



Scalable Earthquake Catalog Generation: Harnessing cloud computing with Kubeflow on Azure

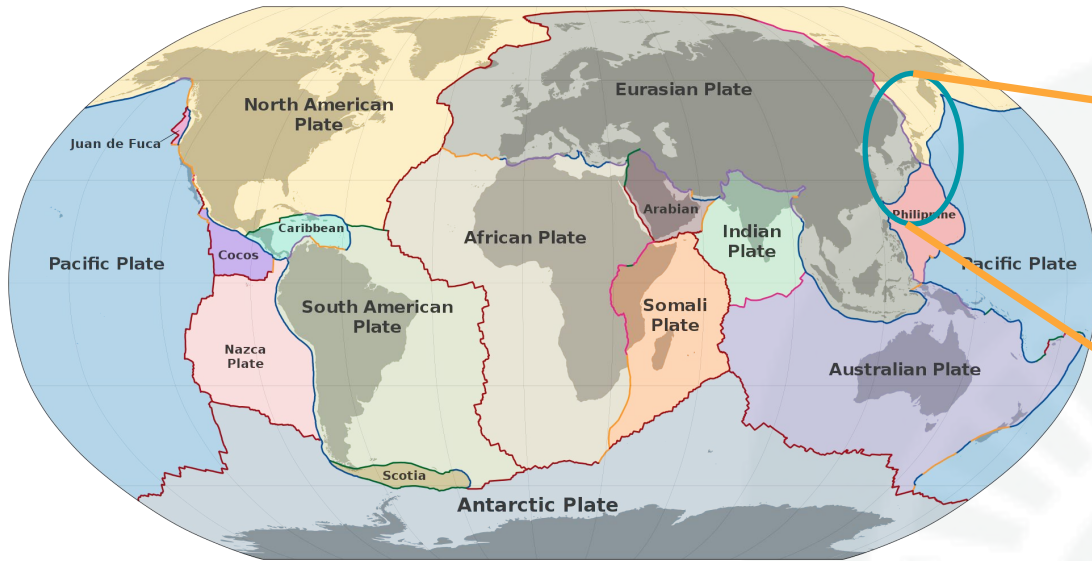
Ziyi Xi

Department of Computational Mathematics, Science and Engineering

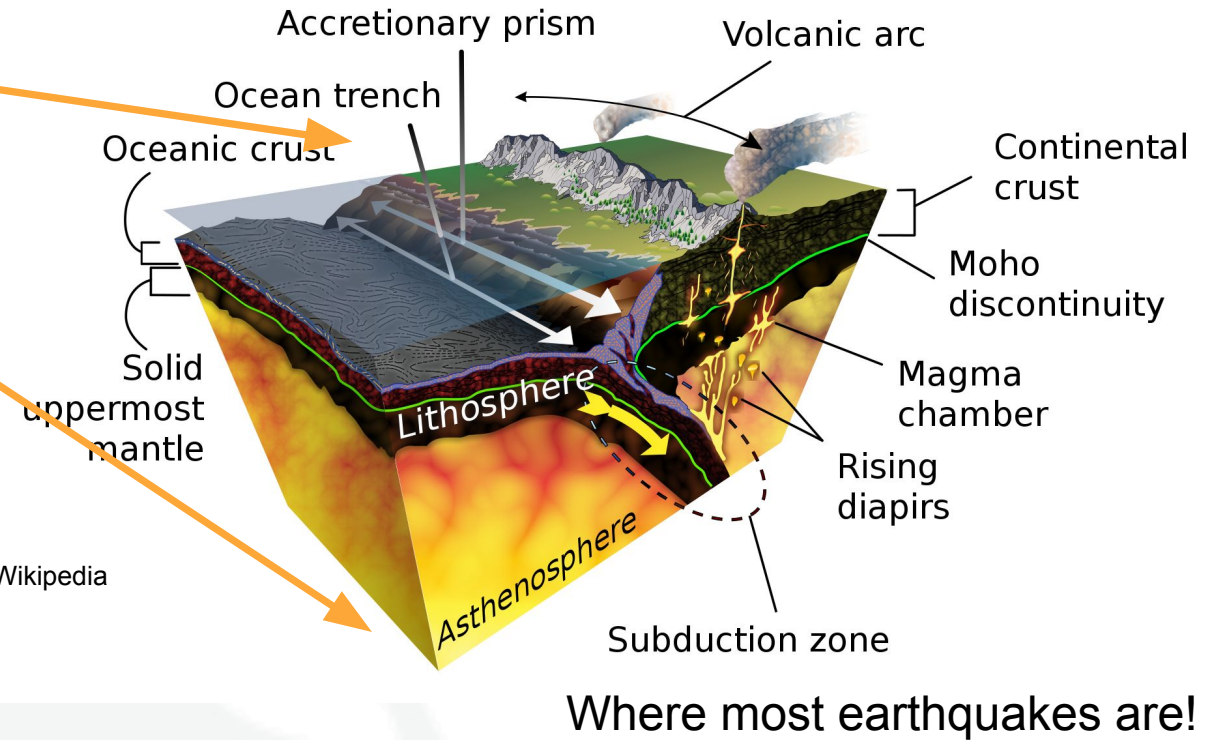
Department of Earth and Environmental Sciences

