Artificial Intelligence for Cold Regions (AI-CORE)

a Pilot to bridge Data Analytics and Infrastructure Development

Long Phan, Julia Christmann, Martin Rückamp

Angelika Humbert, Stephan Frickenhaus, Tilman Dinter

ML for Earth System Modelling and Analytics workshop 04.05.2021

HGF initiative „Helmholtz Incubator, Information & Data Science“

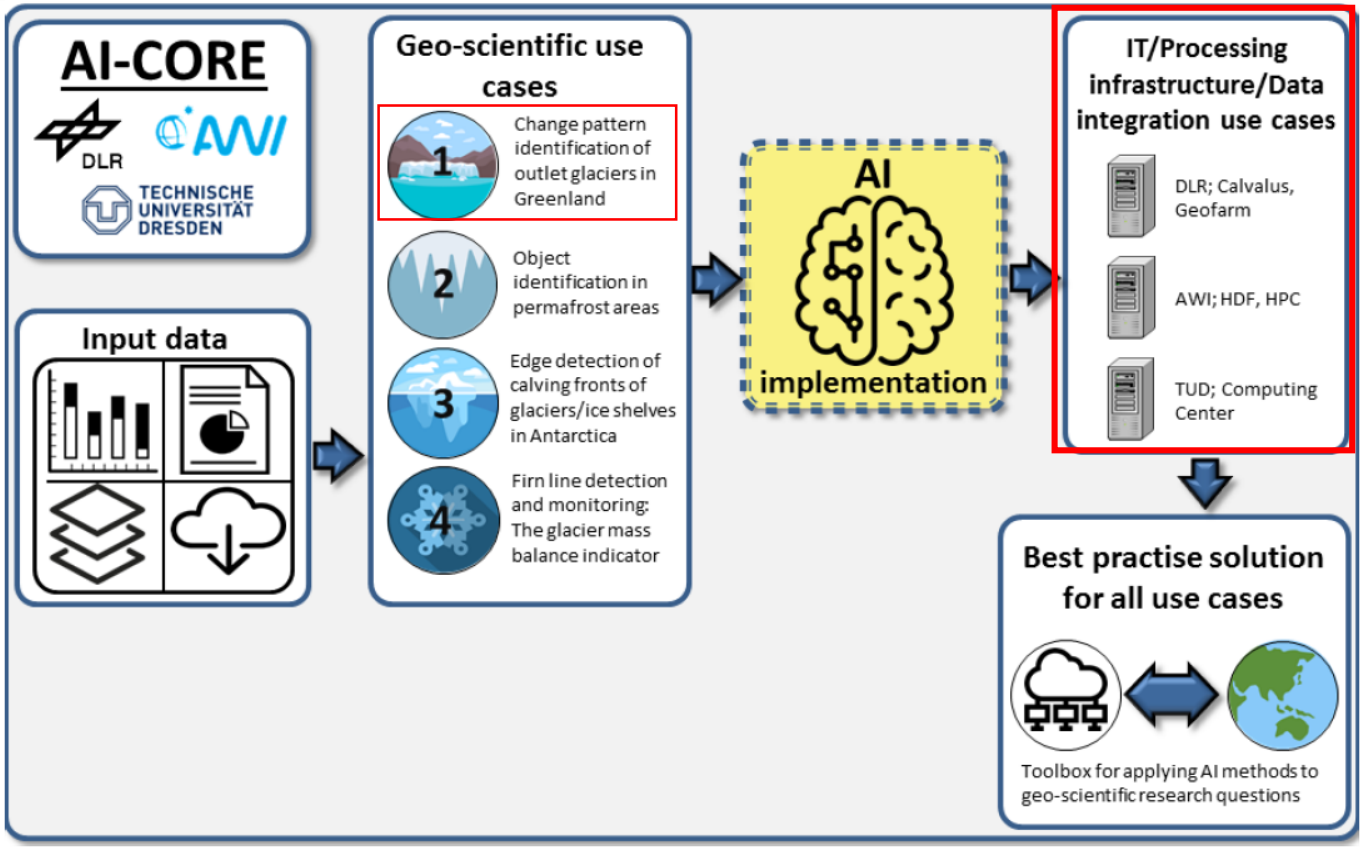


Figure 1: Workflow of AI-CORE

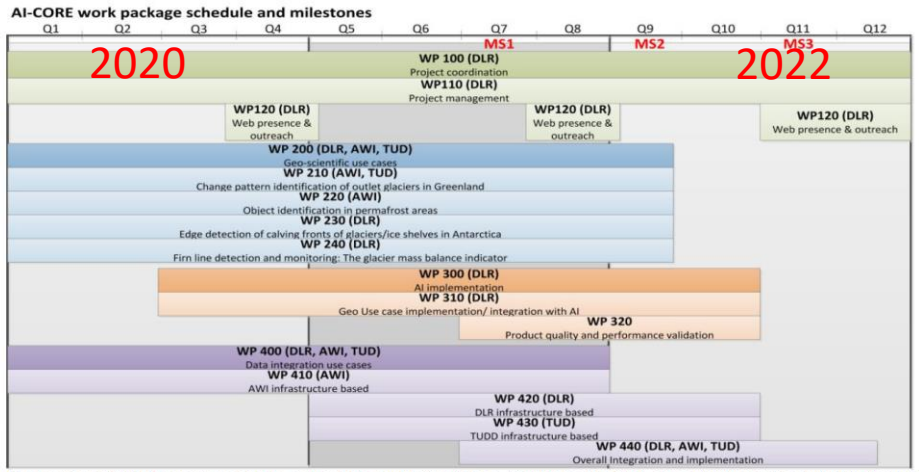


Figure 2: AI-CORE work package schedule and milestones. Partners leading an activity are included in brackets after the work package number

