Cloud Computing
Sky Computing

Minchen Yu
SDS@CUHK-SZ
Fall 2024

香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

SCHOOL OF
DATA SCIENCE
數據科學學院

Utility Computing

Computing may someday be organized as a public utility just as the telephone system is a public utility … Each subscriber needs to pay only for the capacity he actually uses, but he has access to all programming languages characteristic of a very large system … The computer utility could become the basis of a new and important industry

— John McCarthy, 1961

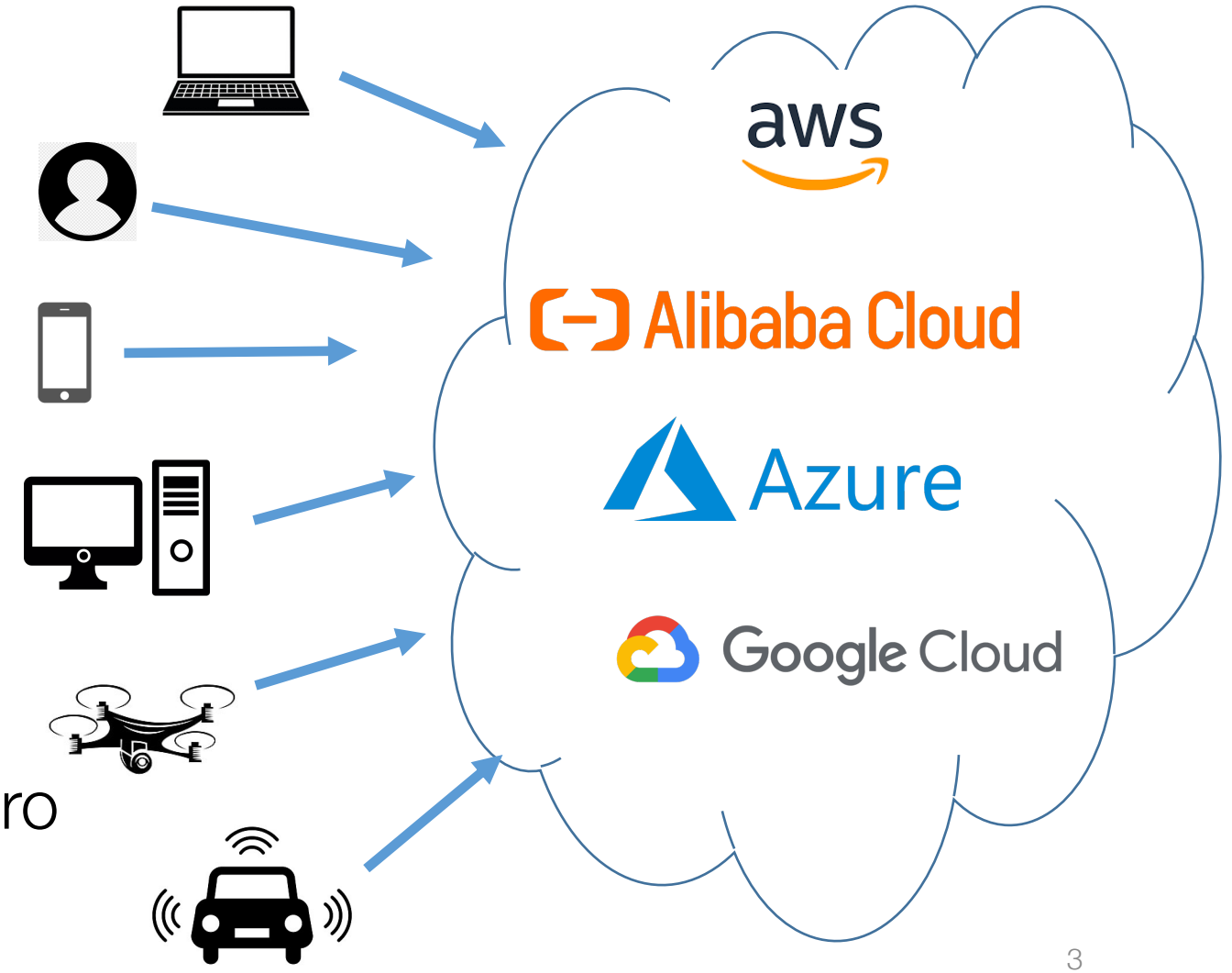
Cloud Computing: One Step Forward

Allows everyone access to

- Unlimited computing resources
- Unlimited storage resources
- All delivered over internet

No upfront cost

- Pay only for what you use



All started in 2006 since the intro of AWS cloud

The Cloud Ecosystem Thrives

A fundamental infrastructure that runs anything at any scale

Today, almost all apps and digital services are built on top of the cloud!

