life balance.

Vox Media is committed to building an inclusive environment for people of all backgrounds and everyone is encouraged to apply. Vox Media is an Equal Opportunity Employer and does not discriminate on the basis of race, color, gender, sexual orientation, gender identity or expression, religion, disability, national origin, protected veteran status, age, or any other status protected by applicable national, federal, state, or local law.