Geoscientific Use-case 1 - Greenland

Experimental design for marine ice sheet-ocean model: MISMIP+

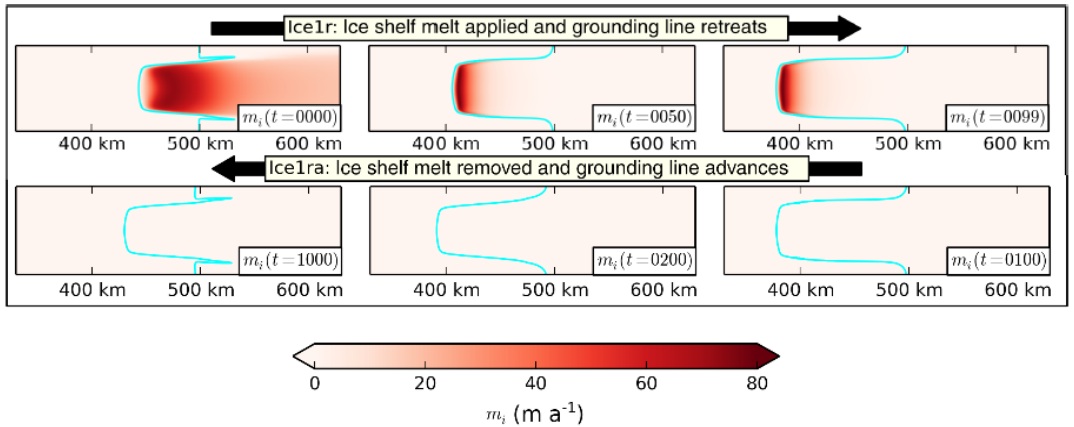
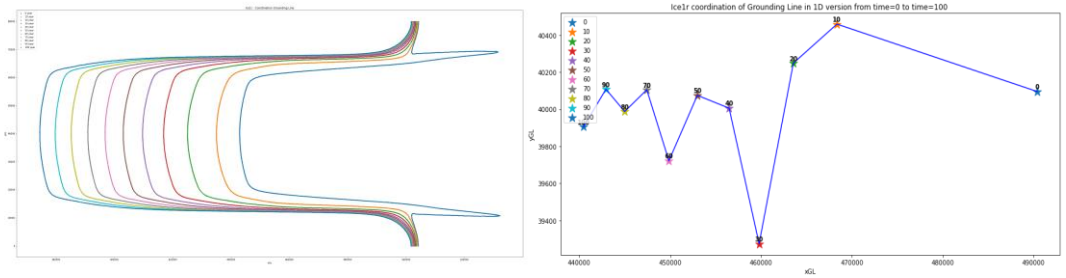
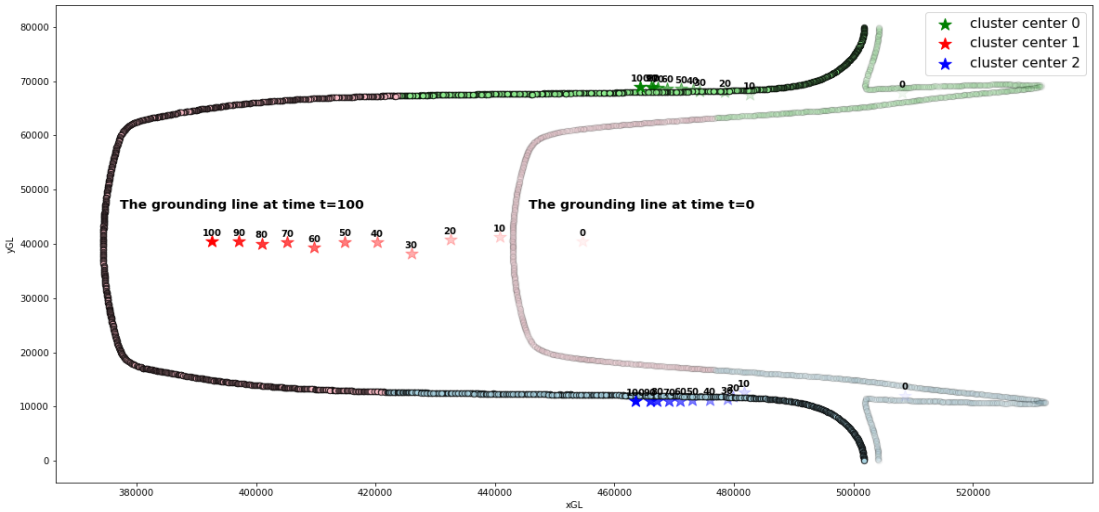


Figure: ice shelf melt rate mi fields during the Ice1r and Ice1ra experiments from a BISICLES run. Melt rates are applied when 0 < t < 100 a, causing the ice shelf to thin and grounding line to retreat. Once t > 100 a, no melt is applied, the ice shelf thickens, and the grounding line advances*

Aim: Clustering regions of small, middle and high change of grounding line migration with time.

