

# Deploy

To deploy a model, you must first have a version of that model. If you have not yet created a version, please refer to the [versions](#) documentation.

To create a new service, users can employ the `deploy` command. This command facilitates the deployment of a machine learning service ready to accept predictions at the `/cairo_run` endpoint, providing a straightforward method for deploying and using machine learning capabilities that can easily be consumed as an API endpoint.

...

Copy

```
giza endpoints deploy --model-id 1 --version-id 1 model.sierra ##### Creating endpoint! [giza][2024-02-07 12:31:02.498] Endpoint is successful ✓ [giza][2024-02-07 12:31:02.501] Endpoint created with id -> 1 ✓ [giza][2024-02-07 12:31:02.502] Endpoint created with endpoint URL: https://deployment-gizabrain-38-1-53427f44-dagsgas-ew.a.run.app
```

...

If a model is fully compatible the sierra file is not needed and can be deployed without using it in the command:

...

Copy

```
giza endpoints deploy --model-id 1 --version-id 1 ##### Creating endpoint! [giza][2024-02-07 12:31:02.498] Endpoint is successful ✓ [giza][2024-02-07 12:31:02.501] Endpoint created with id -> 1 ✓ [giza][2024-02-07 12:31:02.502] Endpoint created with endpoint URL: https://deployment-gizabrain-38-1-53427f44-dagsgas-ew.a.run.app
```

...

For a partially compatible model, the sierra file must be provided, if not an error will be shown.

Example request

Now our service is ready to accept predictions at the provided endpoint URL. To test this, we can use the `curl` command to send a POST request to the endpoint with a sample input.

...

Copy

```
curl -X POST https://deployment-gizabrain-38-1-53427f44-dagsgas-ew.a.run.app/cairo_run \ -H "Content-Type: application/json" \ -d '{ "args": "[2 2] [1 2 3 4]" }' | jq { "result": [0.1234], "request_id": "b14bfbcf250b404192765d9be0811c9b" }
```

...

There is an extra `args.job_size`, that can be used in each request to specify the size of the proving job so it has more CPU and memory available to generate the proof. An example:

...

Copy

```
curl -X POST https://deployment-gizabrain-38-1-53427f44-dagsgas-ew.a.run.app/cairo_run \ -H "Content-Type: application/json" \ -d '{ "args": "[2 2] [1 2 3 4]", "job_size": "M" }'
```

...

Available sizes are `S`, `M`, `L`, and `XL`, each with different usage limits.

List the proving jobs for an endpoint

To list the proving jobs for an endpoint, we can use the `list-jobs` command available for the endpoints. This command will return a list of all the proving jobs for the endpoint with their `request_id` for easier tracking.

...

Copy

```
giza endpoints list-jobs --endpoint-id 1 [giza][2024-03-06 18:13:50.485] Getting jobs from endpoint 1 ✓ [ { "id": 1, "job_name": "proof-20240306-979342e7", "size": "S", "status": "Completed", "elapsed_time": 120., "created_date": "2024-03-06T16:12:31.295958", "last_update": "2024-03-06T16:14:29.952678", "request_id": "979342e7b94641f0a260c1997d9ccfee" }, { "id": 2, "job_name": "proof-20240306-f6559749", "size": "S", "status": "COMPLETED", "elapsed_time": 120.0, "created_date": "2024-03-06T16:43:27.531250", "last_update": "2024-03-06T16:45:17.272684", "request_id": "f655974900d8479c9bb662a060bc1365" } ]
```

...

List the proofs for an endpoint

To list the proofs for an endpoint, we can use the `list-proofs` command available for the endpoints. This command will return a list of all the proofs for the endpoint with the `request_id` for easier tracking.

...

Copy

```
giza endpoints list-proofs --endpoint-id 1 [giza][2024-03-06 18:15:23.146] Getting proofs from endpoint 32 ✓ [ { "id": 1, "job_id": 1, "metrics": { "proving_time": 0.03023695945739746 }, "created_date": "2024-03-06T16:44:46.196186", "request_id": "979342e7b94641f0a260c1997d9ccfee" }, { "id": 1, "job_id": 2, "metrics": { "proving_time": 0.07637895945739746 }, "created_date": "2024-03-06T16:44:46.196186", "request_id": "f655974900d8479c9bb662a060bc1365" } ]
```

...

Download the proof

We can download the proof using the `download-proof` command available for the endpoints:

...

```
Copy ▶ giza endpoints download-proof --model-id 1 --version-id 1 --endpoint-id 1 --proof-id "b14bfbcf250b404192765d9be0811c9b" [giza][2024-02-20 15:40:48.560] Getting proof from endpoint 1 ✓ [giza][2024-02-20 15:40:49.288] Proof downloaded to zk.proof ✓
```

...

The proof id used is the `request_id` returned in the response.

[Previous](#) [Transpile](#) [Next](#) [Prove](#)

Last updated 1 day ago