Michigan State University

The Importance of Studying Seismicity



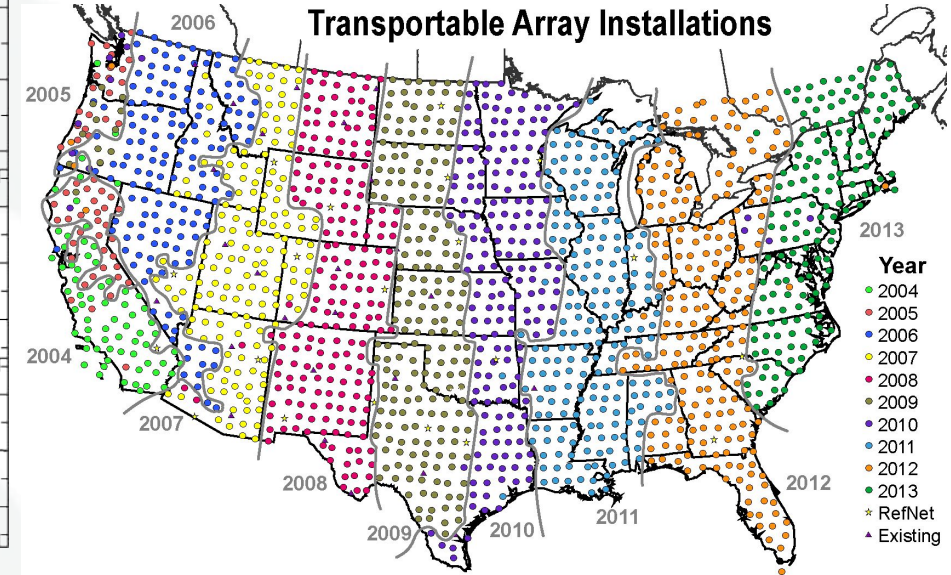
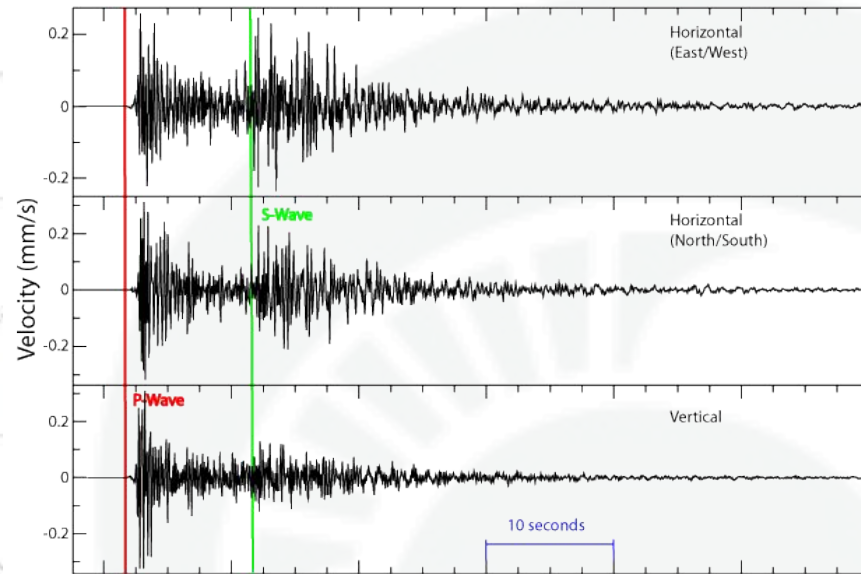
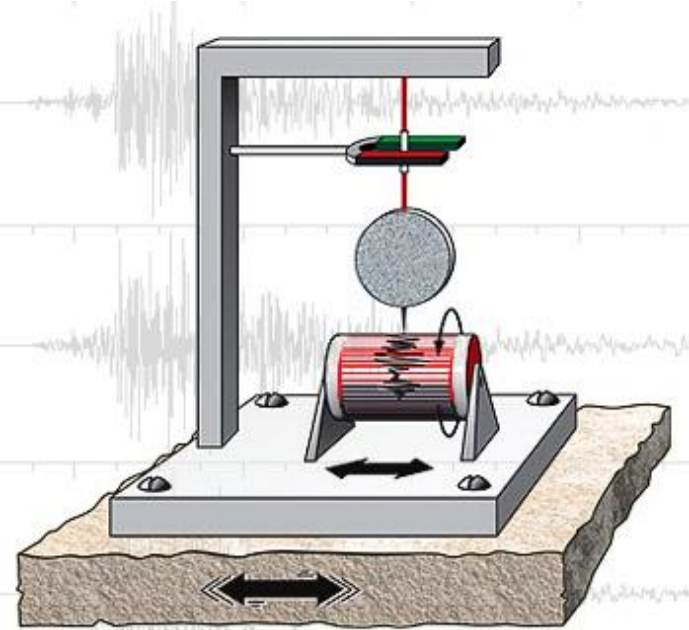
Cited from Wikipedia