Vendor Lock-In

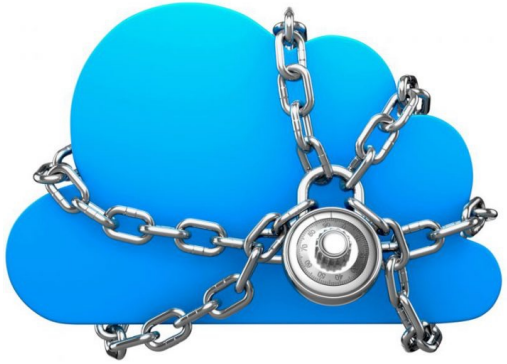
Each cloud provides different services to differentiate itself

- Proprietary services and programming interfaces
  - E.g., AWS has 175+ services, many being proprietary
- Proprietary hardware
  - E.g., TPUs exclusive to Google Cloud, Inferentia exclusive to AWS

Data gravity pricing

- Free to move data into the cloud but expensive to move data out

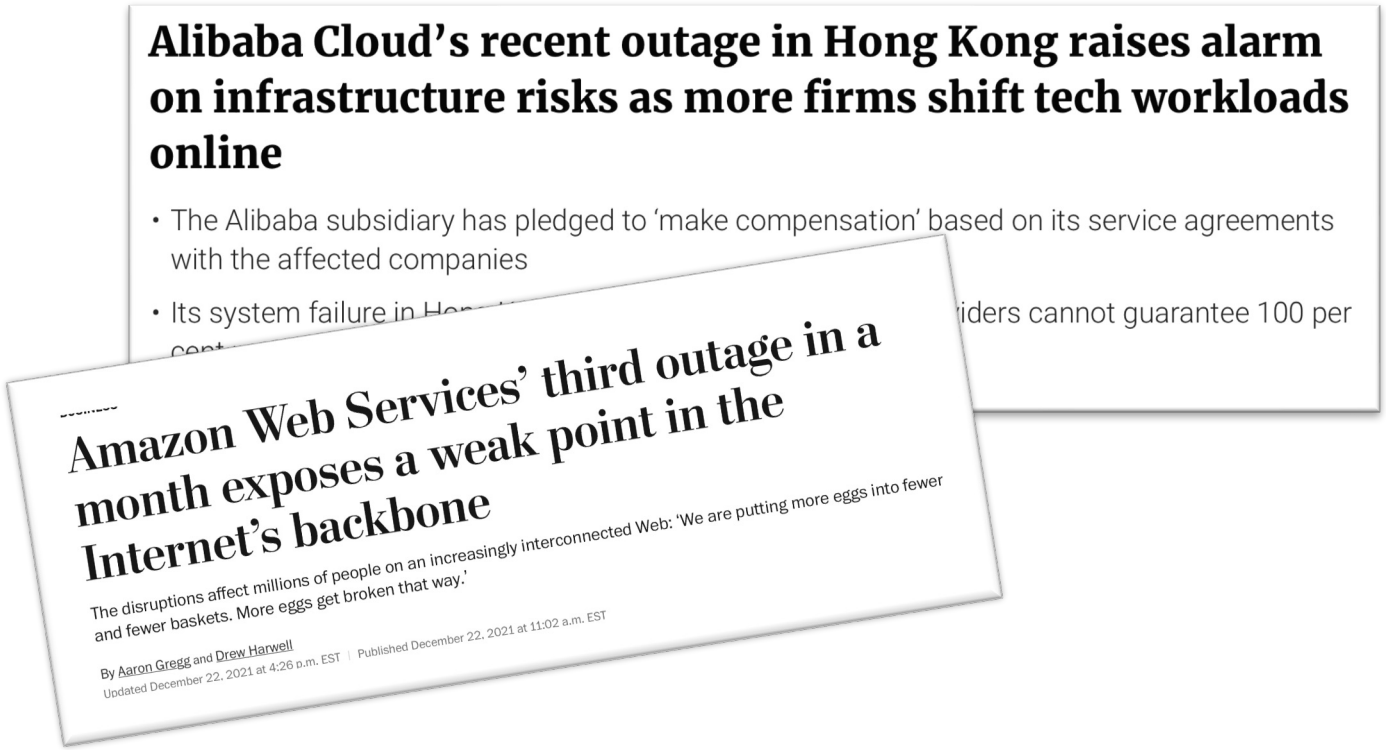
Cloud users often found themselves locked into the current provider!



