



Contribution ID: 531

Type: Oral

Delivering a machine learning course on HPC resources

Monday, November 4, 2019 2:30 PM (15 minutes)

In recent years proficiency in data science and machine learning (ML) became one of the most requested skills for jobs in both industry and academy. Machine learning algorithms typically require large sets of data to train the models and extensive usage of computing resources both for training and inference. Especially for deep learning algorithms, training performances can be dramatically improved by exploiting Graphical Processing Units (GPU). The needed skill set for a data scientist is therefore extremely broad, and ranges from knowledge of ML models to distributed programming on heterogeneous resources. While most of the available training resources focus on ML algorithms and tools such as TensorFlow, we designed a course for doctoral students where model training is tightly coupled with underlying technologies that can be used to dynamically provision resources. Throughout the course, students have access to OCCAM, an HPC facility at the University of Torino, managed using container-based cloud-like technologies, where Computing Applications are run on Virtual Clusters deployed on top of the physical infrastructure.

Task scheduling over OCCAM resources is managed by an orchestration layer (such as Mesos or Kubernetes), leveraging Docker containers to define and isolate the runtime environment. The Virtual Clusters developed to execute ML workflows are accessed through a web interface based on JupyterHub. When a user authenticates on the Hub, a notebook server is created as a containerized application. A set of libraries and helper functions is provided to execute a parallelized ML task by automatically deploying a Spark driver and several Spark execution nodes as Docker containers. This solution automates the delivery of the software stack required by a typical ML workflow and enables scalability by allowing the execution of ML tasks, including training, over commodity (i.e. CPUs) or high-performance (i.e. GPUs) resources distributed over different hosts across a network.

Consider for promotion

No

Primary authors: LEGGER, Federica (Universita e INFN Torino (IT)); BAGNASCO, Stefano (Istituto Nazionale di Fisica Nucleare, Torino); LUSSO, Stefano (Universita e INFN Torino (IT)); Mr FRONZE', Gabriele Gaetano (University e INFN Torino (IT), Subatech Nantes (FR)); VALLERO, Sara (Universita e INFN Torino (IT))

Presenter: LEGGER, Federica (Universita e INFN Torino (IT))

Session Classification: Track 8 – Collaboration, Education, Training and Outreach

Track Classification: Track 8 – Collaboration, Education, Training and Outreach