- Plate Collision => Subduction
- Subduction => Earthquakes

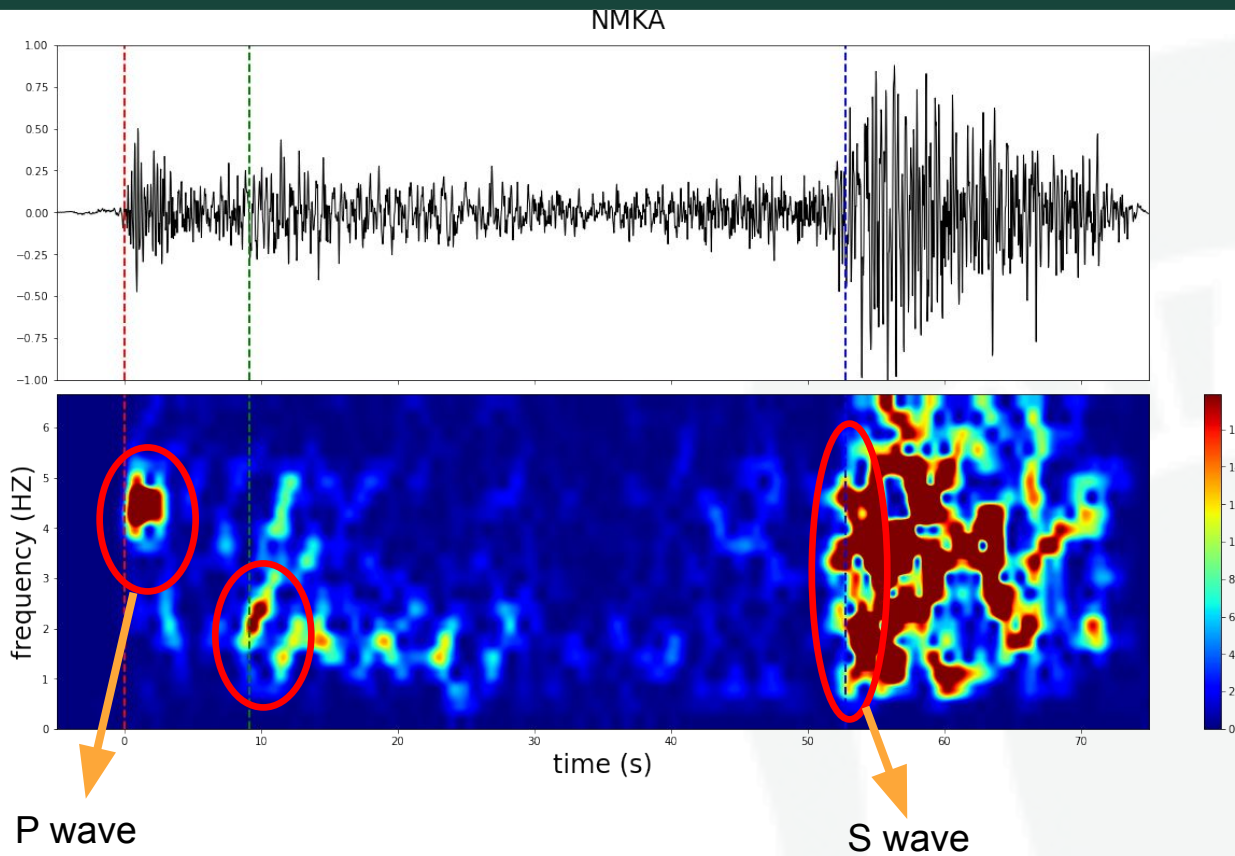
- So, if we know earthquake distribution inside Earth:
 - Where and how subduction happens
 - Better understand plate collision process

Seismic Waves: A Wealth of Data



- Thousands of earthquake monitoring stations are distributed worldwide
- Detect earthquake signals is the first step for further analysis

