

KEY PARTNERS

- **AWS** — Marketplace org billing; Inference Endpoints on AWS GPUs.
- **Google Cloud** — Vertex Model Garden, GKE, Cloud Run GPUs partnership.
- **Microsoft Azure** — named cloud partner path for HF in customer clouds.
- **Open-model labs & community authors** — fill Hub with weights and datasets.
- **Inference Providers network** — multi-provider serverless routing on Hub.
- **GPU / accelerator vendors** — underlying capacity for Spaces and Endpoints.

KEY ACTIVITIES

- Operate Hub storage, CDN, discovery, and collaboration (git ML repos).
- Maintain Spaces (Gradio/Docker/ZeroGPU) and Jobs for demos and training.
- Run Inference Providers (serverless) and Inference Endpoints (dedicated).
- Ship open-source libraries and Hub UX that deepen daily habits.
- Sell and onboard Team / Enterprise (SSO, SCIM, security, support).
- Partner with CSPs so Hub models deploy in customer clouds.

KEY RESOURCES

- **Hub catalog** — 3M+ public models; 500k+ datasets (Spring 2026 / Aug 2026).
- **Community & brand** — default place to share and discover open ML.
- **Spaces directory** — hundreds of thousands of AI apps/demos.
- Open-source libraries and Hub APIs that lock in developer workflow.
- **Compute control plane** — Providers, Endpoints, Jobs, ZeroGPU.
- **Enterprise security stack** — SSO/SCIM, access control, data location.

VALUE PROPOSITIONS

- **Git-based Hub for models, datasets, Spaces** — the open ML catalog.
- **Spaces** — share interactive demos (CPU free; GPU PAYG).
- **Inference Providers** — serverless calls to 10k+ / 250k+ models (credits).
- **Inference Endpoints** — dedicated, autoscaling production deploy.
- **Team/Enterprise** — SSO, storage, billing controls, priority support.
- **CSP paths** — use Hub models inside AWS / Azure / Google Cloud.

CUSTOMER RELATIONSHIPS

- **Community PLG** — free Hub, Discussions, Posts, Articles, Spaces.
- **Self-serve PRO** — badge, credits, Dev Mode for serious individuals.
- **Team** — org seats, spending limits, Dataset Viewer on private data.
- **Enterprise** — custom onboarding, legal/compliance, dedicated support.
- Usage dashboards and credit auto-recharge for compute trust.
- AWS Marketplace / CSP billing for procurement-friendly orgs.

CHANNELS

- **Primary door** — huggingface.co Hub (models, datasets, Spaces, orgs).
- **Spaces as viral demos** — shareable ML apps.
- Docs, blogs, and social Posts / Articles on profiles.
- In-product subscribe for PRO and Team.
- Enterprise sales + annual contracts.
- AWS Marketplace; Azure / Google Cloud partner surfaces.

CUSTOMER SEGMENTS

- Individual developers and researchers (Free + PRO).
- Startups and product teams collaborating on private models (Team).
- Enterprises needing SSO/SCIM, compliance, priority support.
- MLOps / platform teams deploying Inference Endpoints.
- Open-model labs publishing weights to reach users.
- CSP customers discovering Hub models inside cloud consoles.

COST STRUCTURE

- Hub storage + egress/CDN for models and datasets (PB-scale).
- GPU/CPU capacity for Spaces, Jobs, Inference Endpoints, ZeroGPU.
- Inference Providers routing / partner compute economics.
- **Engineering** — Hub, libraries, security, billing systems.
- Enterprise sales, success, and dedicated support.
- Community programs, events, and open-source maintainership.

REVENUE STREAMS

- **PRO** — \$9/month individual subscription (FACT — /pro).
- **Team** — \$20/user/month; Enterprise from \$50/user/month (FACT — pricing).
- **Spaces upgraded hardware** — CPU Upgrade \$0.03/hr; T4 from \$0.40/hr.
- **Inference Endpoints** — dedicated deploy from ~\$0.03/hr CPU; GPU SKUs higher.
- **Inference Providers** — serverless usage; credits included then PAYG.
- **Private storage overage** — ~\$18/TB/mo base with volume discounts.
- **CSP / AWS Marketplace** — compute billed via cloud account.