1. **TPU Google cloud - TPU (Tensor Processing Unit)**

Single Cloud TPU device pricing

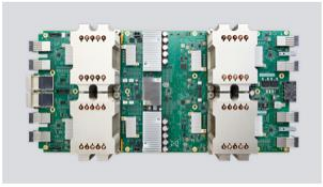
The following table shows the pricing per region for using a single Cloud TPU device.

US EUROPE ASIA PACIFIC

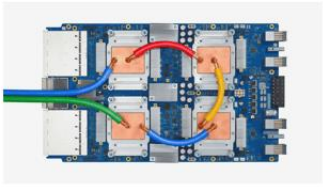
Version	On-demand	Preemptible
Cloud TPU v2	$4.95 / TPU hour	$1.485 / TPU hour
Cloud TPU v3	$8.80 / TPU hour	$2.64 / TPU hour

Zašto smo izabrale on demand a ne preemptible?
You can save money by using preemptible Cloud TPUs for **fault-tolerant** machine learning workloads, such as long training runs with checkpointing or batch prediction on large datasets.

Cloud TPU offering

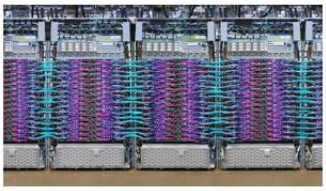


Cloud TPU v2
180 teraflops
64 GB High Bandwidth Memory (HBM)



Cloud TPU v3
420 teraflops
128 GB HBM

Cloud TPU v2 Pod
11.5 petaflops
4 TB HBM
2-D toroidal mesh network



Cloud TPU v3 Pod
100+ petaflops
32 TB HBM
2-D toroidal mesh network

Znači 1 TPU v3 je kao naše 4 grafičke karte. Odnosno za 4 grafičke nam treba 1 tpu – broj teraflops se podudara sa specifikacijom grafičkih koje smo tražili.

2. Услуге - Equinix data center

https://www.dscga.com/colocation-pricing-the-definitive-guide-on-what-to-expect-2019-report/ - cene za data centar
Cene su date za 1U jedinicu. Naš jedan server zauzima 2U. Cene se kreću od 50 do 300e za 1U za 1 mesec. Nama je potrebno na 17 meseci, znači 17meseci*2U*280e = 9500e po serveru.

Tier 1:

- Non-redundant systems (no backups for power, network, etc.)
- 99.671% Uptime
- 28.8 hours of downtime per year

Tier 2:

- Partial redundancy for power and cooling
- 99.749% Uptime
- 22 hours of downtime per year

Tier 3:

- N+1 fault tolerant providing at least 72 hour power outage protection
- 99.982% Uptime
- Maximum of 1.6 hours of downtime per year

Tier 4:

- 2N+1 fully redundant infrastructure with 96 hour power outage protection
- 99.995% Uptime per year
- Maximum 26.3 minutes of annual downtime

Data centar koji smo mi izabrale je **Equinix u Minhenu - tier 3**, ima n+1 redundant infrastructure