Change pattern identification with Descriptive Analytics, Clustering K-Means

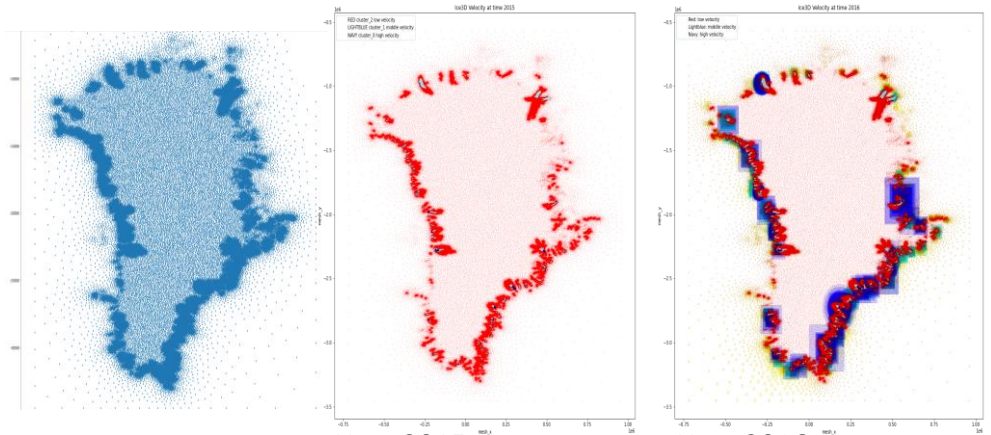




*Source: Experimental design for three interrelated marine ice sheet and ocean model intercomparison projects: MISMIP v. 3 (MISMIPC), ISOMIP v. 2 (ISOMIPC) and MISOMIP v. 1 (MISOMIP1), 2016, Asay-Davis et al.

Geoscientific Use-case 1 - Greenland

Sensitivity of Greenland ice sheet projections to spatial resolution: ISMIP6

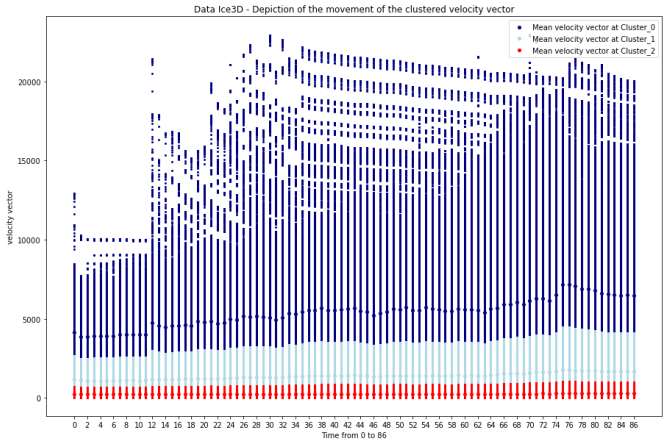


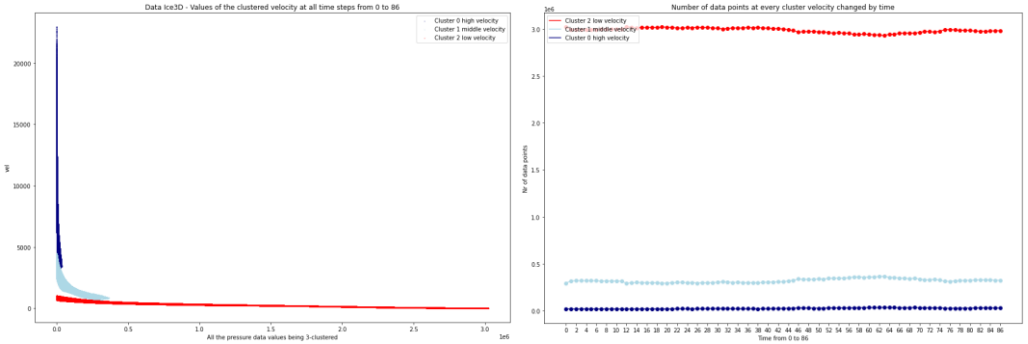
Year 2015 → Year 2016 → Year 2101

The dataset was initially generated from the simulations in the framework of the Ice Sheet Model Intercomparison Project for CMIP6 (ISMIP6) (Goelzer et al., 2020, Rückamp et al., 2020). The Clustering approach is hereby re-applied, and extended to observe the changes of the positions in term of **velocity**, ice mask, pressure, etc. in Greendland.*

Aim: Analyse and detect the temporal change of the ice mask, velocity with Descriptive Analytics and Clustering