Vendor Lock-In: a Thorny Pain Point

Reliability and security

- Adds the risks of service outage and/or data leakage



Alibaba Cloud's recent outage in Hong Kong raises alarm on infrastructure risks as more firms shift tech workloads online

- The Alibaba subsidiary has pledged to 'make compensation' based on its service agreements with the affected companies
- Its system failure in Hon… …iders cannot guarantee 100 per cent…

Amazon Web Services' third outage in a month exposes a weak point in the Internet's backbone

The disruptions affect millions of people on an increasingly interconnected Web: 'We are putting more eggs into fewer and fewer baskets. More eggs get broken that way.'

By Aaron Gregg and Drew Harwell
Updated December 22, 2021 at 4:26 p.m. EST | Published December 22, 2021 at 11:02 a.m. EST

Vendor Lock-In: a Thorny Pain Point

Reliability and security

- Adds the risks of service outage and/or data leakage

Data sovereignty regulations

- Data must be stored and processed in local datacenters operated by the country nationals, which the current provider does not have
  - E.g., Google cloud has no datacenters in China

Short of commercial leverage

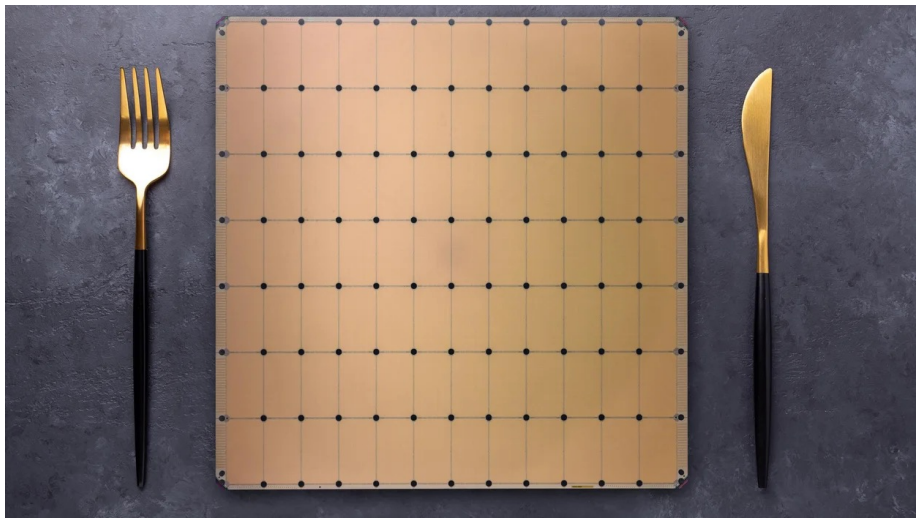
Vendor Lock-In Harms Innovation

Hard to take advantage of innovations across clouds

- E.g., AWS users cannot use TPUs in Google cloud

Hard to take advantage of innovation outside clouds

- E.g., cannot use the latest Cerebras AI chips unless a cloud provider adopts them

