Choose Your Own Adventure

Welcome to the adventure into the Data team at Vox Media. 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. We do this by providing Vox Media Stakeholders with democratized and integrated data solutions that help them gain clarity, insights, and scale with first-party data.

How To Prepare

For the Data Engineer role, you will participate in a 45-minute technical panel interview with the Revenue Data team. Choose any **ONE** of the following prompts and come prepared to present and discuss.

Spend a maximum of 2 hours on this part of the choose your own adventure exercise.

Technical Prompts

Prompt One: Around the world in 45 minutes?

This is wild west style and we're taking a trip around the Data Engineering world with stops at the following attractions:

- The Zoo of Serpents: SQL and Python
    - What is your level of experience with SQL and Python?
    - What tools have you used in the past to develop data solutions?
    - Have you worked in ML teams before? What libraries are used to label and clean datasets?
- Le Data Modelé: for data warehousing and Big Data
    - Explain why staging areas are useful.
    - What is Data Mining?
- Express Train Lane: Or just... ETL
    - What is structured versus unstructured data?
    - What is ETL?
    - What happens after a data pipeline is available?
- Institute of Mathematica:

    - What the benefits of using MapReduce functions?
    - Explain the Pub/Sub model.
- The Museum of Visual Arts:
    - What's your experience in software engineering background?
    - Do you have experience or an interest in Javascript or Data visualization?

Review these questions and come prepared to share your answers in the technical panel.

Prompt Two: Is it the end of Data Warehousing?

The data warehouse architecture has been in the IT industry for at least three decades now. Read the Lakehouse paper and come ready to discuss the following:

- What are some of the considerations Data Engineers should have when architecting data platforms and pipeline processes for consumption by data stakeholders? (We will ask about reducing costs and optimizing query computation)
- What are some other methods and designs for achieving Lakehouse goals?
- How will you measure the impact and performance of data management?

Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics: http://cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf

Prompt Three: Project Mania- A Trip Down Memory Lane

We'd love to see any relevant past work you have completed! Use this time to determine what to demo. Come ready to share your role in the design and implementation of the data solution.

Feel free to share any slides, links, or documentation to the work ahead of time by sending an email to rp.dataproduct@voxmedia.com with your full name in the subject line.

Here is an outline of the panel topics and kinds of questions to expect:

- What was the use case and scope of the project?
- What data sources were used?
- How data was retrieved, ingested, transformed, and loaded from data?

- What tools did you use and was this process automated, if not how could it be achieved?
- How did you use other software programming or ML skills to do this work?
- What errors and challenges did you run into?
- And any future work and considerations you'd like to highlight?

Questions?

Reach out to your recruiter contact or tiffany.jachja@voxmedia.com