Interview Preparation: Data Analyst – Data & Web Scraping

Brahim Ouhammou

Personal Introduction

Hello, my name is Brahim Ouhammou. I am a student in Artificial Intelligence and Machine Learning at the Faculty of Sciences, and I am also currently studying at 1337 Coding School.

I have a strong foundation in programming and problem solving, with skills in Python, C, and C++, as well as experience in machine learning, deep learning, and data analysis. I have worked on practical projects involving CNN-based models, AI-driven systems, and software development, including backend development and Docker-based infrastructures.

As part of my training at 1337, I am seeking a 6-month internship starting at the end of February, where I can apply my skills, learn from real-world projects, and continue developing as an AI and software engineer.

Technical Questions – Web Scraping & Python

1. **What Python libraries do you use for web scraping?**

   - `requests`, `BeautifulSoup`, `Scrapy`, `Selenium` for dynamic content.

2. **How do you handle websites with dynamic JavaScript content?**

   - I use `Selenium` or `Playwright` to simulate a real browser.

3. **How do you avoid getting blocked while scraping?**

   - I use realistic headers, rotate proxies, add delays, and respect `robots.txt`.

4. **What is a REST API and how do you use it in Python?**

   - It's an interface for data exchange. I use `requests.get()` or `requests.post()`.

5. **How do you structure scraped data?**

   - I store it in CSV, JSON, or databases like SQLite or PostgreSQL.

Technical Questions – Data Analysis

1. **What Python libraries do you use for data analysis?**

   - `pandas`, `numpy`, `matplotlib`, `seaborn`.

2. **How do you handle missing or incorrect data?**

   - I use `fillna()`, `dropna()`, or interpolation, depending on the context.

3. **How do you detect outliers?**

- Using visualizations (box plots) or statistical methods (IQR, Z-score).

4. **What is an SQL join? What types do you know?**

   - INNER, LEFT, RIGHT, FULL. Used to merge tables.

5. **How do you present your findings?**

   - Dashboards (Power BI, Tableau), Jupyter reports, or simple presentations.

Project & Experience Questions

1. **Tell me about a scraping project you've worked on.**

   - I scraped plant data for a CNN classifier using `BeautifulSoup` and stored it in CSV.

2. **How do you ensure the quality of scraped data?**

   - I validate formats, check for duplicates, and run unit tests on extractors.

3. **Have you worked with large datasets?**

   - Yes, in ML projects (images, logs). I use `pandas` with optimization or batch processing.

4. **How do you handle website structure changes?**

   - I use robust selectors, run regular tests, and set up failure notifications.

5. **Have you used cloud or no-code scraping tools?**

   - No, I prefer code for flexibility and control.

Soft Skills & Behavioral Questions

1. **How do you organize teamwork?**

   - I use Git, project tools like Trello/Notion, and maintain clear communication.

2. **How do you stay updated with technology?**

   - I follow tech blogs, Reddit, take MOOCs, and work on personal projects.

3. **What is your greatest strength?**

   - Fast learning and independent problem-solving.

4. **Why do you want to work here?**

   - To apply my scraping and data skills in a real-world business context with impact.

5. **Do you have any questions for us?**

   - Yes:
     - What are the main data challenges your team is facing now?
     - What's your current tech stack for scraping and analysis?