Experience & Technical Skills

Can you describe a past project where you worked with GIS tools like OpenStreetMap, QGIS, PostGIS, or GeoPandas? If not, which GIS tools would you choose and why?
No, I don't have any experience using these GIS tools. I would choose GeoPandas since it has a simple and strong integration with Python libraries. I guess it offers a relatively smooth learning curve.

What are your experiences developing back-end applications in Python? If you primarily use another language, how would you transfer your skills to Python?
Yes I have experience developing back-end in Flask(Python framework) during one of my previous internships and in my last full-time position I used python and crontab for daily vulnerability updates. I use Node.js for back-end primarily and languages are not the main skills for coding. As long as I have back-end development skills, it is easy to transfer skills to Python. It only takes a short time to go through documents for Python libraries.

Have you used Docker? If so, in which scenarios? How did it improve your development or deployment process?
Yes, I have used Docker in my last full-time job. I used Docker and nginx to deploy the portal page including multiple services. I made 4 containers for PostgreSQL, crontab, software developments and nginx. It helped me a lot in both the development and deployment process. When I make changes to my back-end services, I do not need to shut down other containers since every container is individual. Additionally, I can have multiple databases for development and production at the same time with Docker.

Do you use GPT (or other LLMs) or Copilot for coding? Please be honest. If yes, how do you typically use it? What are its strengths and limitations?
Sure I have used LLM for coding a lot even in my previous job. By providing a clear description of the desired function and the expected output type, LLMs could save a significant amount of development time. The strength is that LLMs are good at data processing and it could save my time. However, the solutions they provide can sometimes be incomplete or incorrect, requiring careful validation and adjustments

Could you describe the steps to add a new function to a backend system, from requirements gathering to deployment?
The first step is having a meeting with stakeholders to understand the purpose of the new function, afterwards I need to design the function and potential database changes/migration. After implementation, writing unit tests and API documents is necessary. The last step is to deploy the changes.

Can you share an example of how you optimized a data processing pipeline? What tools or strategies did you use? How did you handle SQL database optimizations?
I used to process data for over 60 thousand users, which is time-consuming. I used batch processing which loads 1000 users per batch and the performance gets much better. I do not

have much experience for database optimizations. But in database design, breaking a huge table into multiple smaller tables and joining them if you need will be helpful for performance.

Are you a good Git user? What are your habits when using Git? Would you consider yourself an expert?
I believe I am a good Git user. For using Git in back-end development, I usually use **git checkout -b branch_name** to create a new branch to add new changes and **git commit -m "message"** for each commit. Afterwards, I use **git push origin branch_name** to push my local changes to remote branches for code review. For merge, merge conflict and rebase, it is not that hard. I would not consider myself an expert since it is just a version control tool for software development.

What factors do you consider when choosing between multiprocessing, multithreading, or asynchronous execution for a backend?
If we have individual data processing or parallel jobs, I would prefer multiprocessing. For fetching the APIs, I think async executions are more appropriate. Web servers that take multiple requests may need multithreading, but I do not have experience with it.

Imagine the backend system you are working on suddenly starts processing sensor data much slower than expected. What steps would you take to diagnose and resolve the issue?
I would consider to add print() function for each part of processing data and figure out the most time-consuming part. And check the CPU/memory usage to see if we need multiprocessing there. Also it could be related to the way we process data. We can check the complexity of data processing.

If you were designing a system to process and analyze real-time roadway sensor data from thousands of vehicles, what key architectural decisions would you make? How would you ensure scalability and reliability?"
For real-time sensor data, I feel we should use stream-based data pipeline to handle it. And using multiprocessing since there are thousands of vehicles and the data should be independent. I would use Docker and Kubernetes for scalability.

Motivation & fit

What interests you most about this position and working with the GT Smart City Infrastructure research team?
The opportunity to use technologies I have not learned like GIS related technology and LiDAR and gaining hands on backend development experiences interest me the most.

How do you see this role contributing to your academic and professional growth?
As this role could be my first research assistant position, I would say it is a good chance for me to understand how the research team works. And I also could learn more about back-end development with Python libraries to process data

Have you worked in a multidisciplinary team? How did you ensure effective communication between technical and non-technical members?
Yes, I have worked in many multidisciplinary teams. In my previous interns I’ve worked with investment brokers and . I believe having concise documentations and frequent updates are important factors to ensure communication between technical and non-technical members.