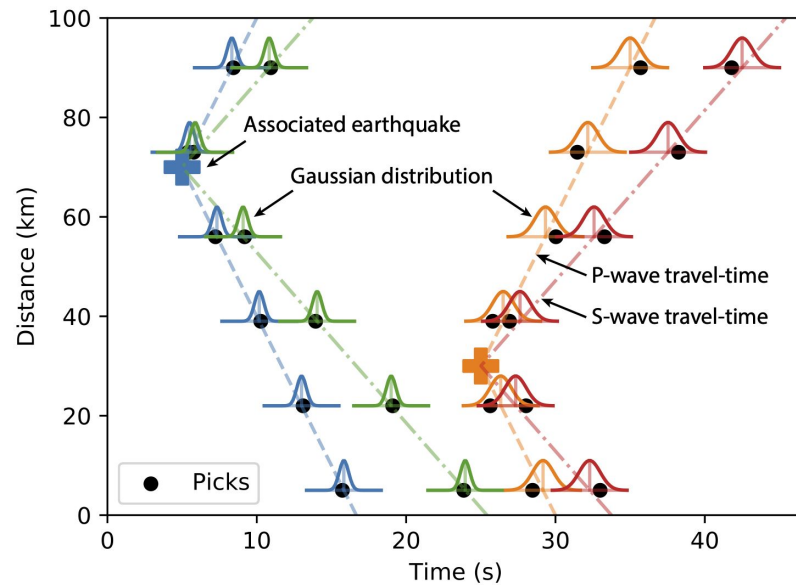
Analyzing Spectrograms for Signals



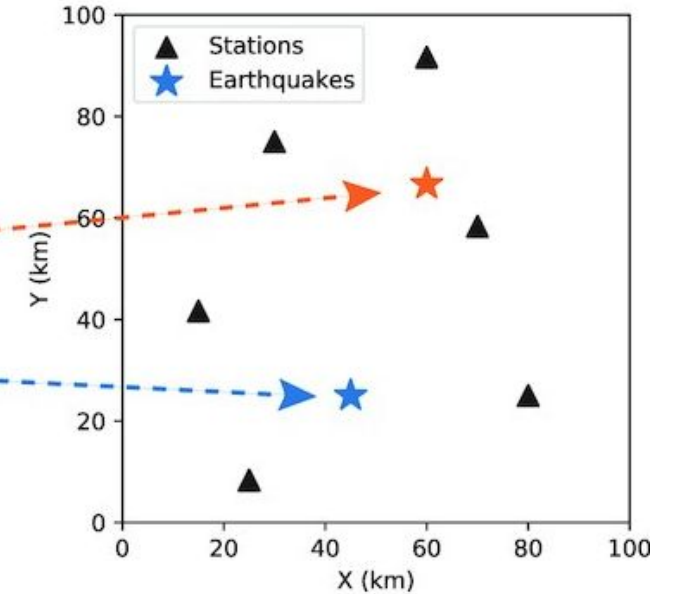
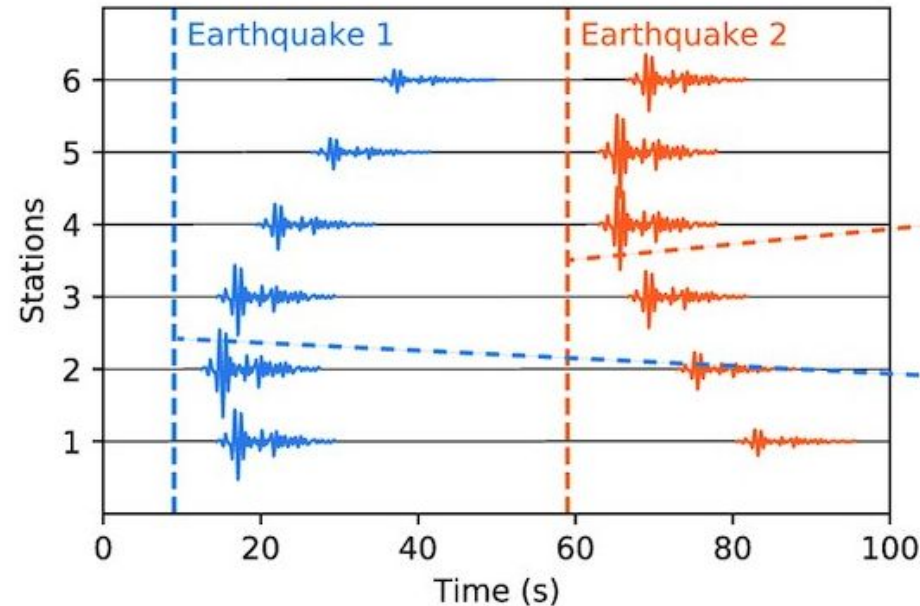
Example of image segmentation: Segment Everything – Meta AI

- Spectrogram domain image segmentation => distinguish seismic signals with noise