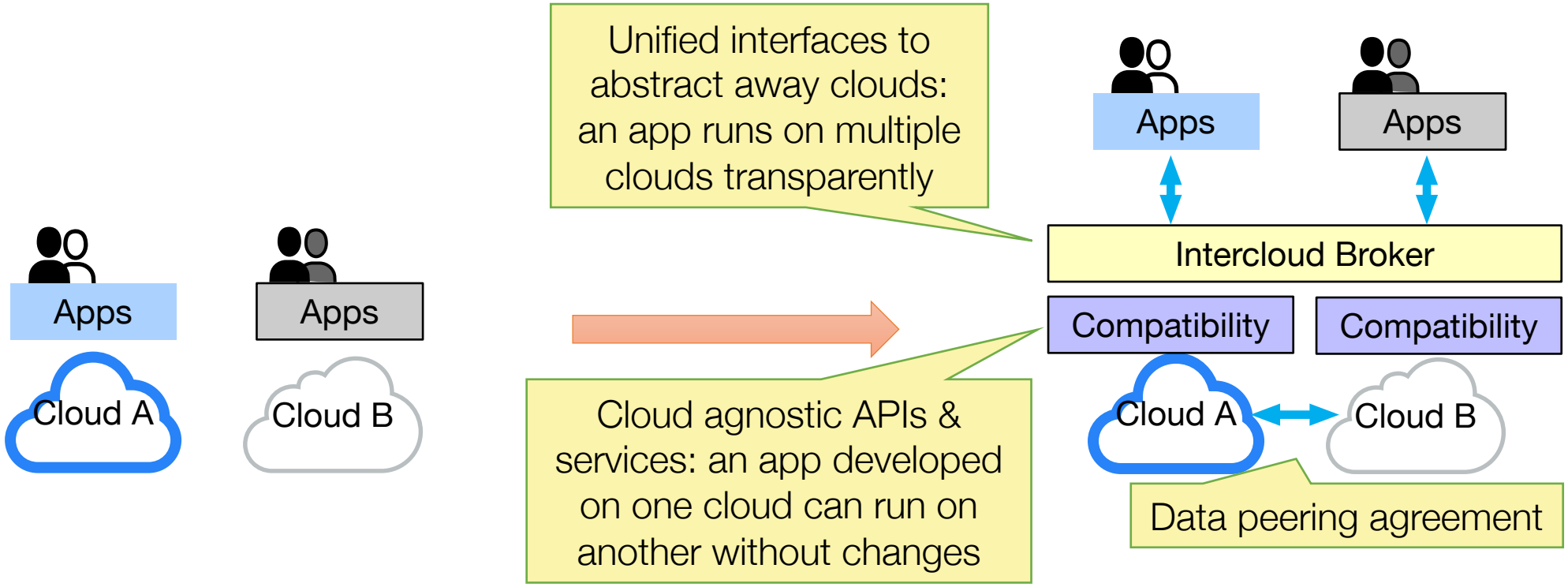


cerebras

From Cloud Computing to Sky Computing

Sky: Cloud of Clouds

- Use an intercloud broker to create a two-sided market between users and cloud providers, thus addressing the vendor lock-in problem



Sky's Benefits

Better reliability & security

- Resilient to the outage of one cloud
- Distributing trust across multiple clouds eliminate the single point of attack

Compliance to the data sovereignty regulation

Better commercial leverage

Enabling access to cloud-exclusive hardware

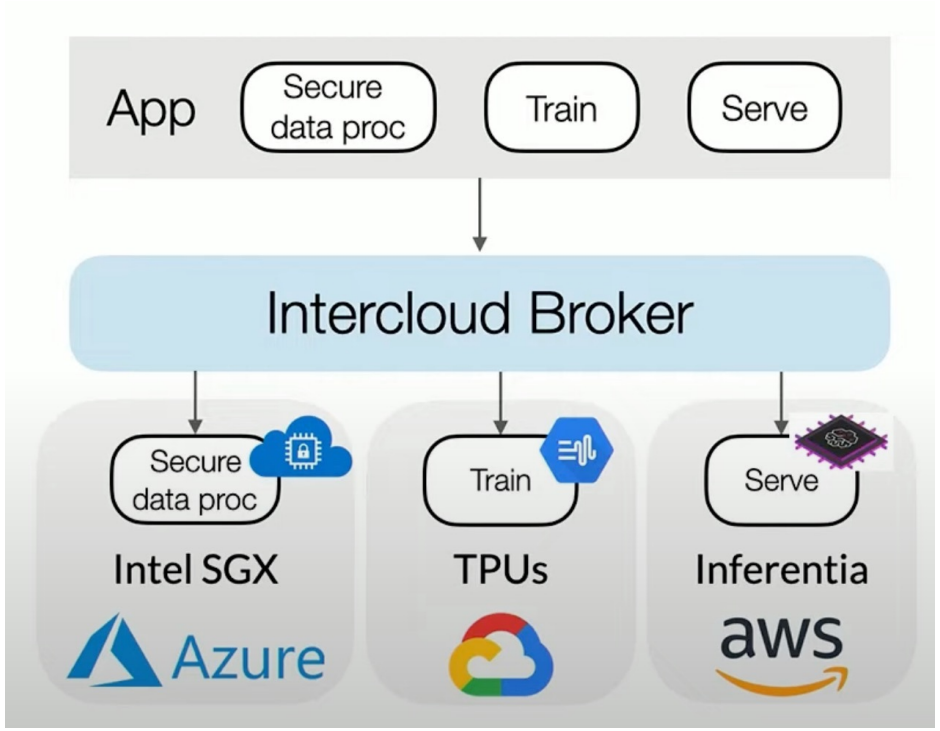
- E.g., AWS users can use TPUs in Google cloud

