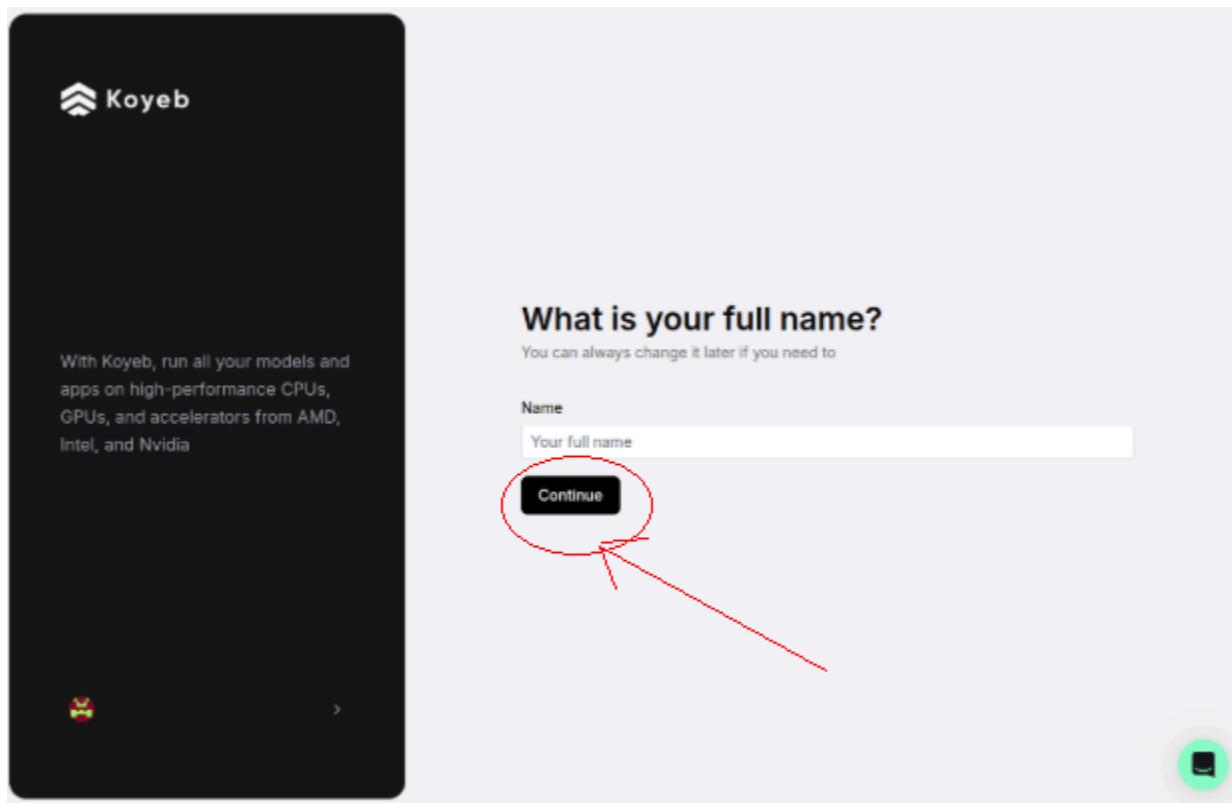
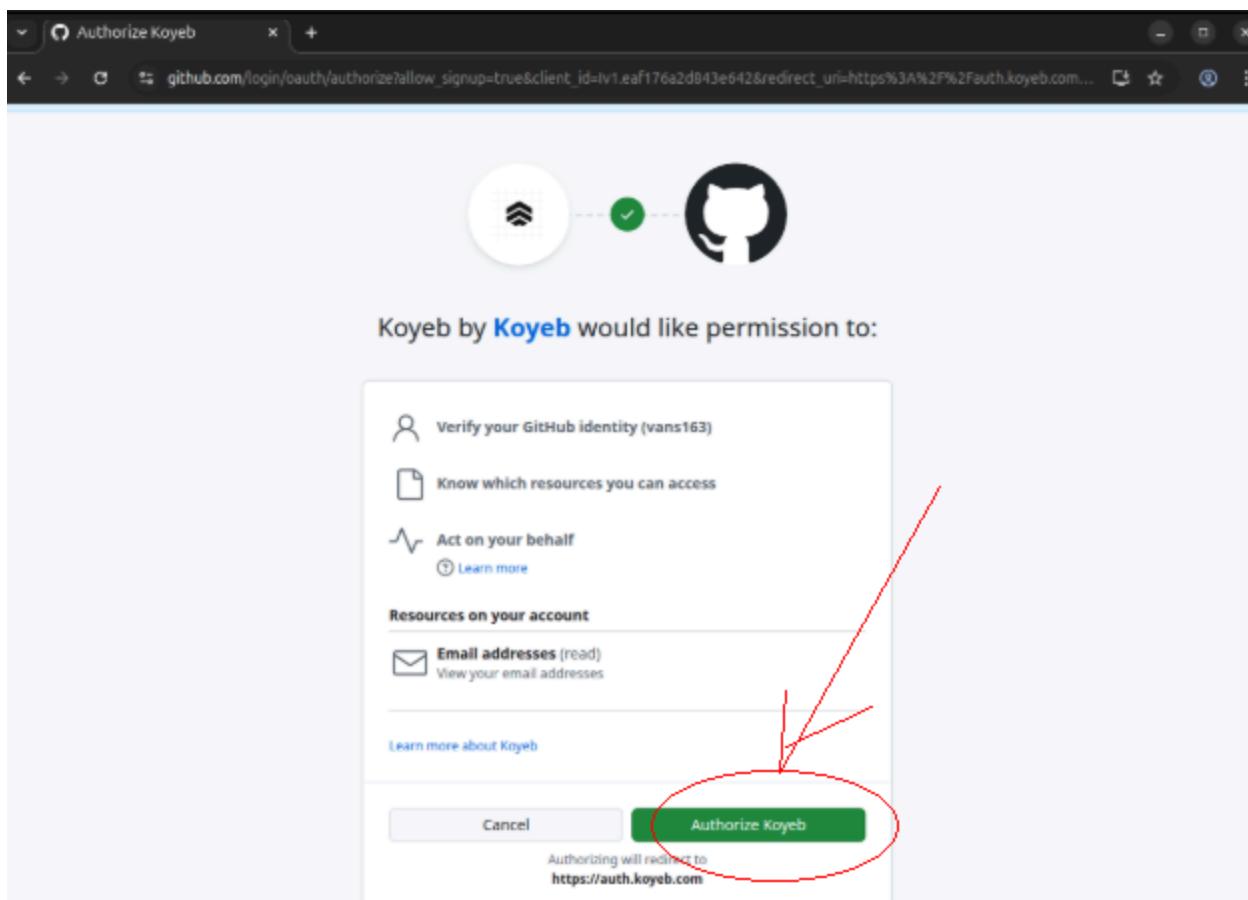