Inversion Problem: Earthquake Detection

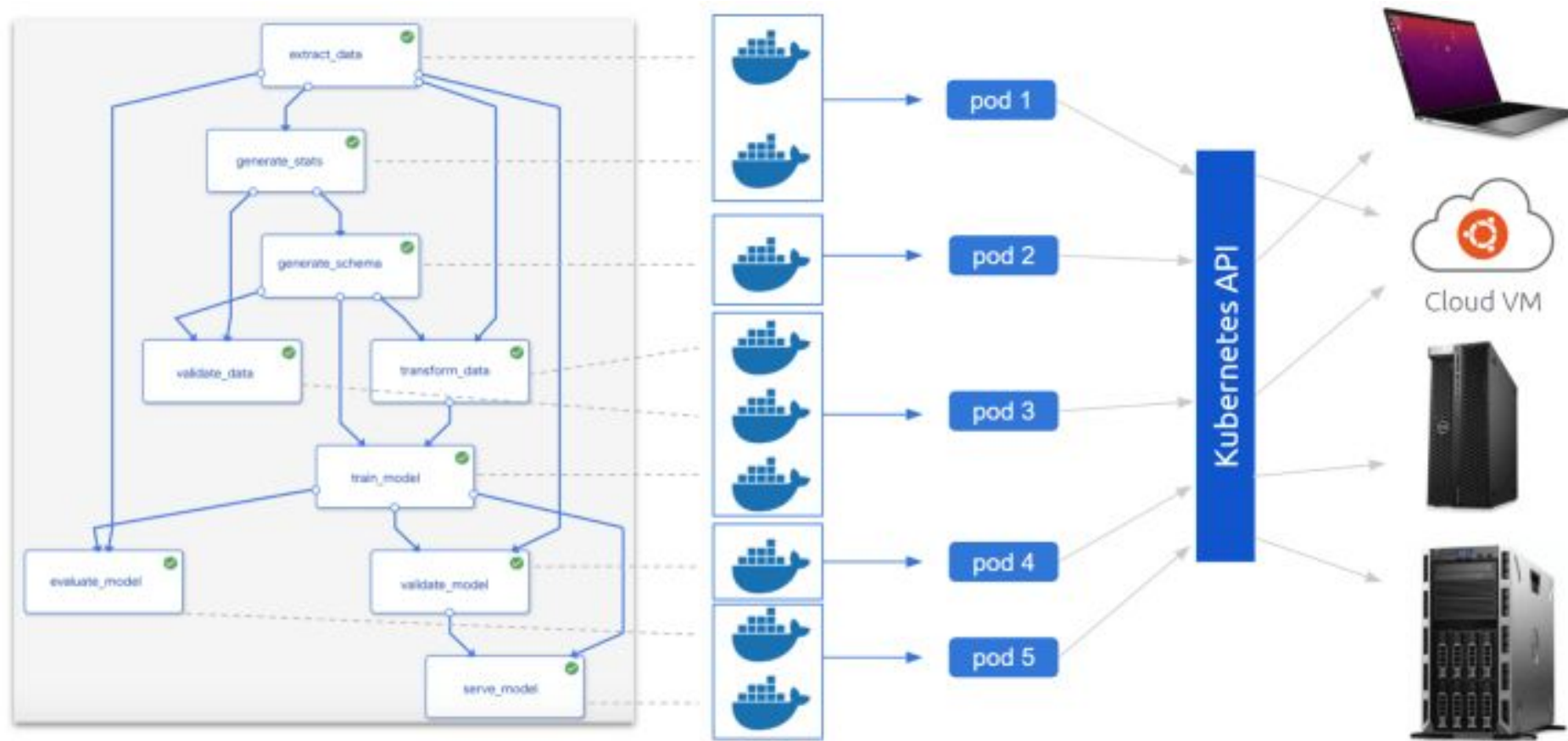


Zhu et al. 2022, JGR Solid Earth



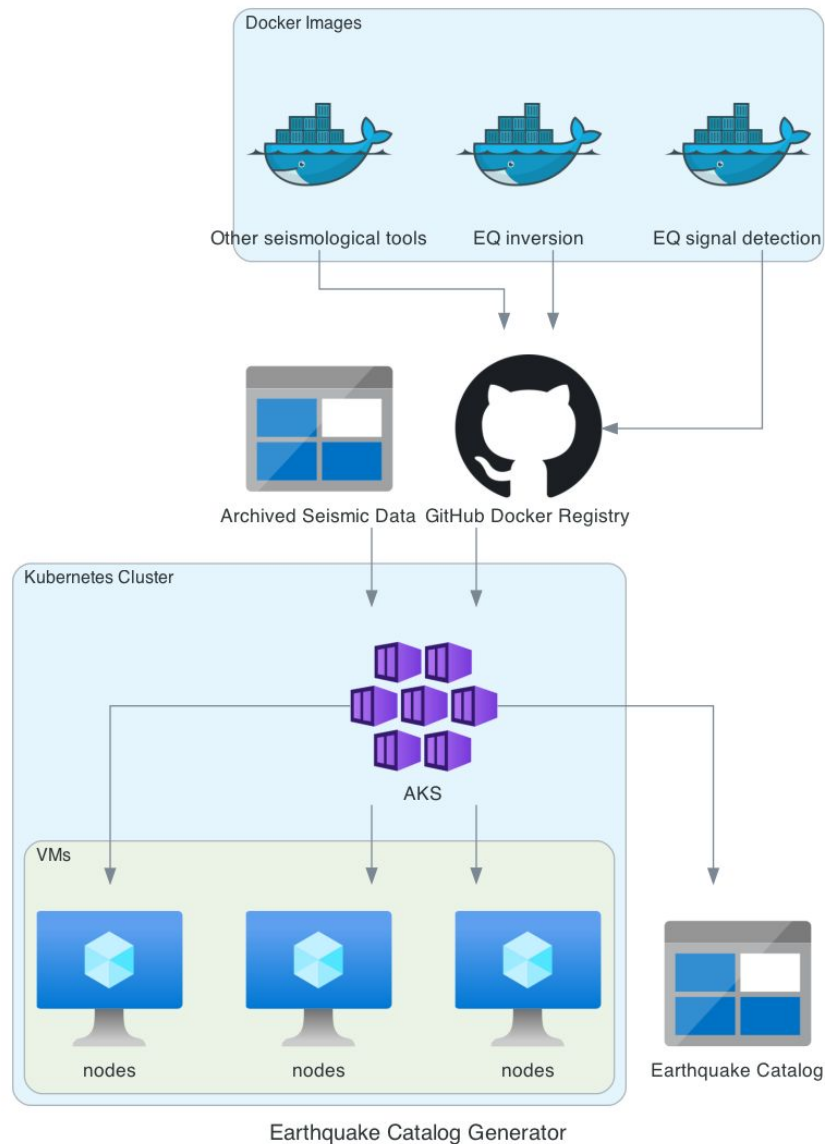
- Earthquake locations/time $\Rightarrow$ forward problem $\Rightarrow$ Earthquake signals locations/time
- Earthquake signals locations/time $\Rightarrow$ inversion problem $\Rightarrow$ Earthquake locations/time

Bring it all together using Kubeflow!



- Kubeflow builds execution pipeline above Kubernetes
- The pipeline can run end to end without controlling any hardware!

