Société canadienne de CFD
CFD Society of Canada
161 LOUIS PASTEUR MCG CBY A331
Ottawa, ON K1N 6N5
https://www.cfdcanada.ca/

Société canadienne de CFD Society of Canada

DRI Needs Assessment for the Computational Fluid Dynamics (CFD) Research Community

December 14, 2020

The Computational Fluid Dynamics (CFD) Society of Canada is a non-soliciting charitable organization devoted to advancing research and education in CFD within Canada. CFD involves numerically solving sets of partial differential equations governing the motion of fluids, which also includes aerodynamics, hydraulics, heat transfer, two phase flows, acoustics, reactive flows, fluid structure interactions and other related disciplines. CFD is an important approach across multiple engineering disciplines as well as natural and applied sciences, including atmospheric sciences and condensed matter physics. CFD has become an instrumental tool in many industries, such as aerospace, wind energy, automotive, civil, materials, nuclear, medicine and many others that impact not only Canada's economic competitiveness, but also certification, security, and environmental targets as well as other critical subjects such as the transmission of airborne infectious diseases. With more than 100 active members across academia and industry, the CFD Society of Canada is submitting this White Paper in response to the NDRIO's Call for Future Needs Assessment to highlight the unique digital research infrastructure (DRI) requirements of our discipline.

Current state

Overall, the current DRI resources are generally well adapted to the needs of the CFD community, and are essential for sustaining CFD research within Canada. In particular, large high-performance parallel computing systems that are supported and upgraded to remain internationally competitive are essential for CFD researchers to remain competitive with peer researchers around the world.

In the 2020 RAC competition, 71 CFD-related projects were allocated a total of 28,777 core-years. The bulk of CFD-related projects (45%) are awarded fewer than 160 core-years, although there are a small number of CFD researchers receiving allocations of more than 1000 core-years. In total, CFD-related projects received over 16% of the CPU core-years allocated in the 2020 RAC. In contrast, CFD-related projects received less than 2% of total allocated GPU-years. This highlights that the CFD research community currently favors parallel jobs on homogeneous clusters employing large numbers of CPU cores, and GPU resources are less widely used at present. This is mainly due to the specificity of fluid dynamic sets of partial differential equations which involves highly non-linear terms and large numbers of variables to be solved in double precision. Of course, as GPU-accelerated software tools become more prevalent within the CFD community, this may change over time. In particular, some research CFD tools are making the efforts to include hardware accelerator directives in strategic sections of the algorithms but these developments likely require some years before they are widespread. New CFD concepts such as Lagrangian or Lattice Boltzmann techniques based on simpler sets of partial differential equations are already designed to take advantage of the GPU resources, but require further development to reach the fidelity of traditional approaches that have been in development for almost 30 years.

Project storage allocations are also critical for CFD researchers; in 2020, CFD-related projects received nearly 7% of the total allocated project storage. CFD research often investigates the temporal evolution of spatial fields related to a flow, which can generate very large datasets of temporally-sequenced results, and thus generate high Input/Output (IO) loads. Integrating large parallel compute clusters with high-performance IO resources is essential to efficiently generate and store CFD research data. Access to large, persistent storage allocations are essential for ensuring the valuable data researchers have collected in previous years remain useful and accessible.

Associated with very large datasets produced by CFD-related projects is the need for remote, parallel visualization of the generated data. Clusters that are setup for remote, parallel visualization of large datasets are also a key factor in supporting CFD research. Related to this is the need for ongoing training on remote parallel visualization of large datasets.

DRI challenges particular to CFD

As stated above, CFD projects involving unsteady problems (for example, aerodynamics, aeroacoustics, turbulence, etc.) possess very large IO requirements and large project storage requirements. Moreover, many CFD projects require long wallclock times to compute meaningful statistical measures of temporally and spatially evolving flow variables. Consistency in maximum wallclock times across similar compute resources will improve access for CFD researchers in different regions and improve uniformity.

CFD projects based on massively parallel solvers require thousands of processors to be allocated during computation. These massively parallel jobs are becoming increasingly important as industrial simulations move to higher levels of physical fidelity and thus much higher computational cost. Scheduler policies need to be tailored to not only favor the maximum usage of the machine, but also account for the particular needs of such large jobs to avoid excessive queue times.

An ongoing challenge facing many CFD researchers in academia is the training of students and competence retention in a research group. Significant differences exist in the training opportunities provided across regional consortia. Training is often offered at a level that is too novice to be useful for training graduate students, or not specifically tailored to the tools and workflows relevant to CFD.