This screenshot shows the 'Sign up for Koyeb' page. It has a white header with the 'Sign up' title. The main form contains fields for 'Your first name', 'Your last name', and 'Your email address', followed by a 'Continue' button. Below this is a 'Continue with GitHub' button, which is circled in red. A red arrow points from this circle to the 'SIGN UP' button on the Koyeb homepage above. The right side of the page features a dark sidebar with the Koyeb logo and the tagline 'DEPLOY INTENSIVE APPLICATIONS ACROSS GPUS, CPUS, AND ACCELERATORS IN MINUTES - SCALE IN 50+ LOCATIONS'. It also lists several features: 'AUTOSCALING WITH SUB-200MS COLD-START', 'SERVERLESS CONTAINERS', 'SANDBOXED CODE EXECUTION', and 'GLOBAL AVAILABILITY AND FAILOVER'. At the bottom, it says 'Trusted by the most ambitious teams' and lists logos for Databricks, Stanford, Strapi, PhotoRoom, scenario, and eToro. A small note at the bottom of the sidebar says 'Need help or want to learn more, talk to an expert'.



Your Pro trial ends in 6 days. To maintain access to Pro features, upgrade now.

Settings

General Usage and billing Plans API Registry configuration

Usage summary

You are currently trialing the Pro plan. You still have 6 days left. [View all plans and upgrade →](#)

Download usage

Manage billing

Access your invoices and manage your payment methods

Manage billing

Billing alerts

Receive email notifications when your monthly usage reaches 80% and 100% of the budget threshold set for your organization

\$ / month Save

Promo codes

Redeem a promo code

Redeem

TTDEPLOY25FADEV2M

Your Pro trial ends in 6 days. To maintain access to Pro features, upgrade now.