Better performance & cost

- Using the best-of-breed services across clouds

Case study: SkyPilot

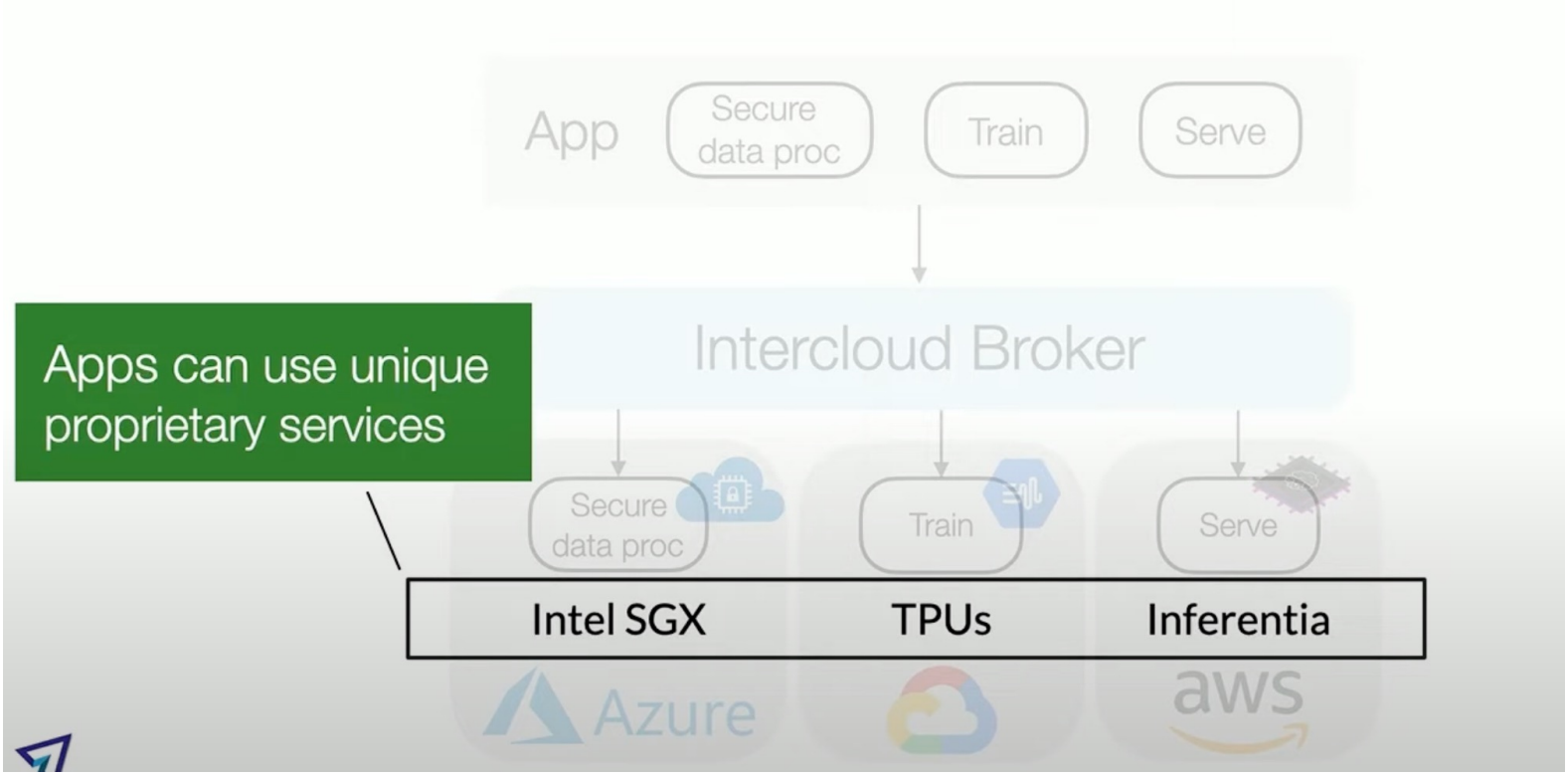
Cross-cloud deployment and execution and ML pipeline



Credit: Zongheng Yang, et al “SkyPilot: An Intercloud Broker for Sky Computing”, NSDI 2023.

