IGARSS, July 16, 2023

End-to-End Machine Learning with Supercomputing and in the Cloud

Introduction and Motivations

Manil Maskey, Gabriele Cavallaro, Iksha Gurung, Muthukumaran Ramasubramanian, Rocco Sedona

HDCRS

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

UNIVERSITY OF ICELAND

GRSS

ESI

IMPACT Interagency Implementation and Advanced Concepts Team

Motivations

Build capacity around latest machine learning approaches for remote sensing data

Utilize various computing resources to optimize machine learning

Promote open science via collaboration

Provide a venue for sharing experiences and lessons learned

Promote collaboration amongst machine learning experts, domain experts, and software developers

What we hope to do today

- High level introduction to geospatial foundation model
- Fine tune foundation model in **HPC** for various downstream applications
- Inference on trained model in **cloud** using new data from archive

Expected outcomes

Participants are expected to:

- Learn the fundamentals of foundation model
- Understand machine learning life cycle
- Fine-tune foundation models for downstream applications
- Inference on trained models

Everyone is expected to:

- Exchange ideas
- Foster collaboration

NASA EARTH FLEET

OPERATING & FUTURE THROUGH 2023

SWOT (CNES)
LANDSAT-9 (USGS)
SENTINEL-6 Michael Freilich/B (ESA)
TROPICS (6)
GEOCARB
NISAR (ISRO)
MAIA
TSIS-2
TEMPO
PREFIRE (2)
PACE (NSO)
GLIMR
ICESAT-2
GRACE-FO (2) (DLR)
CYGNSS (8)
NISTAR, EPIC (DSCOVR/NOAA)
CLOUDSAT (CSA)
TERRA (JAXA, CSA)
AQUA (JAXA, AEB)
AURA (NSO, FMI, UKSA)
CALIPSO (CNES)
GPM (JAXA)
LANDSAT 7 (USGS)
LANDSAT 8 (USGS)
OCO-2
SMAP
SUOMI NPP (NOAA) (JAXA)