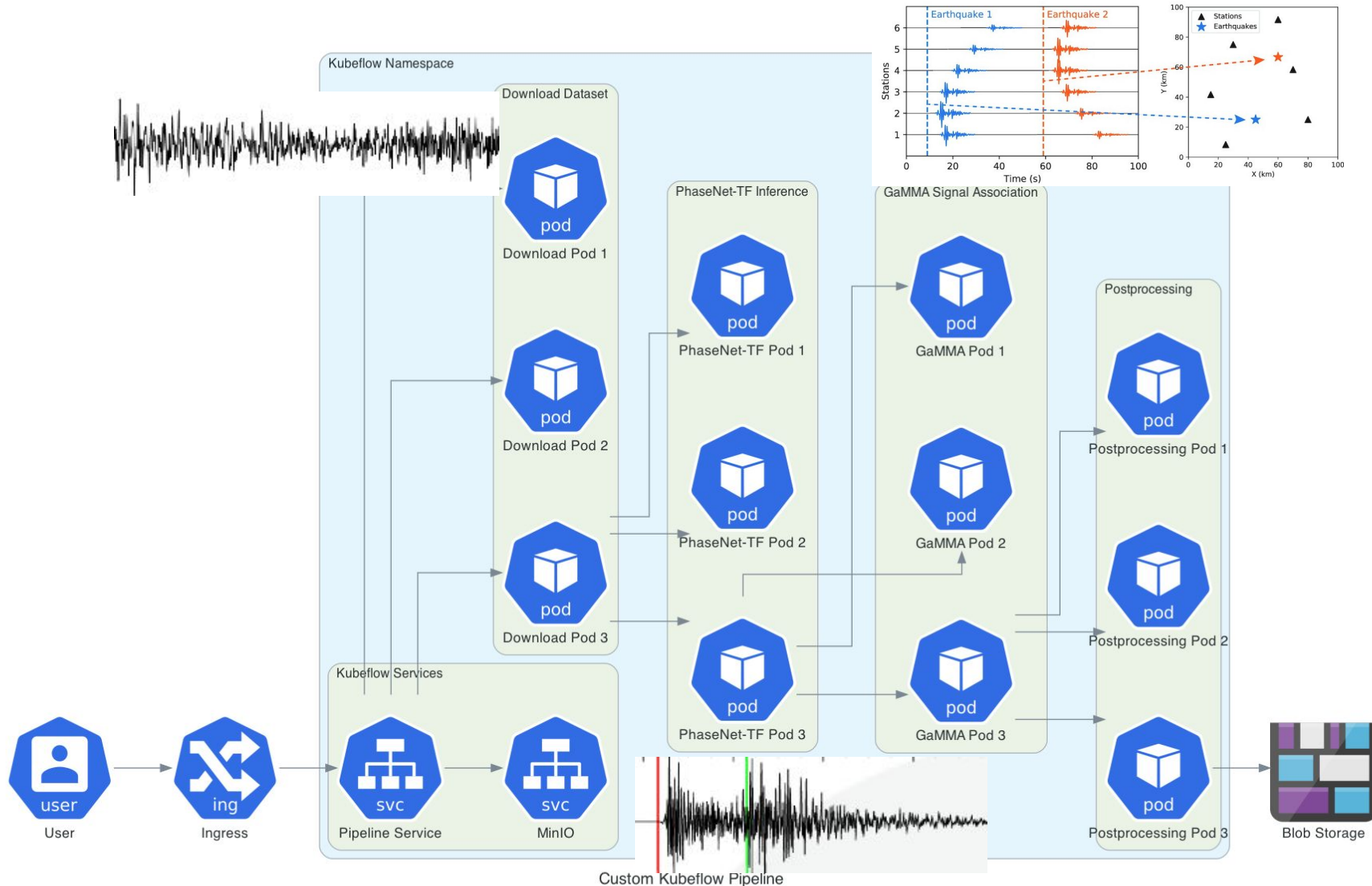
Cloud Computing for Seismic Analysis



Azure/GitHub Services in this project

- GitHub docker registry:
 - PhaseNet-TF (seismic signal detection)
 - GaMMA (signal association and earthquake time/location inverter)
 - Other seismological tools to refine the result
- Blob storage:
 - Store the seismic wave dataset