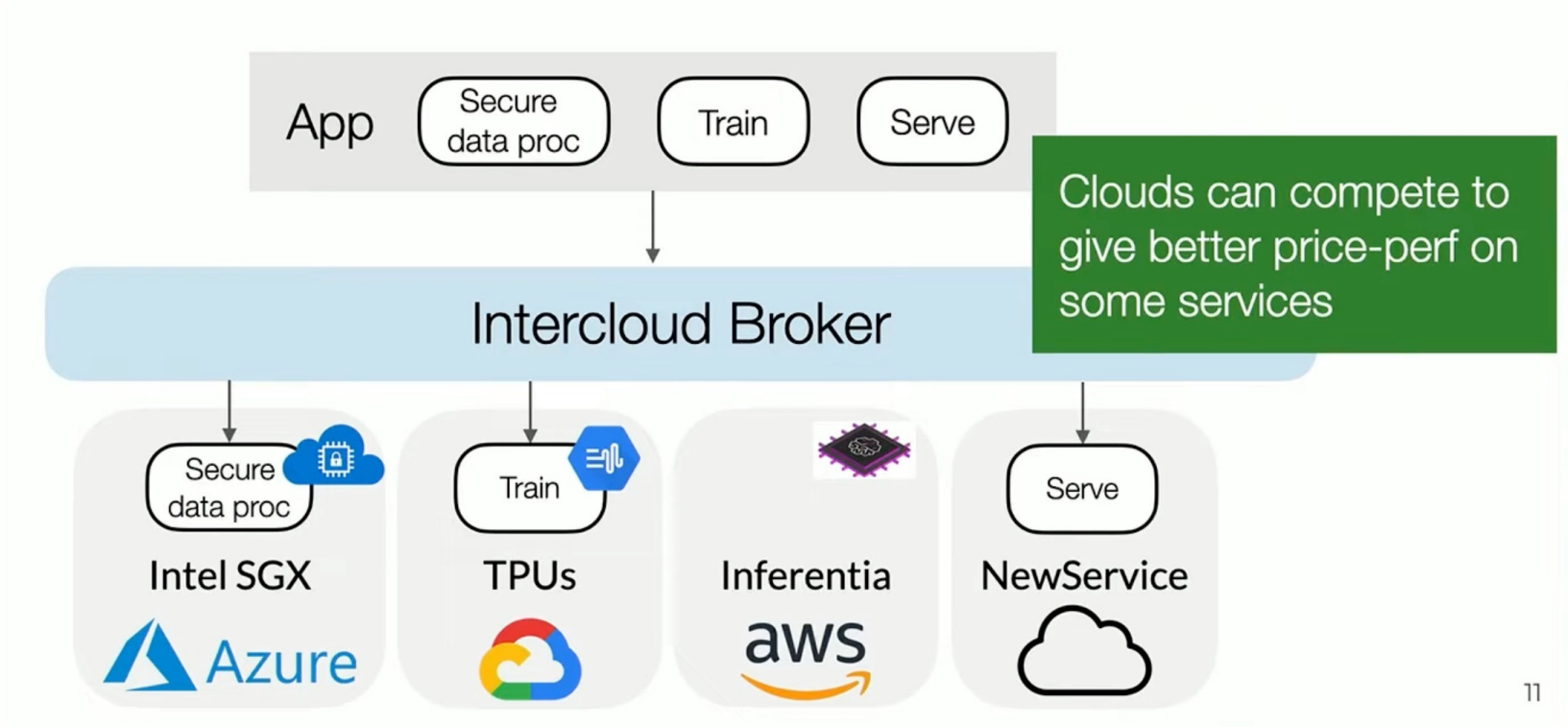
Case study: SkyPilot

Cross-cloud deployment and execution and ML pipeline



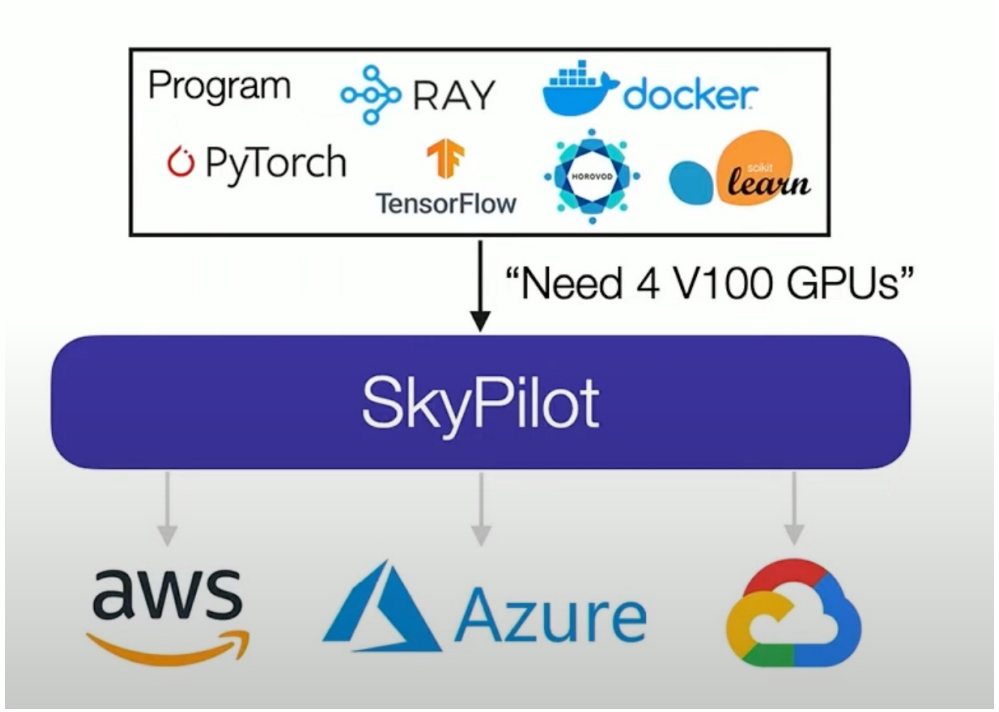
Case study: SkyPilot

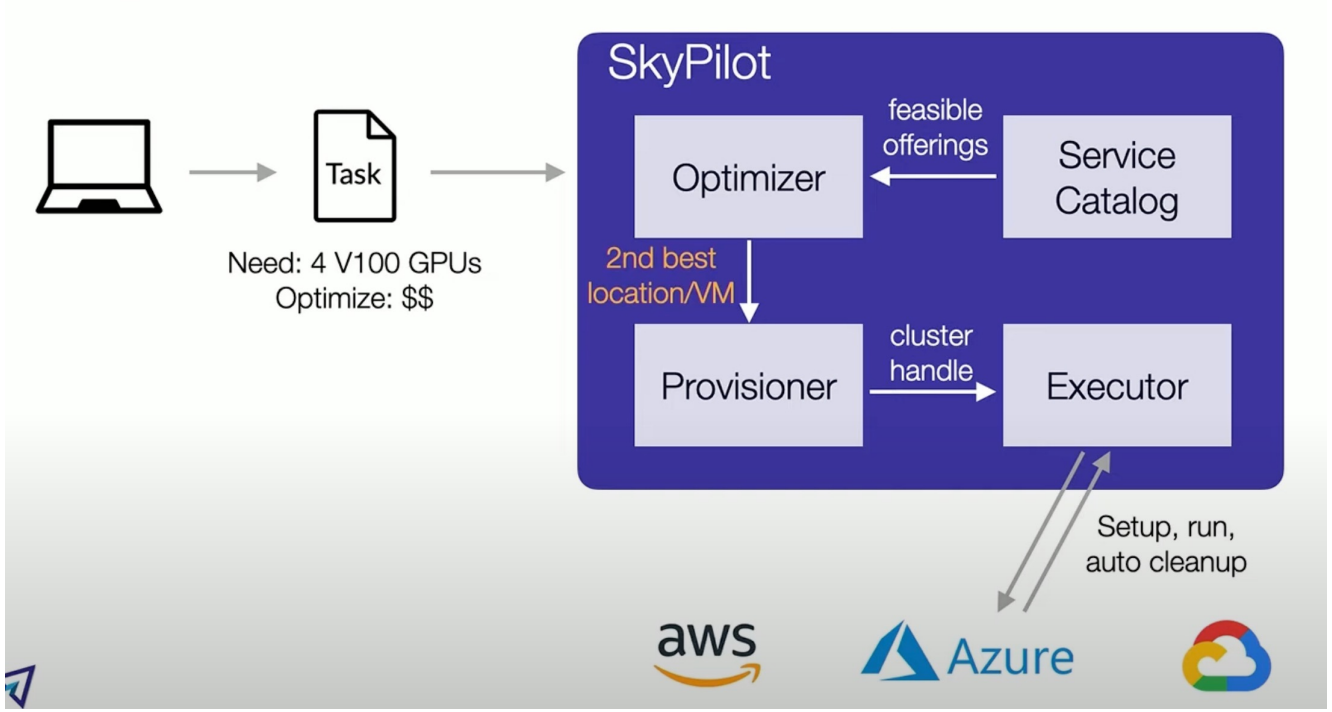
Open to new clouds and services



Case study: SkyPilot

Configure cloud services with user-defined requirements