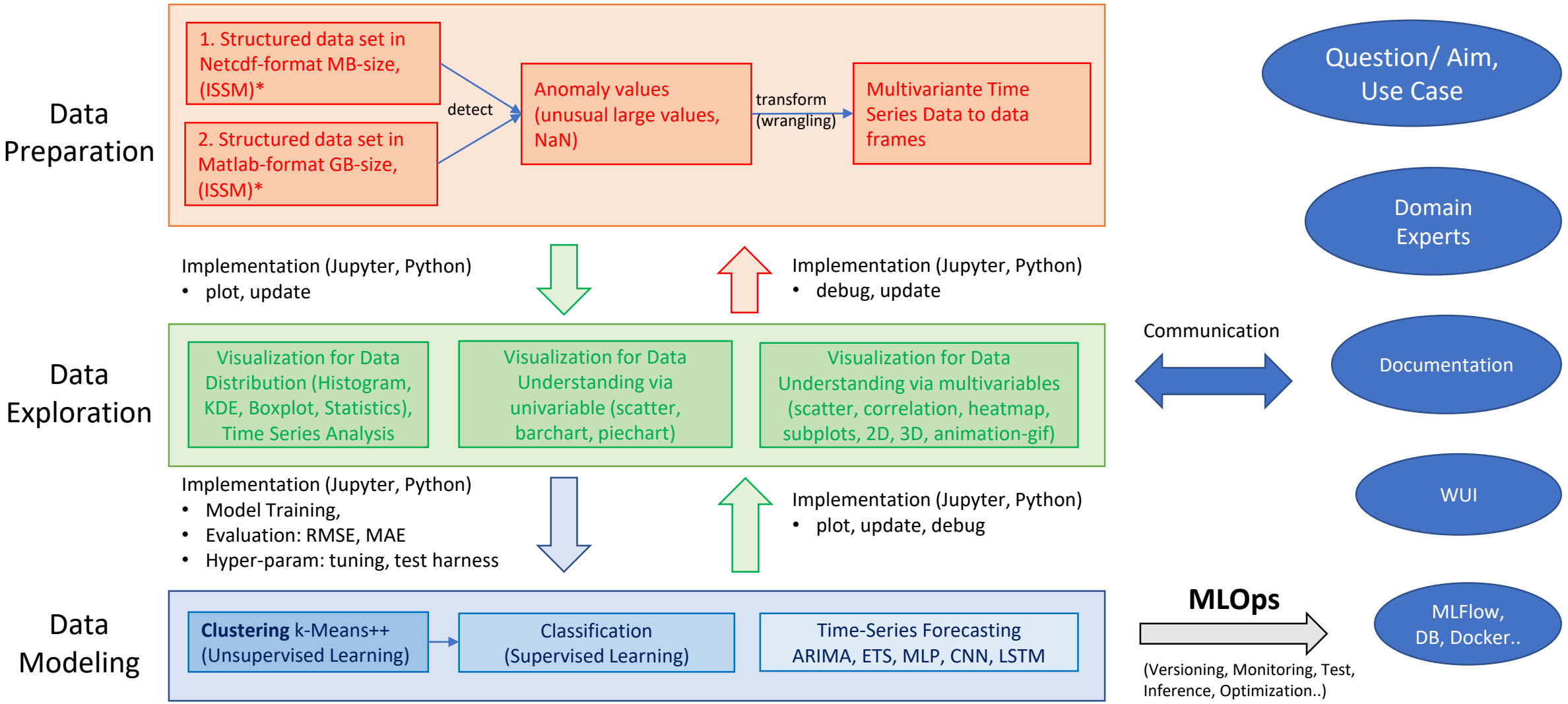
Change pattern identification in Greenland





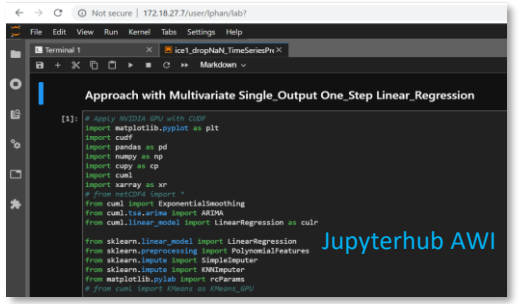
*Source: Sensitivity of Greenland ice sheet projections to spatial resolution in higher-order simulations: the Alfred Wegener Institute (AWI) contribution to ISMIP6 Greenland using the Ice-sheet and Sea-level System Model (ISSM) Rückamp et al.