ISS INSTRUMENTS

EMIT
CLARREO-PF
GEDI
SAGE III
OCO-3
TSIS-1
ECOSTRESS
LIS

JPSS-2, 3 & 4 INSTRUMENTS

OMPS-Limb
LIBERA

03.24.20

INVEST/CUBESATS

RainCube
CSIM-FD
CubeRRT
TEMPEST-D
CIRiS
HARP
CTIM
HyTI
SNoOPI
NACHOS

(PRE) FORMULATION
IMPLEMENTATON
PRIMARY OPS
EXTENDED OPS

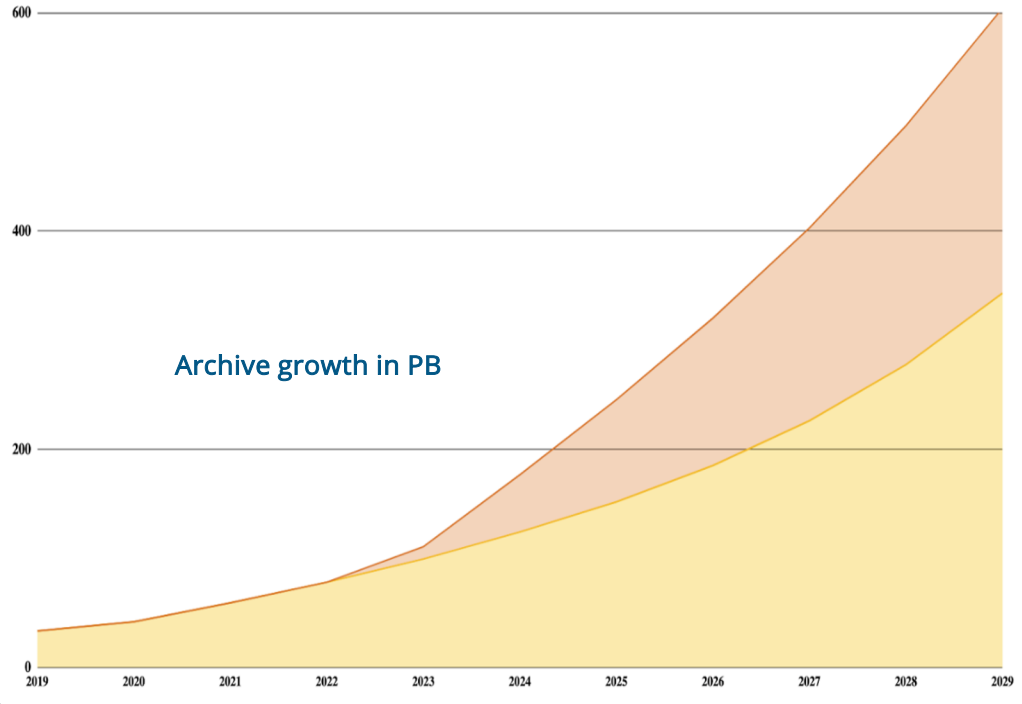
By the number

62 PB
Archive

2 Billion
Data products

3.56 Billion
Files in archive

1.7 Million
Distinct data users



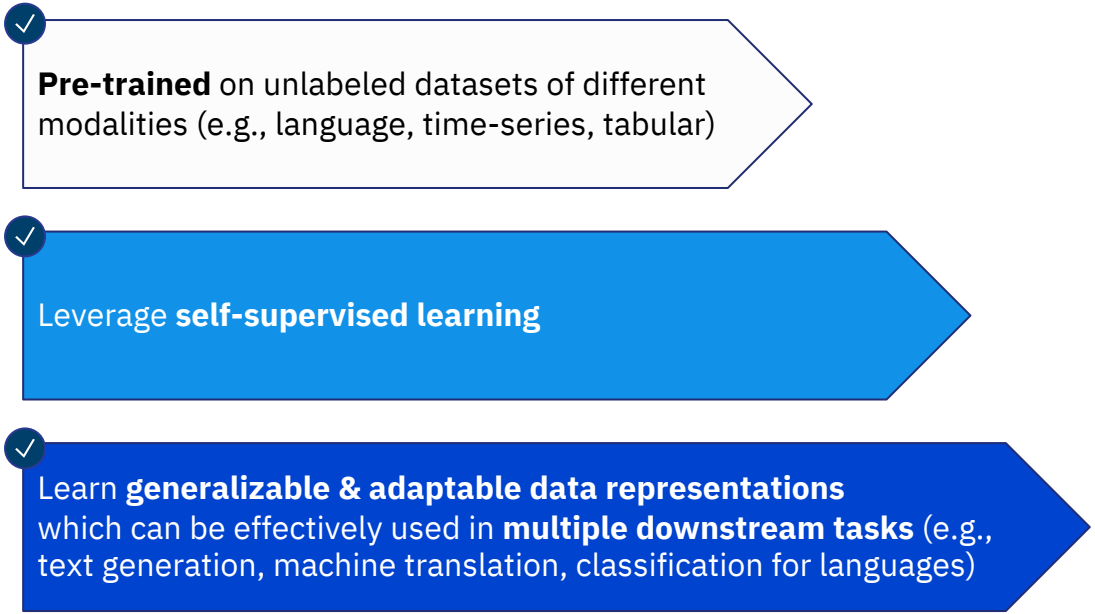
Now

Future

Evolving technologies

- Rapid advancements in AI
- Computing platforms

Foundation models



Note: while transformer architecture is most prevalent in foundation models, definition not restricted by model architecture

Source: Raghu Ganti

Why Foundation Model?

Advancing Application of Machine Learning Tools for NASA's Earth Observation Data