 - Store the final earthquake catalog
- Azure Kubernetes Service:
 - Perform computing tasks

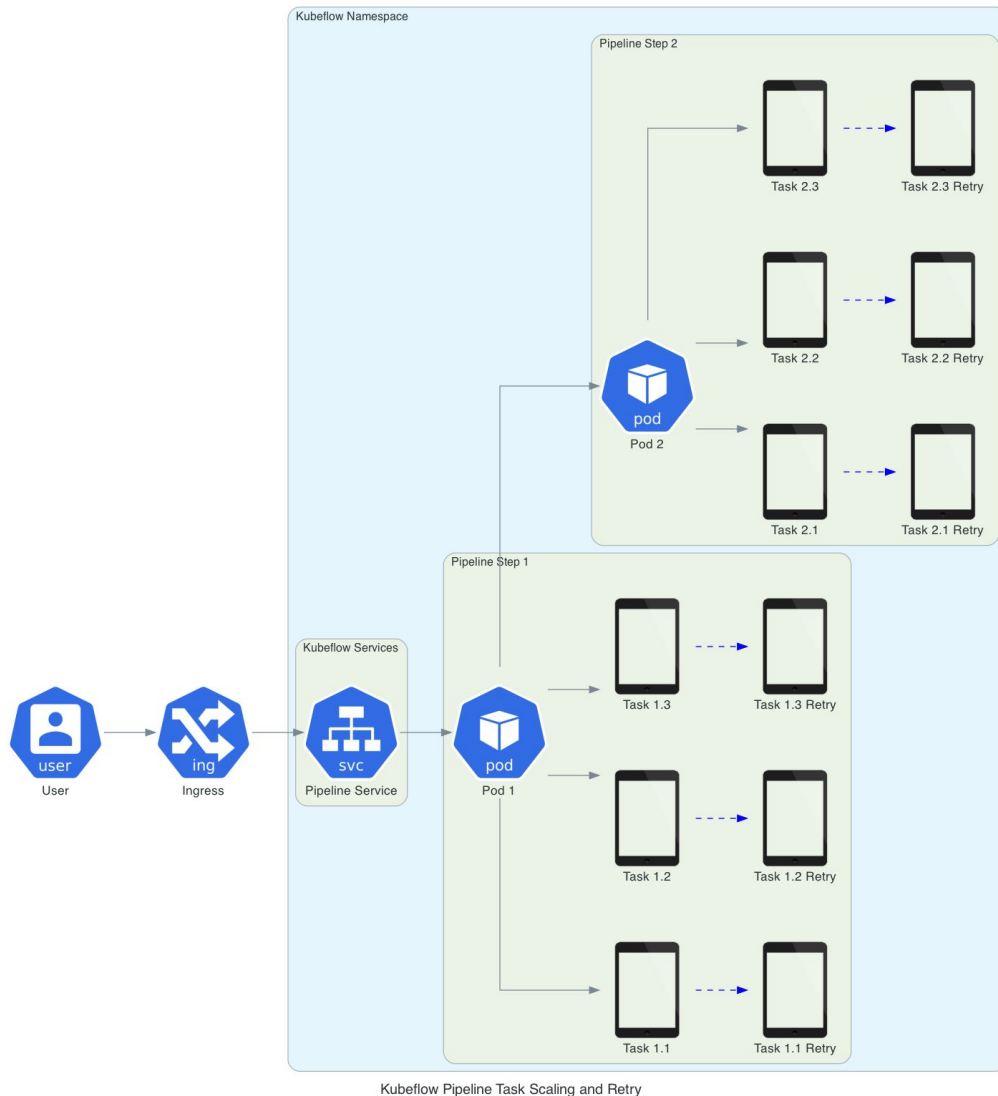
Kubeflow and K8s: workflow pipeline



Kubeflow is used to construct the pipeline

- Can automatically schedule tasks on pods.
- Pods can be regarded as virtual nodes on top of physical nodes