Settings

General Usage and billing Plans API Registry configuration

API access tokens

Name	Description	Created	⋮
test		3 minutes ago	

[Create API token](#)

Create new service - Koyeb | tenstorrent docker image | Build AI Models with Ten... | +

app.koyeb.com/services/new?service_type=private

Your Pro trial ends in 6 days. To maintain access to Pro features, upgrade now.

[Create new service](#)

Let's deploy a new service

Create service

- Web service** (selected)
- Private service
- Sandbox
- Worker
- Database

Deploy one-click app

- Bun
- Cal.com
- Ollama

[More one-click apps →](#)

Create private service
Only exposed to your other services via the internal private network

- GitHub**
Connect, build and deploy your code from a GitHub repository
- Docker**
Deploy docker images from any public or private registries

Create new service - Koyeb | tenstorrent docker image | Build AI Models with Ten... | +

app.koyeb.com/services/new?service_type=web&step=importProject&type=docker

Your Pro trial ends in 6 days. To maintain access to Pro features, upgrade now.

Create a new service

Deploy a Docker image from a container registry

Import project 1 of 3

Deploy a Docker image

Image
The Docker image and tag you want to deploy
 ✓ The image is accessible

Registry
If your image is private, create or select the registry configuration secret to allow Koyeb to pull your image

[Back](#) [Next](#)

ghcr.io/tenstorrent/tt-xla/tt-xla-ird-ubuntu-22-04:latest

Your Pro trial ends in 6 days. To maintain access to Pro features, upgrade now.

Create a new service

Pick the instance and regions to deploy

Choose instance and regions 2 of 3

GPU (circled in red)

RTX-4000-SFF-ADA (New) Low supply \$0.50/h
6 vCPU 20 GB VRAM 44 GB RAM 240 GB Disk

Deploy in regions
Europe (selected) 102ms • Low supply Metropolitan Continental

N300s (New) Private preview High supply \$0.00/h
4 vCPU 24 GB VRAM 32 GB RAM 320 GB Disk

L4 (New) Requires higher quota Low supply \$0.70/h
6 vCPU 24 GB VRAM 32 GB RAM 160 GB Disk

Request increase

The screenshot shows the 'Create new service' step of the Koyeb deployment process. It lists three instance types: RTX-4000-SFF-ADA (selected), N300s, and L4. The RTX-4000-SFF-ADA instance is highlighted with a red oval and has a red arrow pointing to its 'Low supply' price of \$0.50/h. The N300s instance is also highlighted with a red oval. The L4 instance is shown with a red arrow pointing to its 'Requires higher quota' note. The interface includes a sidebar with various icons and a progress bar indicating '2 of 3' steps completed.

Deployment
Configure deployment

Entrypoint ⓘ
entrypoint.sh Override

Command ⓘ
sh -c "python3 -m http.server 8000 & sleep infinity" Override
CMD ["sh", "-c", "python3 -m http.server 8000 & sleep infinity"]

Privileged ⓘ

sh -c "python3 -m http.server 8000 & sleep infinity"

Review and deploy 3 of 3

Service type
Web service

Source
ghcr.io/tentorrent/tt-xla/tt-xla-ird-ubuntu-22-04:latest

Deployment ⓘ (Default configuration)

Environment variables and files
No variables, no files

Instance
N300s [4 vCPU, 32GB RAM, 320GB Disk] GPU North America 🇺🇸

Scaling
Autoscaling (0-1 instances/region)

Volumes
No volumes configured

Ports
Ports 1 configured

Health checks
TCP health check configured on port 8000

Service name
tt-xla-ird-ubuntu-22-04

Deploy

Estimated Cost
Below are the minimum and maximum estimated costs based on your current service configuration. Resources are billed by the second.

0 to 1 N300s instances
Autoscaling enabled

min.	max.
\$0.00/hr	\$0.00/hr
\$0.00/mo	\$0.00/mo

1 region
You will have between 0 and 1 N300s instance running per region

Total cost
min. max.
\$0.00/hr \$0.00/hr
\$0.00/mo \$0.00/mo

Deploy to Koyeb button ⓘ

[! Deploy to Koyeb](https://www.koyeb.com/services/tt-xla-ird-ubuntu-22-04/deploy)

DEPLOY TO KOYEB

Your Pro trial ends in 6 days. To maintain access to Pro features, [upgrade now](#).

tt-xla-ird-ubuntu-22-04 ⓘ

Web service | crowded-emmalynn-amaproto...-428ab29d.koyeb.app/ Redeploy
Koyeb CLI C

