Integrating HPC and Deep Learning in converged workflows

Andrew Grant[1]

[1]VP, HPC and Quantum, Atos UK & Ireland

In the last few years the use of AI and specifically deep learning techniques has emerged as a key pillar in scientific discovery. While many of the underlying techniques have been around for some time the computational power and data volumes required to make them effective have only recently become available. Deep Learning provides new methods to improve predictive accuracy, response times and insight into new phenomena, often using data sets that would previously have been considered unmanageable.

HPC, on the other hand, has a long history of modelling and simulation of physical phenomena and a track record of enabling grand challenge scientific discovery with a proven return on investment in multiple science domains.

As the requirements of deep learning environments grow, the underlying architectures start to share many characteristics with HPC systems, leading to a general convergence in architectural models. However, they also share many of their issues. With processor frequencies and single thread performance now plateauing, the only way to satisfy the insatiable demand for the additional performance required to drive increasingly large AI workloads is through increased parallelism.

This talk will firstly discuss different approaches to converged HPC and AI workloads exploring the following methods: Augmentation – where experimental/simulated data is used to train the deep learning neural networks, replacing significant portions of a conventional simulation, Modulation – where deep learning is used to reduce the number of ensemble runs needed in a parameter sweep in order to steer a simulation, and Transformation - where live data streams are used to continuously improve an HPC simulation.

Secondly, the talk will discuss converged architectural models to manage the hybrid workloads above by describing emerging software frameworks to ease the burden on the data scientist.

CompBioMed Conference 2019