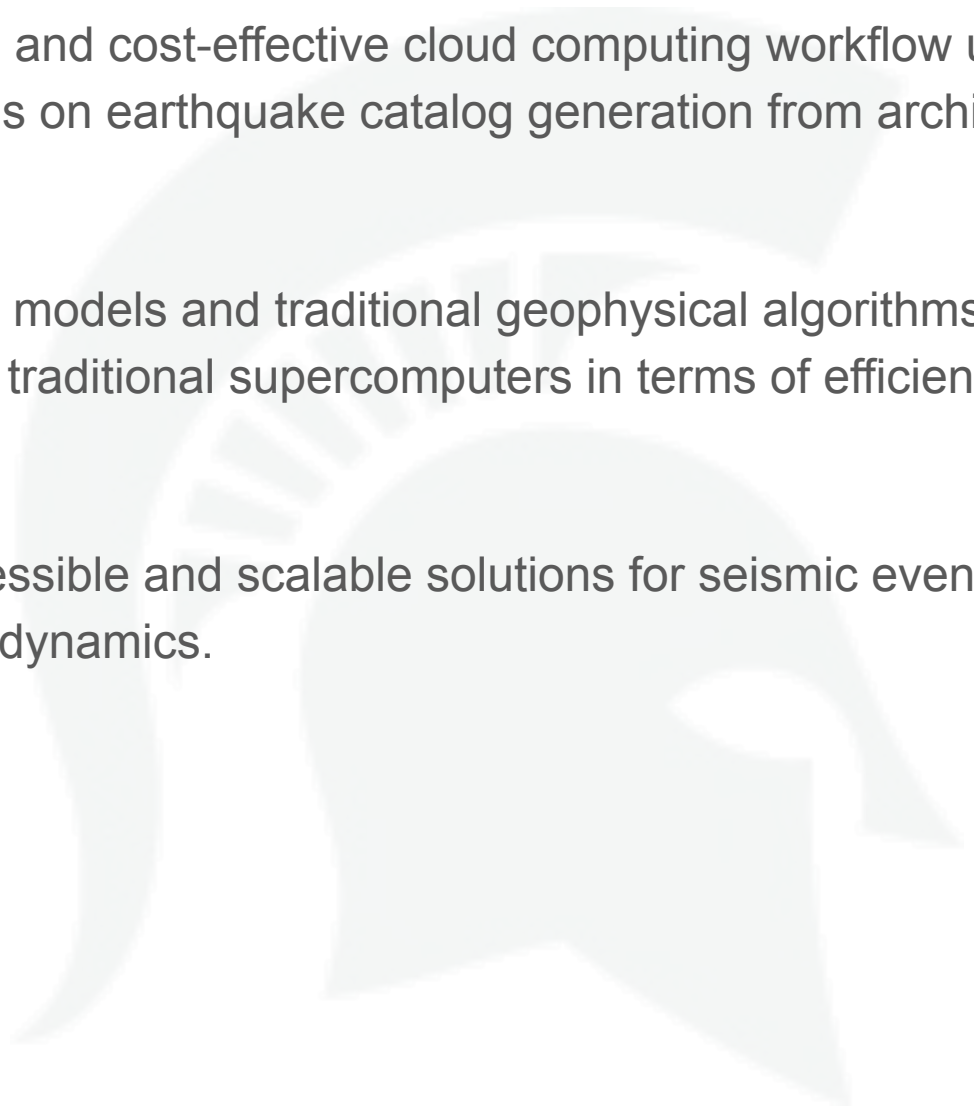
Advantages of Kubeflow and K8s



With the nature of Kubernetes, Kubeflow can:

- ❑ Auto-scaling: Different stages need different computing power.
- ❑ Fault-tolerance: Retry when one step fails.
- ❑ Set up once, run in the future:
 - ❑ We only need to update docker images in the future if we have updated algorithms to test
- ❑ No need to bind Azure:
 - ❑ An extra abstraction layer with Kubernetes, no restriction for the underlying platforms

Summary & Conclusions



- We developed a scalable and cost-effective cloud computing workflow using Kubeflow on Azure Kubernetes Service, focus on earthquake catalog generation from archived seismograms.
- Combining deep learning models and traditional geophysical algorithms, our approach has unique advantages compared to traditional supercomputers in terms of efficiency and scalability.
- This project delivers accessible and scalable solutions for seismic event analysis, contributing to a better understanding of Earth's dynamics.