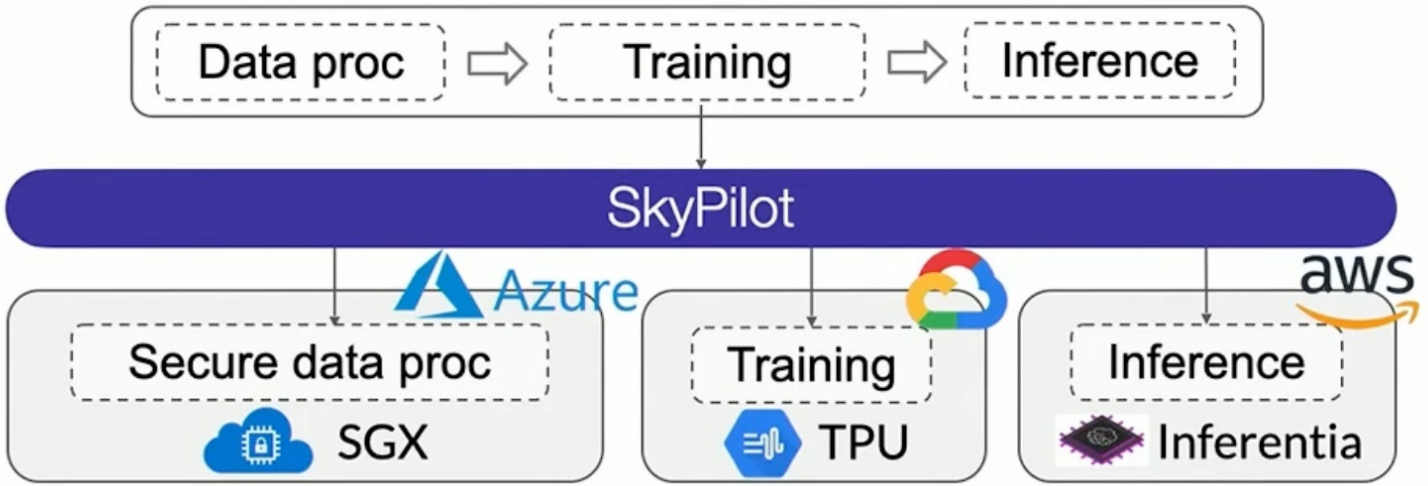




Cross-cloud ML pipeline

BERT on user review data

- Requirement: use SGX to for data processing



		proc	train	infer	egress	**Total**
Time	**Azure**	0.6	13.3	1.5	–	15.4
(hr.)	**Broker**	0.6	3.8 -71%	1.4 -7%	0.03	5.8 -62%
Cost	**Azure**	0.8	163	1.2	–	165
($)	**Broker**	0.8	32 -80%	0.5 -58%	0.1	33.4 -80%

Reduce both latency and cost

Grand Challenges of Sky Computing

What is the functionality of the intercloud broker and how can it be built?

How to incentivize data peering between two clouds?

What are the “killer apps” of Sky computing?