

How would you implement authentication in this system?

To enforce access control on the endpoint, there are two main methods:

- Custom token: the server expects the client to send a specific token that acts as an access key. This token doesn't carry any user information.
- JWT (JSON Web Token): the server generates a digitally signed token that contains encoded data such as username, permission, user id, etc. It's useful for handling authentication and authorization in user-based systems.

Since we don't have a user system, the most practical solution here is to use a custom token.

If you had to scale this system up to serve 1000s of requests per hour, how would you do it?

To scale our system we would need to move beyond the basic Flask development server.

There are two main options to consider:

- Azure Functions (Serverless Architecture): is a serverless compute service that automatically handles infrastructure, scaling and concurrency. Can't keep models loaded in memory.
- Gunicorn+Docker (Containerized Deployment): Gunicorn is a production-ready WSGI server for Python applications which supports multiple workers and concurrent processing. We use it with Docker to containerize the app, making sure this way that the environment is isolated, portable and easy to scale. This allows us to keep the model to be loaded in memory all the time.

Given our system (which needs to have a consistent model running) the best solution is to use Gunicorn+Docker.

How would you implement logging in your API?

No front end for this exercise.

How would you test this API?

To test the API we could use frameworks made for this purpose, like pytest.

Sources

https://medium.com/@ceo_44783/the-easiest-way-to-scale-your-api-to-thousands-of-users-74b9e5dcdd94

<https://learn.microsoft.com/en-us/azure/azure-functions/functions-overview>

<https://learn.microsoft.com/en-us/azure/devops/pipelines/agents/scale-set-agents?view=azure-devops>

<https://docs.gigaspace.com/latest/dev-dotnet/scaling-agent-example.html>

<https://flask.palletsprojects.com/en/latest/testing/>

<https://stackoverflow.com/questions/47892413/how-to-secure-a-rest-api-on-flask>