Jan. 21-23, 2020 | Washington, D.C.
Workshop Report

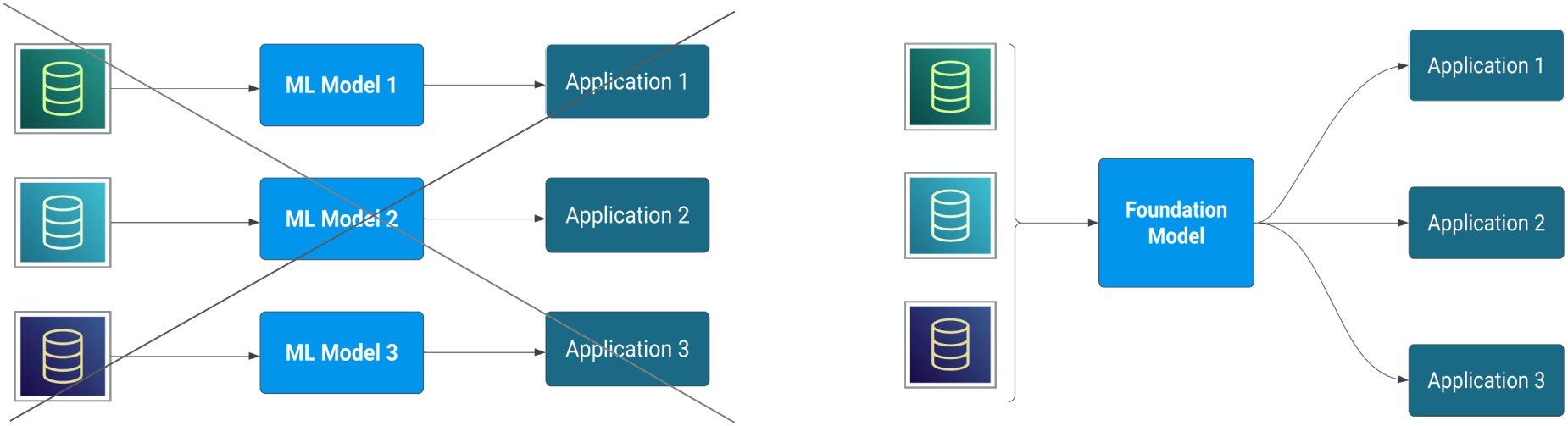
- Training data is the main component of supervised machine learning techniques and is increasingly becoming the **main bottleneck to advance applications of machine learning** techniques in Earth science.
- Geoscience models must **generalize across space and time**; however, for supervised learning one needs large training datasets to build generalizable models.

Potential innovative solutions in which to invest include techniques that require less data, such as transfer learning, representation learning, semi-supervised learning, and unsupervised learning.

Maskey, et al., Advancing AI for Earth science: A data systems perspective, Eos, 101, doi: 10.1029/2020EO151245. November 2020.

Advancing Machine Learning Tools for Earth Science: Workshop Report | Earthdata (nasa.gov)

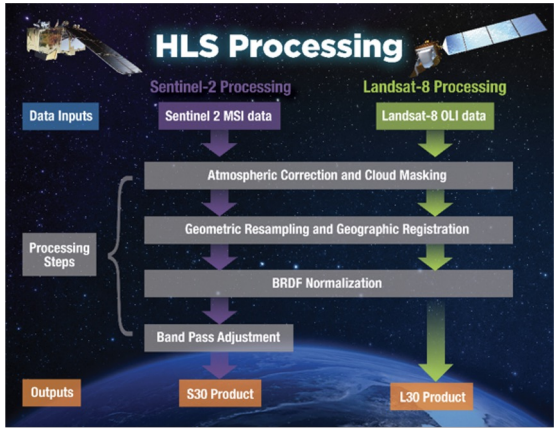
Foundation Models

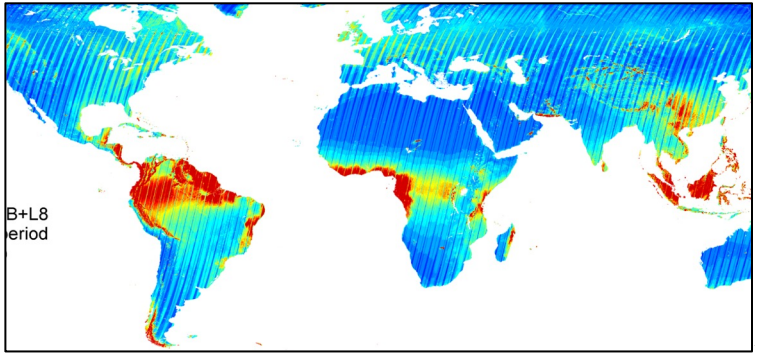


Foundation AI Model for Optical Remote Sensing Data Collaboration between NASA and IBM Research