Data Science Workflow in AI-CORE WP-210

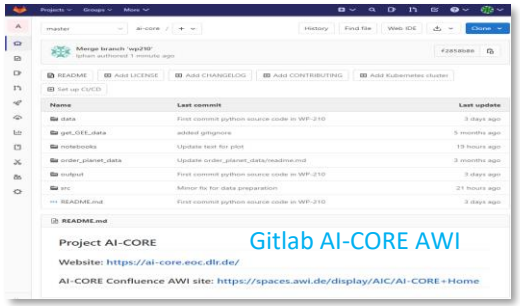


*ISSM: Ice-sheet and Sea-level System Model

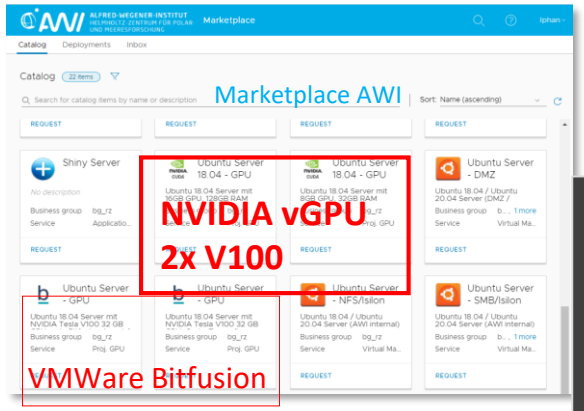
Overview ML-Infrastructure
First results in 2020-2021 WP-410 Infrastructure



Next: **Data Integration**
HPC with GPU

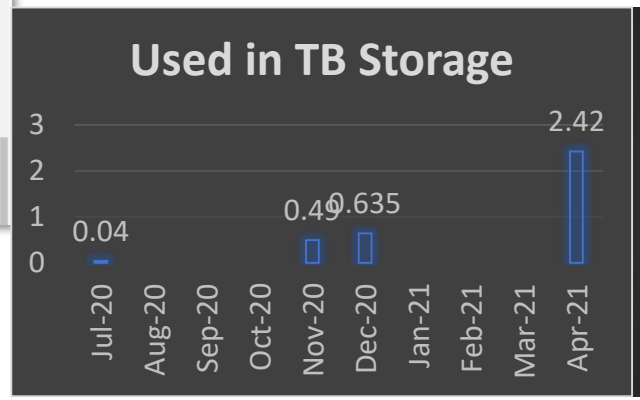


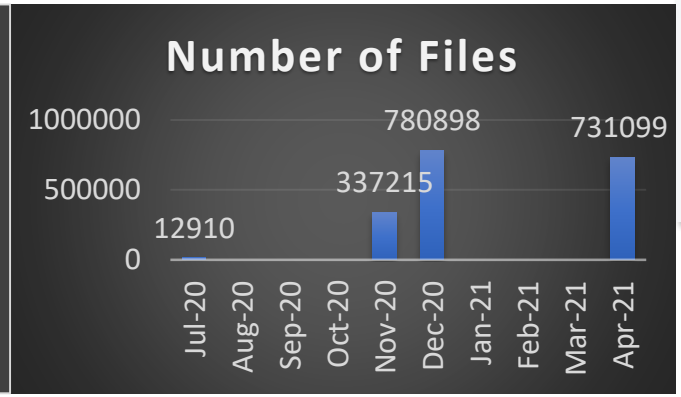
AWI_RZ_Marketplace GPU System → AWI_RZ_JupyterHub Software Development → AWI_RZ_Online Storage → AWI_Gitlab AI-Core → AWI_Confluence