Currently, allotment of DRI resources occurs through an annual RAC competition. However, most CFD researchers operate through multi-year industrial or Tri-council projects, which require access to DRI over multiple years. Most project planning time frames are much longer than the annual RAC competitions. This makes planning for project resources very challenging; there is no certainty what resources will be available in the future. Moreover, it is highly problematic for research projects requiring a certain level of DRI access to have their allocation suddenly reduced at mid-project due to uncertainties in the annual RAC competition. This is particularly incoherent when the project has already received peer review and scrutiny by NSERC or another Tri-Council agency. This is even more critical for storage space. CFD projects involving unsteady problems generate large amounts of data which become reference inputs for the development of reduced and statistical models in many engineering fields. These databases must be stored for a long period of time, typically three to five years. Some of them need to be shared across research groups or even through international collaborative teams.

Future DRI State

As emphasis grows in emerging areas like data science, machine learning, and artificial intelligence, it may affect the nature of the computing infrastructure managed by NDRIO. Despite these exciting advancements, it is important that NDRIO also retains internationally competitive infrastructure employing more traditional methodologies that build on large, parallel, homogeneous clusters. As highlighted above, the bulk of CFD-related projects involve large parallel jobs requesting large numbers of CPU cores. Therefore, the renewal of large, parallel, homogeneous cluster resources, with high performance IO servers and associated network facilities, is important for on-going and future CFD research. In particular, systems that are set up for large jobs (at least 100s of processors) and the expansion of data storage space to researchers are essential.

Support for integrating cloud services for code repositories and data sharing will be increasingly important as research data management and open data initiatives are requiring more complex archival and sharing of scientific datasets.

How to Bridge the Gap

Below are suggestions for meeting the DRI needs of the Canadian CFD community.

- To tackle modern engineering challenges, researchers are using more and more high-fidelity approaches. This is contributing to a continuous increase in CPU and storage demands. Hence, continuous investment in hardware is required to keep the competitiveness of the Canadian CFD research community.
- Improve training offerings: there are a lot of beginner training sessions presently. But there is a gap between beginner and HPC expert levels to be filled. Intermediate to advanced training on schedulers, parallel programming with MPI, openACC, version control, and remote parallel visualization of large datasets. Training on diagnostic tools to evaluate performance, communication and read/write latency of massively parallel simulations would provide students the knowledge to use the DRI resources in a wiser manner.
- Offering thematic schools based on disciplines rather than summer schools based on tools. CFD is a sufficiently broad discipline (as indicated by the number of 2020 RAC awards conducting CFD projects) that a thematic school targeting “HPC for CFD” would generate significant interest, and introduce toolchains and best practices

covering much of the training requirements for a large number of Canadian CFD research groups.

- Canadian hackathon around computational or graphical challenges. The “Visualize This” challenge held by Compute Canada for the past few years is an example of such a challenge which should be generalized in more aspects.
- Enhance links between universities (professors), discipline associations (such as CFDSC) and Compute Canada support services (experts) to provide specialized training on CFD-specific topics. Linkages between faculty and NDRIO can be leveraged to provide CFD-specific workshops. Participation of the NDRIO in the CFD Society of Canada annual meetings could be a good communication channel to keep the CFD research community informed of the NDRIO’s activities, initiatives, plans, asset purchases, and training opportunities.
- Facilitate academia-industry collaborations by introducing tri-sector partnerships between industry, CFD researchers at universities, and DRI by providing new RAC application programs. This could pave a path for industrial partners to directly contribute to improvement of DRI computational resources for advancement of Canadian industries such as aerospace, agriculture, mining, bioengineering, and quantum computers.

Signatories

François Morency (École de Technologie Supérieure, President CFDSC)
Joshua Brinkerhoff (UBC-Okanagan, Treasurer and Secretary CFDSC)
Marlène Sanjosé (École de Technologie Supérieure)
Claire Deschênes (Université Laval)
Guy Dumas (Université Laval)
Jean-Pierre Hickey (University of Waterloo)
Sébastien Houde (Université Laval)
Rajeev Jaiman (UBC-Vancouver)
Éric Laurendeau (École Polytechnique de Montréal)
Yvan Maciel (Université Laval)
Stéphane Moreau (Université de Sherbrooke)
Siva Nadarajah (McGill University)
Ugo Piomelli (Queen's University)
Jérôme Vétel (École Polytechnique de Montréal)
Xiaohua Wu (Royal Military College of Canada)
Alireza Najafi-Yazdi (Anyon Systems Inc.)
Mostafa Najafiyazdi (MAYA HTT, Toronto)