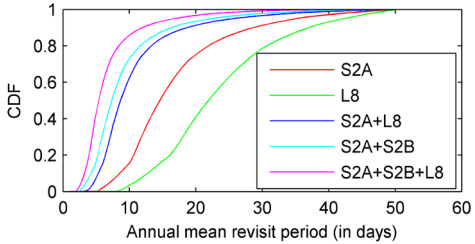
Harmonized Landsat Sentinel (HLS)

- "Seamless" near-daily 30m surface reflectance record including atmospheric corrections, spectral and BRDF adjustments, regridding
- Merges Sentinel-2 and Landsat data streams and can provide 2-4 day global coverage



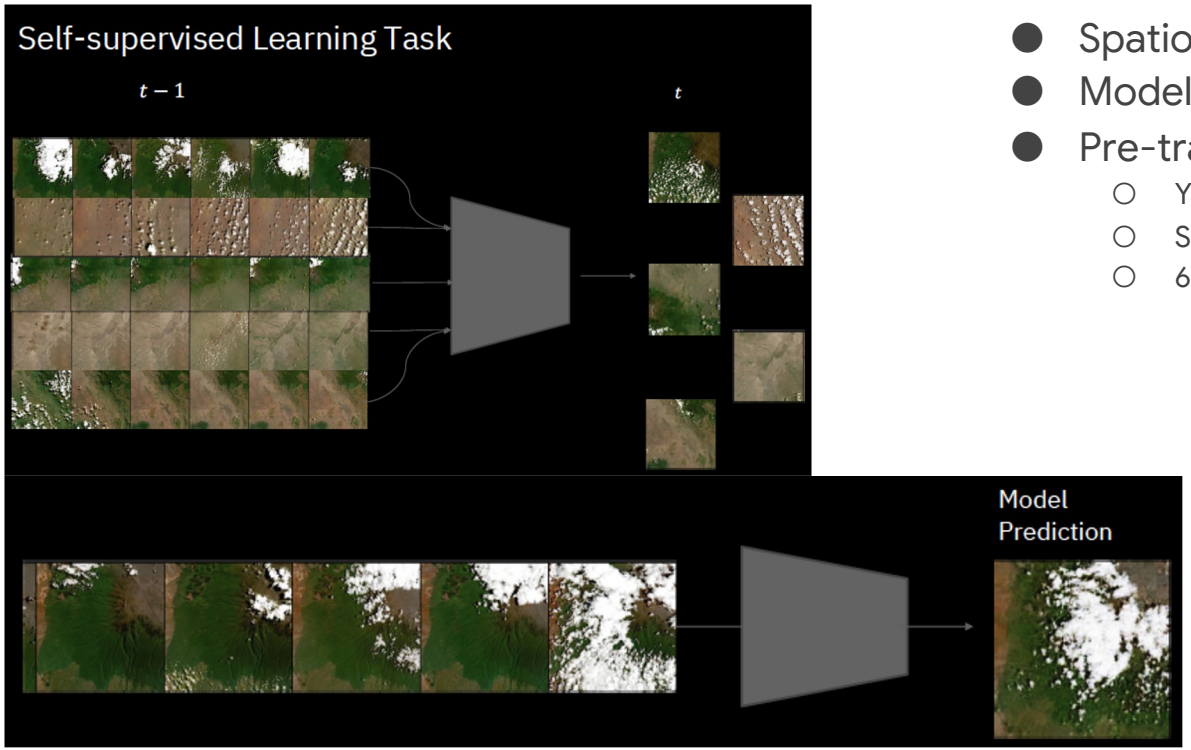


Potential Revisit using different Virtual Constellations



Foundation Model for Harmonized Landsat Sentinel (HLS) Data

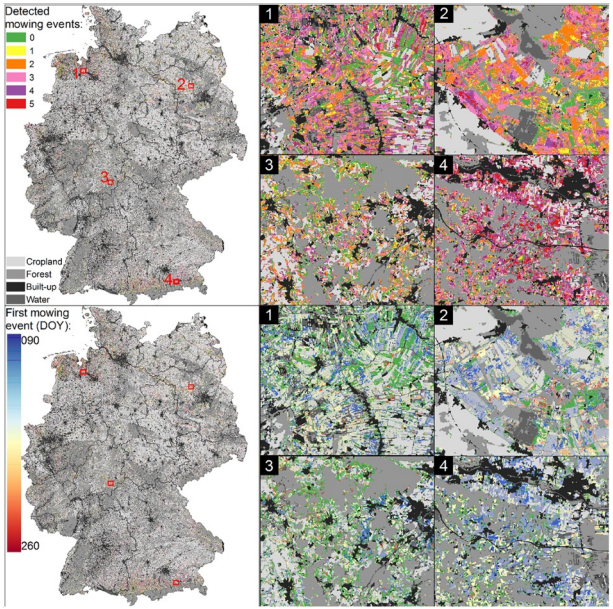
(IBM Research: Dr. Raghu Ganti)