Console (selected) Overview Metrics Settings

View active deployment | View latest deployment

f20b281a ⓘ (Healthy) 16m ago

Initial deployment

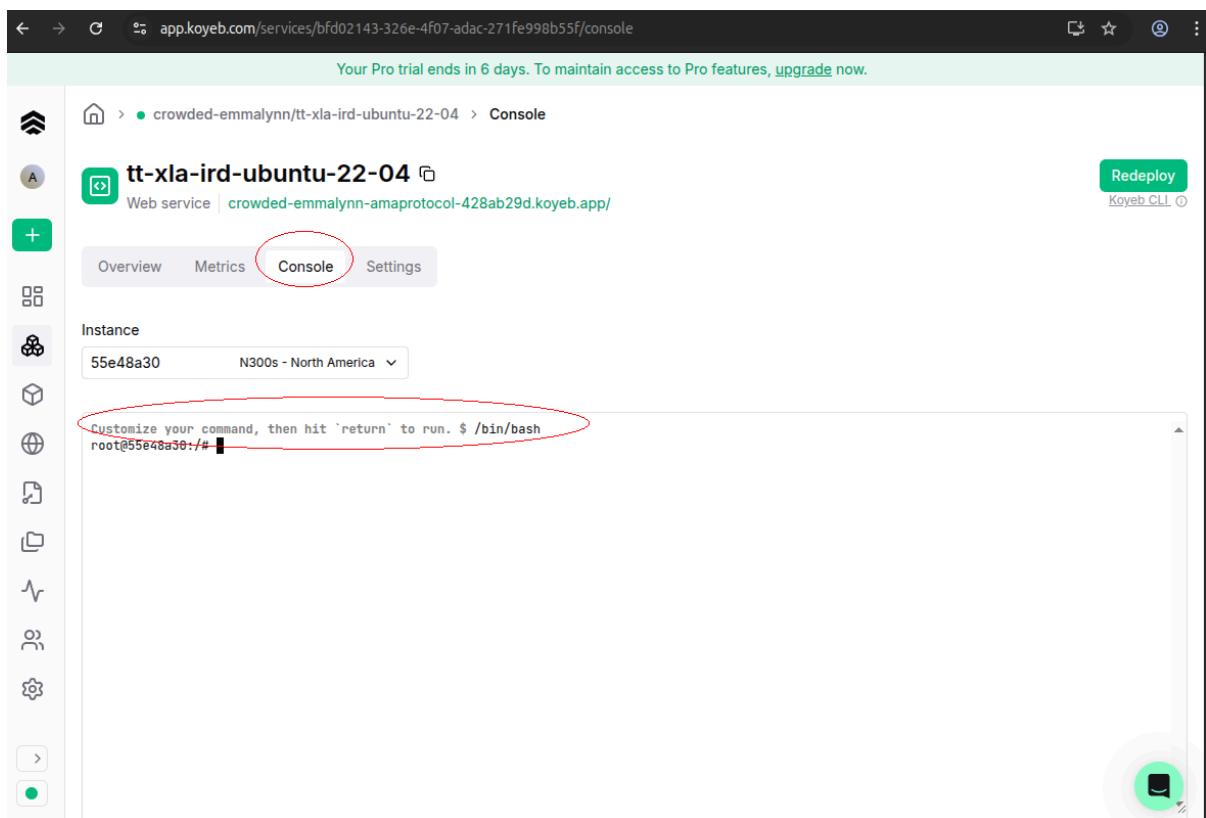
Overview Web service

Public URL: crowded-emmalynn-amaproto...-428ab29d.koyeb.app (forwarded to port 8000) Private Address: tt-xla-ird-ubuntu-22-04.crowded-emmalynn.internal:8000

