# Leveraging Cloud GPUs

## Major Cloud Providers

**AWS**
EC2 P4/P5

**GCP**
A2/A3

**Azure**
NC/ND series

## On-Demand

$$$$

Pay per hour - Expensive but flexible

e.g., $32/hr for A100

## Spot Instances

70-90% off

Much cheaper, can be interrupted

Good for experiments

## Managed Services

- SageMaker (AWS)
- Vertex AI (GCP)
- Azure ML

## GPU Marketplaces

- Lambda Labs
- RunPod
- Vast.ai

**Cost Optimization**
Small GPUs for debug, scale up for training

**Free Credits**
Research/education credits from providers

**Always Shut Down**
Stop instances when not in use, monitor spending