- Spatio-temporal attention
- Model size 100 million parameters
- Pre-trained data
  - Year 2017
  - Spatial extent CONUS
  - 6 bands

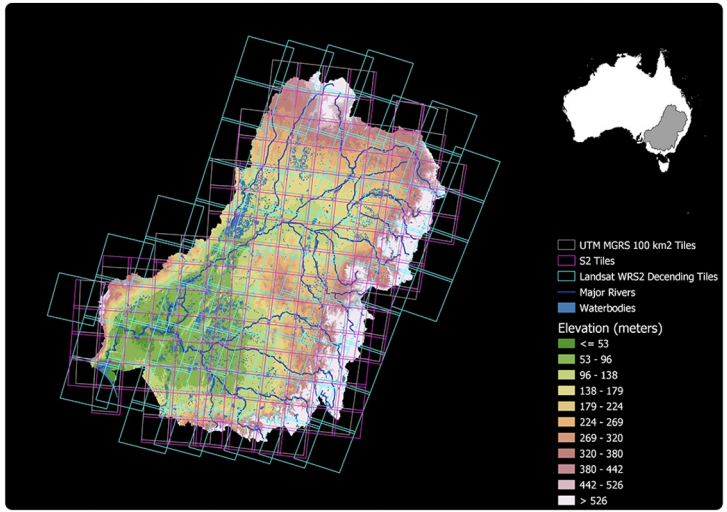
Downstream Applications of HLS Data

National-Scale Grassland Management Using HLS Data in Germany



Griffiths, P., et al, 2020, Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series

Ephemeral Floods Detection in Southeast Australia using HLS Data

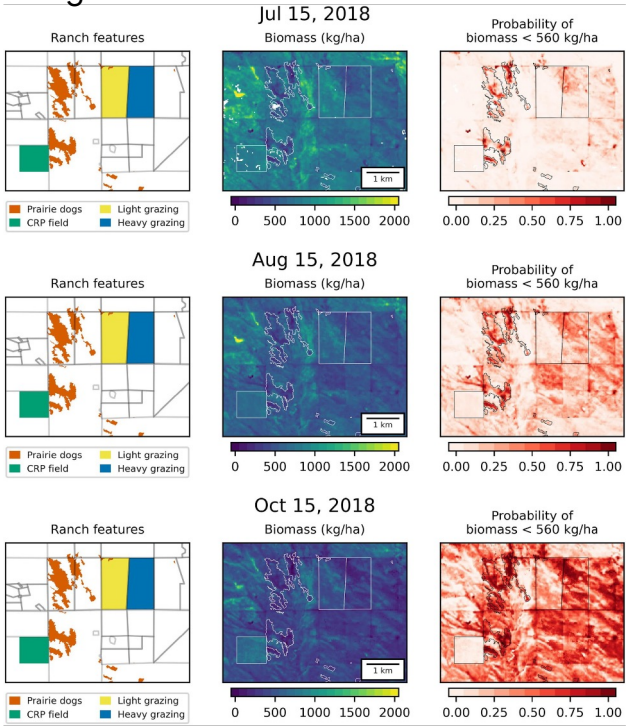


The Murray-Darling Basin is located in southeast Australia. This image shows satellite coverage of Landsat and Sentinel-2 over the study area. The basin's major rivers and water bodies are shown in blue and elevation derived from the Shuttle Radar Topography Mission range from green (low) to red (high). Image credit: Tulbure et al., 2022

Tulbure, MG, et al, 2022, ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING

