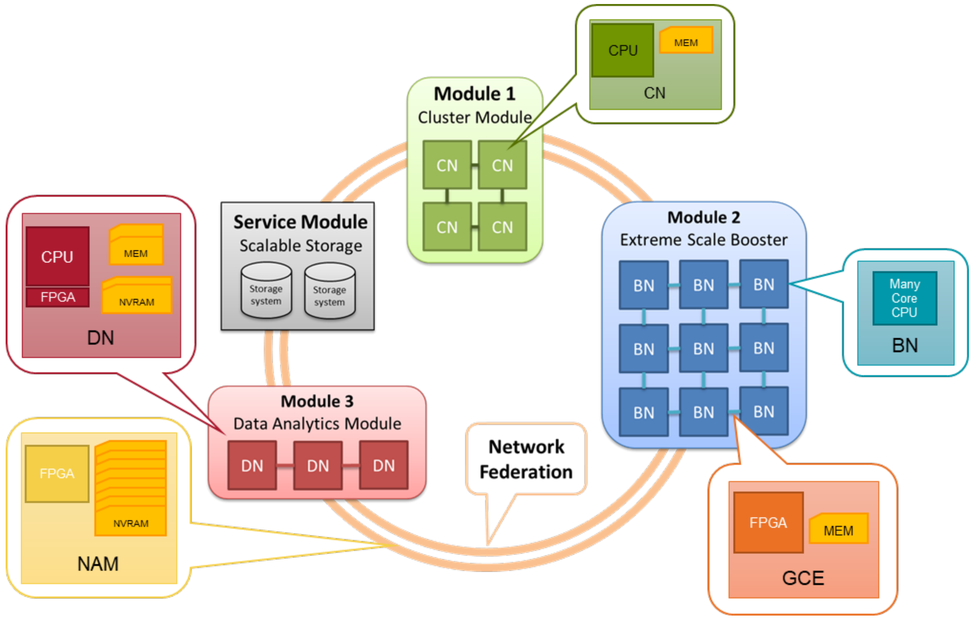
# Enhancing Deep Learning towards Exascale with the DEEP-EST Modular Supercomputer Architecture

Ernir Erlingsson[1] , Gabriele Cavallaro[2] , Morris Riedel[1,2], Helmut Neukirchen[1]
[1]Unversity of Iceland [2]Jülich Supercomputing Centre, Germany

## The DEEP-EST Modular Supercomputing Architecture (MSA)




### Cluster Module (CM)

The CM consists of several Cluster Nodes (**CNs**) with powerful CPUs with a moderate amount of memory and standard interconnects. It delivers a fast single-thread performance and is ideal for computationally expensive tasks that scale poorly.

### Extreme Scale Booster (ESB)

The ESB is suitable for highly scalable tasks, having numerous Booster Nodes (**BN**) with manycore moderate CPUs and memory, but fast interconnects. Additonally, the Global Collective Engine (**GCE**) can be used to speed up communication between the nodes.

### Data Analytics Module (DAM)

The DAM has a small number of Data Nodes (**DNs**) that have a lot of fast memory complemented with good CPUs and an accelerator that for example uses Field Programmable Gate Arrays (**FPGAs**) or General Purpose Graphic Processing Units (**GPGPUs**).

### Network Attached Memory (NAM)

The NAM is a device with which exists in the fast fabric interconnection between all other modules. It has an abudant amount of high-performance memory and an FPGA, making it suitable for many roles, such as a storage target with fast checkpointing or near-data-processing.

### Enhancing Deep Transfer Learning

As members of the DEEP-EST consortium the University of Iceland and Jülich Supercomputing Centre will enhance Machine Learning (**ML**) towards Exascale by exploiting the MSA. One example involve applying Transfer Learning (**TL**) methodology on Convolutional Neural Networks (**CNNs**) that have been pre-trained with large image datasets. These CNNs are used as the foundation for the training of new CNNs, extending the classification labels facilitated by previously unseen datasets.

(1) One or more pre-trained CNNs are loaded into the NAM module from the storage module

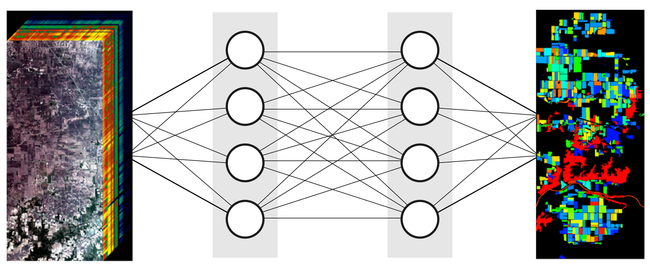
(2) New CNNs are trained in the DAM, or alternatively the ESB module, by using existing CNNs as fixed feature extractors for new datasets. Additionally, the DAM's FPGA can be utilized to accelerate the time needed to prepare existing CNN for training with the new datasets

(3) The trained models are stored in the NAM module to speed-up the evaluation and comparison phase

(4) Model evaluation is embarrassingly parallel and is therefore suitable for the ESB module and its many-core architecture. Specific testing datasets are combined with the trained CNN models residing in the NAM

(5) Evaluation results are written to the NAM module and its FPGA is used to determine the model accuracies.

(6) The described process steps can then be repeated any number of times to improve the model and reduce the likelihood of a local minimum convergence.

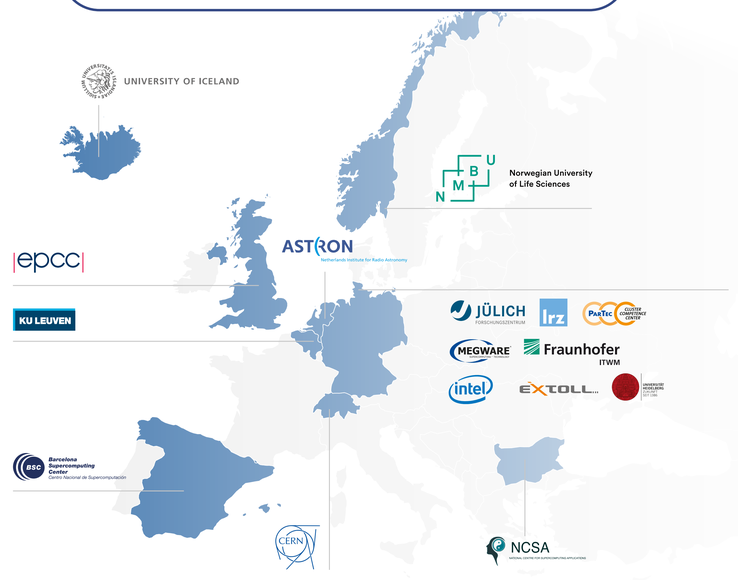



### DEEP Projects

The DEEP projects are a series of HPC research projects, funded by the European Commission, that focus on delivering the blueprints for future Exascale performance. The current DEEP project, Dynamic Exascale Entry Platform - Extreme Scale Technologies (**DEEP-EST**), is the third in line with sixteen partners involved, from both leading international research institutions and industry.

DEEP-EST aims at creating a co-designed heterogenous modular supercomputing architecture (**MSA**) which supports state-of-the-art simulations and a growing variation of supercomputing tasks. The project will have developed and evaluated a prototype by 2020.

DEEP Projects




UNIVERSITY OF ICELAND

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

DEEP Extreme Scale Technologies

This project received fundingfrom the European Commission's Horizon 2020 Programme for research, technological development and demonstration under Grant Agreement no 754304