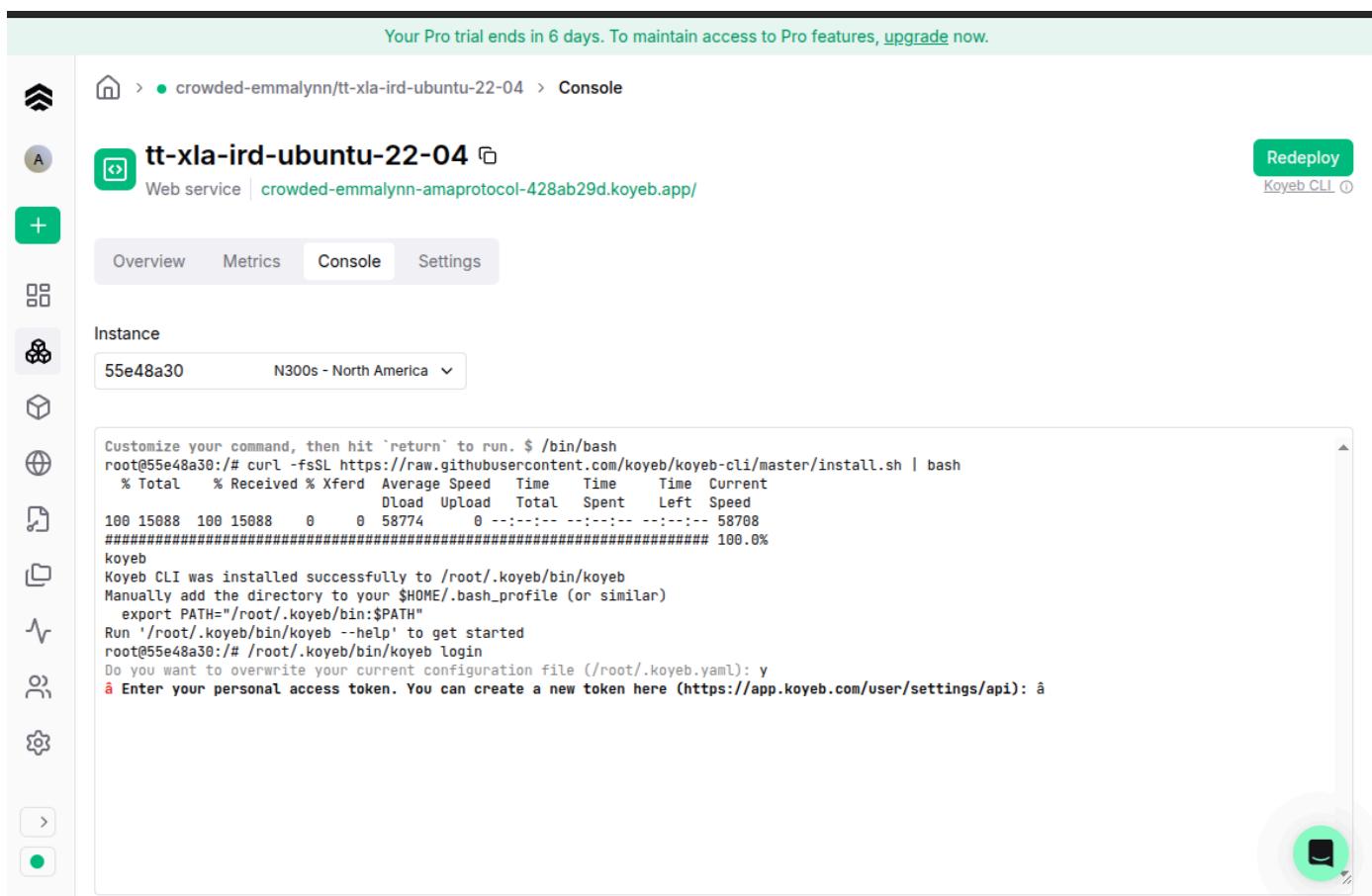
Docker image: ghcr.io/tentorrent/tt-xla/tt-xla-ird-ubuntu-22-04:latest

Instance	Scaling	Regions	Environment	Volumes
N300s	1 of 1 running	North America	0 variables	0 volumes attached
Configure				

[View more](#)



```
curl -fsSL https://raw.githubusercontent.com/koyeb/koyeb-cli/master/install.sh | bash
/root/.koyeb/bin/koyeb login
COPY API KEY
```



Install Locally

```
curl -fsSL https://raw.githubusercontent.com/koyeb/koyeb-cli/master/install.sh | bash  
/root/.koyeb/bin/koyeb login  
COPY API KEY  
koyeb services list
```

```
server returned an internal_server_error (code 20)  
user@ama:/tmp$ /home/user/.koyeb/bin/koyeb services list  
ID          APP           NAME          TYPE    STATUS   CREATED AT  
bfd02143    crowded-emmalynn tt-xla-ird-ubuntu-22-04 WEB    HEALTHY 24 Dec 25 04:19 UTC  
user@ama:/tmp$ /home/user/.koyeb/bin/koyeb service exec bfd02143 /bin/bash  
root@55e48a30:/#
```

```
/root/.koyeb/bin/koyeb service exec <ID> /bin/bash
```