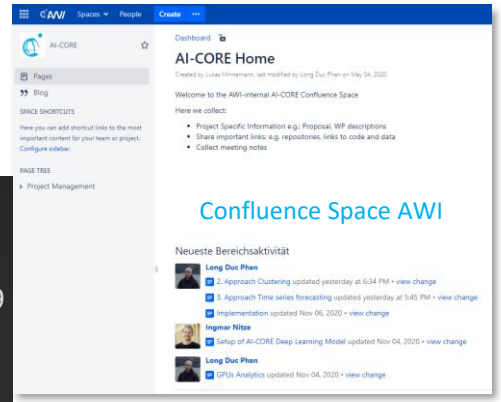


Type	Directory paths	Quota	Used	Files
Online storage	Windows | Mac OS | Linux	10.00 TB	2.42 TB (24.22%)	731099

Cloud Online Storage AWI







Software Stack – Data Science Toolkits

Overview System VM

CPU/ GPU-compatible AI-Toolkit

High-Level Programming: Python, Matlab, ..

Others: Jupyter, Visualization, streamlit, MLflow..

NVIDIA CLI Toolkit (nvidia-smi, ..)

CUDA Toolkit (cuda-memcheck, cuda-gdb, nvprof, visual profiler, ..)

Low-Level Programming: CUDA in C/C++

NVIDIA Driver Enterprise & licensing vCS

OS Ubuntu 18.04

VMWare 7

Setup Host environment (Network, Storage, Monitoring ..)

CPU

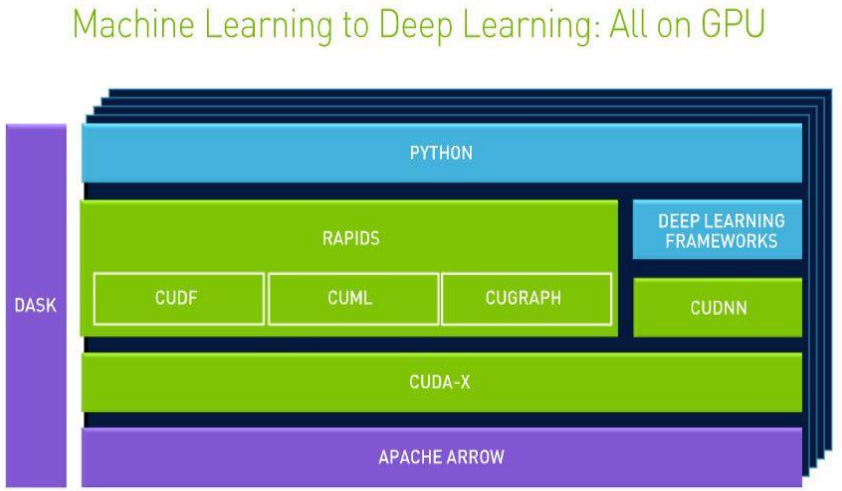
- Descriptive Analytics: Pandas, Numpy,
- Visualization: Matplotlib, seaborn
- Machine Learning: Scikit-Learn, Prophet
- Deep Learning: Tensorflow
- Distributed/ Parallel computing: xarray, Dask
- Data format: csv, json, netcdf, matlab

NVIDIA GPU

- Rapids: cuDF, cuPy,
- Rapids: cuxfilter, bokeh
- Rapids: cuML
- Tensorflow-GPU
- Dask-CUDA, CUDA in C/C++
- Apache Arrow, cuIO (int. NVIDIA)

Challenges/ further Tasks

- Many frameworks/ libraries, Algorithms
- Larger data set, limited (GPU-) memory size, database
- Different requirements to Versioning, Monitoring, Auto Pipeline, expect higher performance (CPU vs GPU)
- (De)bug, code improvement, data integration
- Complex scientific use case



https://www.nvidia.com/en-us/deep-learning-ai/solutions/data-science/

AI-CORE

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

Presentation of first results via Web User Interface (WUI)

Thank you for your attention ☺

long.duc.phan@awi.de

AI-CORE Homepage: www.ai-core.eoc.dlr.de