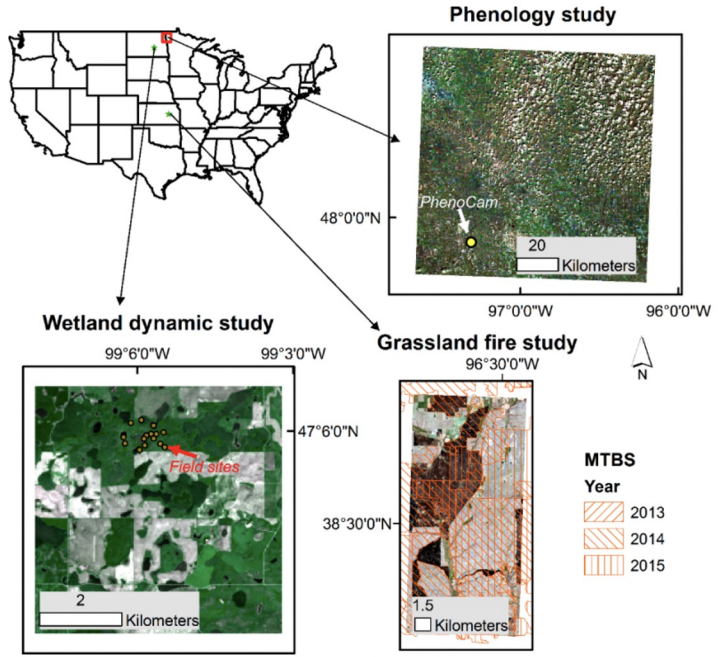
Downstream Applications of HLS Data

Biomass Estimation for Semiarid Rangeland Management in the US



Sean Kearney, et al, March 15, 2022, Remote Sensing of Environment

Wetland Dynamic, Grassland Fire, and Phenology Monitoring in Central US



Zhou, Q., et al, 2019. Monitoring Landscape Dynamics in Central US Grasslands with Harmonized Landsat-8 and Sentinel-2 Time Series Data. *Remote Sensing*

Downstream Applications of HLS Data

New satellite mapping with AI can quickly pinpoint hurricane damage across an entire state to spot where people may be trapped

8

Zhe Zhu, Assistant Professor of Natural Resources and the Environment, University of Connecticut and Su Ye, Postdoctoral researcher in environment and remote sensing, University of Connecticut See less

Fri, October 7, 2022 at 7:49 AM · 3 min read

Disaster response and recovery efforts

Cloud computing

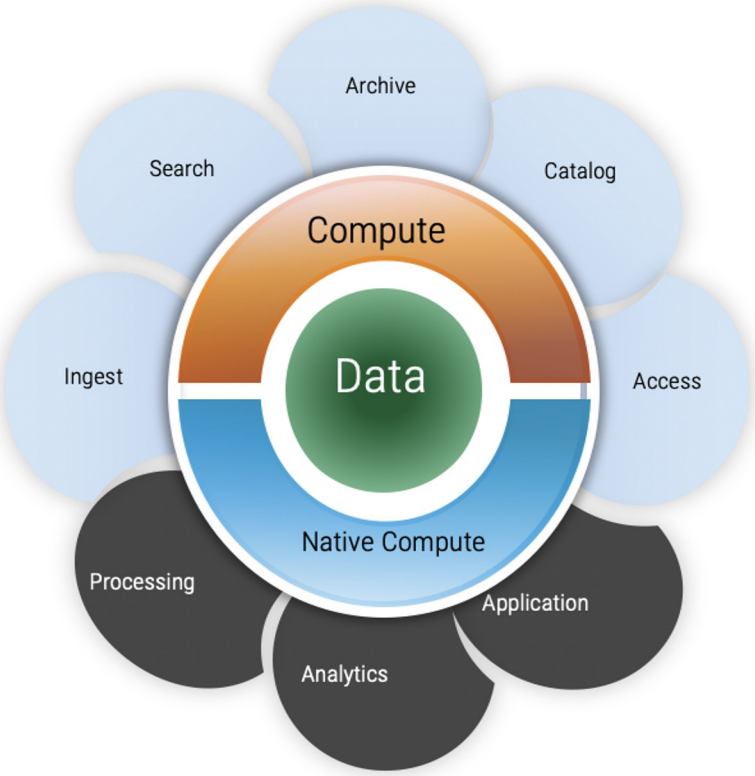
On-demand computing services: storage, compute, software,... as a service

