



Data Engineer Take Home Challenge

Overview

Problem Statement

What We're Looking For

Guidelines

Tips

Submission

Overview

At Nansen, we help crypto investors discover alpha and make better investment decisions using onchain data. Your job is not just to write code, but to think like a builder: understand the problem space, propose a scalable architecture, plan a path forward, and demonstrate hands-on skills by executing one part of the pipeline.

Design and implement a data pipeline that generates valuable labels for blockchain addresses, showcasing your data engineering expertise, architectural thinking, and ability to build scalable solutions.

This challenge evaluates your technical skills and your approach to handling ambiguous requirements - a key aspect of working at Nansen. You may use AI coding assistants to help with implementation.

Problem Statement



"Create a pipeline that produces valuable labels for blockchain addresses. These labels should be useful to an investor who wants to better understand interesting addresses on chain, and help them make smarter decisions."

We intentionally keep the problem open-ended to simulate what working at Nansen is like: we value initiative, structure, and clarity of thought in the face of ambiguity.

What We're Looking For

We'll assess your submission across **four dimensions**:

1. Architectural Thinking

- Can you propose a clear, well-structured architecture for the pipeline?
- How do you handle scaling, maintainability, and data quality?

2. User-Centric Design

- Can you reflect on what makes a label valuable?
- Have you thought about the end-user and what they'd want?

3. Project Management Approach

- Can you break the problem into milestones, steps, and isolated tasks?
- Are you able to prioritise what should be built first?

4. Execution (Code Quality)

- Choose **one part** of your proposed system and implement it.
- This could be a data ingestion component, a labelling algorithm, a database model, or an API layer.

Guidelines

- Timebox your work to max 4 hours total. The deliverable is not meant to be exhaustive, use the task breakdown to highlight areas you would have dived deeper into
- You're welcome to use any stack you're comfortable with (e.g., Python, dbt, SQL, notebooks etc.)

Tips

- Be opinionated, and explain your choices and trade-offs.
- Think from the user's perspective, what would you want to see if you were using this data to trade or invest?
- You can find publicly available onchain data in BigQuery in `bigquery-public-data.crypto_ethereum`
- If you looking for inspiration for labels, take a look at [this](#) blog post

Submission

Send your work as a GitHub repository with complete code and documentation.

Your submission should include:

- A working code sample of one chosen part of the pipeline.
- README with:
 - Setup instructions
 - Architecture overview
 - Technology choices and trade-offs
 - Future improvements roadmap

After we've reviewed your submission, we'll invite you to a 60 minute walkthrough interview. You'll get **20 minutes** to present your work, before we'll move into questions and discussion.

If you're stuck or have questions, feel free to reach out.
