REQUEST FOR ADDITIONAL RESOURCES IN THE CURRENT YEAR FOR AN EXISTING EMI R&D PROJECT

Please submit the completed form via https://www.ecmwf.int/en/support (choose *Software and Computing*, then *Gain access to computing resources (create ticket)*) or send to special_projects@ecmwf.int

Country/Organisation: Italy/CMCC

Principal Investigator: Denis Eiras

Affiliation: CMCC

Address: Viale C. Berti Pichat 6/2 40127 Bologna

Project title: Towards a land/atmosphere/ocean/ice coupled data assimilation system: 20-year coupled reanalysis to initialise longer time-range forecasts, specifically seasonal forecasts (towards C3S multi-model project)

Project account: **SPITCARD**

Additional computing resources requested for year		**2026**
High Performance Computing Facility	[SBU]	80000000
Graphics Processing Unit Cluster - A	[GBU]	60000000
Graphics Processing Unit Cluster - B	[GBU]	
Total DHS Data storage capacity	[GB]	200000
EWC resources		
Number of vCPUs	[#]	96
Total memory	[GB]	1880
Storage	[GB]	200000
Number of vGPUs[3]	[#]	24

Continue overleaf

Technical reasons and scientific justifications why additional resources are needed

System Overview

The proposed system is a machine-learning-based coupled land–atmosphere–ocean data assimilation framework based on **Graph Attention Networks** (GATv2). Unlike the original SPREADS ensemble data assimilation system, which runs on CPUs, the training of graph neural networks requires GPU resources. The framework is designed to produce a dynamically consistent coupled analysis by jointly assimilating land and atmospheric observations on a global grid.
The system is organised in two phases. **PHASE0** generates the land background using a graph-based spatio-temporal model trained on the temporal evolution of soil moisture. **PHASE1** performs the coupled assimilation by combining the land background, the atmospheric background and the available observations to produce a corrected land and atmospheric analysis. Once trained, the system can be used to generate long coupled reanalyses at a much lower computational cost than a traditional ensemble-based data assimilation system.

Justification of the requested resources

GPU resources (24 GPUs, 1.88 TB total memory).
The requested GPU resources are required for training the graph neural network. The computational cost is dominated by the size of the coupled global graph, which includes all land nodes (for all soil layers) and atmospheric nodes (for all vertical levels). The resulting graph requires approximately **1.88 TB of memory**, far exceeding the memory available on a single 80 GB GPU. The use of 24 GPUs is therefore primarily a memory requirement rather than a means of reducing execution time. The training is performed over a 10-year period using 6-hourly analysis cycles.

CPU resources (96 CPUs).
CPU resources are required to prepare the graph datasets, process observations, perform data loading and manage communication with the GPUs during training. Allocating four CPUs per GPU ensures efficient utilisation of the GPU resources.

Storage (200 TB).
The requested storage is needed for global atmospheric and land background datasets, forcing fields, multi-instrument observation archives, graph datasets generated for training, model checkpoints and diagnostic outputs produced during both PHASE0 and PHASE1.