- Amazon web services, Azure, Google cloud

Common characteristics:

- Elasticity: the ability to scale resources both up and down as needed
- Reliability: implies that the service is available and works as intended
- Pay as you go: only users pay for what they use
- Resource pooling: allows a cloud provider to serve its users in a multitenant model
- Minimal management effort: users can use and procure cloud services without much difficulty

Cloud computing

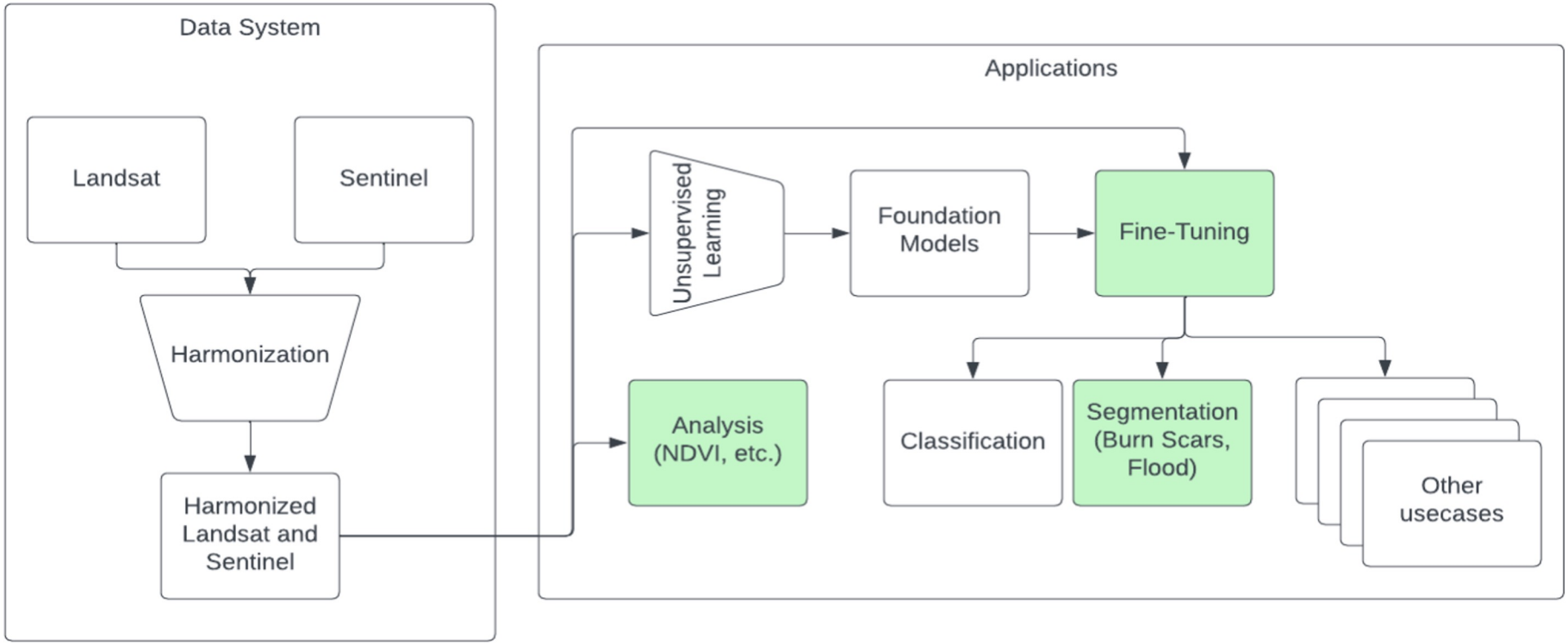


- Big data close to compute
- Data storage
- Scalable compute
- Cloud native

Cloud computing services

SaaS	**Browser**	**Cloud Applications** User interface, Reporting, Content management
PaaS	**Development Environment**	**Cloud Platform** Programming languages, Editors, Frameworks
IaaS	**Console**	**Cloud Infrastructure** Servers, Storage, Load balancers

Geospatial foundation model and applications



Thank you

manil.maskey@nasa.gov

- IEEE GRSS
- Earth Science Informatics TC
- HDCRS WG
- Jülich Supercomputing Centre
- IBM Research
- All the participants