POST-PROCESSING OF LARGE-SCALE DEM SIMULATIONS FOR END-USER ANALYTICS AND VISUALISATION

A. Janda, G. Filippone, K.J. Hanley, S. Papanicolopulos, J.P. Morrissey, P. Totoo, J.Y. Ooi

The University of Edinburgh, Edinburgh EH9 3JL, UK

SUMMARY

Continual advances in computational power and increasing access to HPC or Cloud infrastructures is enabling researchers to use discrete element modelling (DEM) for particle-scale simulation of large-scale engineering processes. These scenarios may involve millions of particles and/or time-steps and therefore the amount of simulation data produced can be of the order of terabytes or more. As a consequence, the post-processing of the results can become so time-consuming and the data storage requirements so substantial that it is no longer possible to conduct them in a single workstation.

VELaSSCo (Visual Analysis for Extremely Large-Scale Scientific Computing) is an EC FP7 project involving a consortium of seven European partners: CIMNE, the University of Edinburgh, SINTEF, INRIA, Fraunhofer, Jotne EPM and Atos. VELaSSCo aims to provide new visual analysis methods for large-scale simulations serving the petabyte era. This is done by adopting Big Data tools/architectures for the engineering and scientific community, leveraging new ways of in-situ processing for data analytics and hardware-accelerated interactive visualisation. The main output of the project is the VELaSSCo platform which has been designed and developed to perform distributed post-processing and visualisation of the results of very large engineering simulations including DEM, finite element method (FEM) and computational fluid dynamics (CFD) simulations.

The architecture of the VELaSSCo platform is based on the open-source Hadoop software stack, a Java-based framework for distributed storage and processing of Big Data. It is composed of two main layers: the Data layer which is responsible for storing, accessing and translating the simulation data and the Query Engine Layer which is in charge of receiving the user requests (VELaSSCo queries) from the visualisation client, extracting and/or analysing the simulation data and returning the results in a GPU-friendly format for fast visualisation. One complex VELaSSCo query is the discrete-to-continuum transformation

(D2C). This applies temporal and spatial coarse graining to DEM simulation data in order to compute bulk quantities (e.g., solid fraction) that are projected onto a continuum field. In this abstract, the platform is described and preliminary evaluation results of the D2C query are shown to demonstrate the considerable potential of the platform.

1: Overview of Research and Platform Architecture

VELaSSCo has been developed for supporting new visual and analysis methods serving the petabyte era. Modern simulation solvers have the capability of exploiting shared or distributed memory parallelism but, in contrast, post-processing and visualisation tools have not matched these developments which is the issue addressed by VELaSSCo. The VELaSSCo platform exists in two versions: a fully open-source version and a proprietary version which use Apache HBase and Open SimDM as database systems, respectively. The open-source version shown in Figure 1 is being revised continually to increase efficiency.

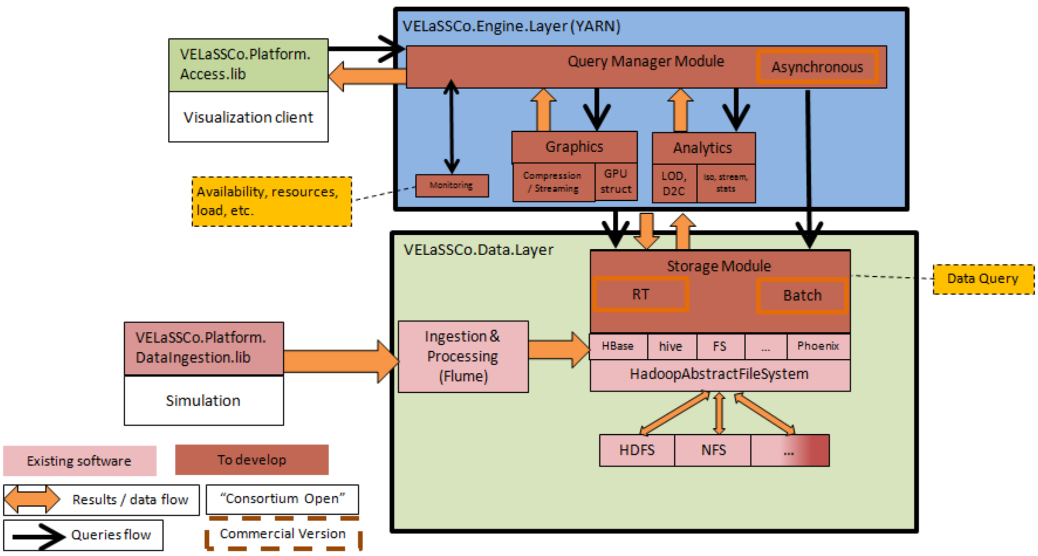

Figure 1: The VELaSSCo architecture used for the open-source platform.

VELaSSCo Queries (VQueries) are the means of traversing the architecture, providing end user functionality and, thus, connecting visualisation with data storage, graphics and analytics modules. VQueries are triggered by user interactions with the visualisation client. These are sent by the Access Library to the Query Manager Module (QMM) of the Engine Layer. The QMM, which supports asynchronous queries, is the core module of the platform and is composed of a Thrift server supplemented by some purpose-written classes. Its function is to provide communication services to the clients of the platform, and to dispatch VQueries to the correct modules. Simple data access VQueries are passed to the Storage module of the Data Layer. More complex VQueries, e.g., analytics transformations such as calculating the boundary of a mesh or a cut plane, are implemented as MapReduce jobs in the Analytics module of the Engine Layer. D2C is another example of a complex VQuery which applies temporal and spatial coarse graining methods (Goldhirsch, 2010; Labra et al., 2013; Weinhart et al., 2015) to DEM simulation data in order to compute bulk quantities (e.g., solid fraction, stress tensors) that are projected onto a continuum field. These computationally-expensive jobs are distributed over the Hadoop nodes.

The Graphics Module is in charge of formatting stored data into a GPU-friendly format. Data are stored in GPU Vertex arrays or Vertex Buffer Objects and returned to the client through the QMM. In the open-source version of the platform, the Data Layer is composed of both standard tools such as Hadoop with HDFS, Apache Flume and HBase, and a bespoke Storage module which is based on a HBase Thrift server. Having separate Thrift server applications for the engine layer, data layer and HBase storage allows a high degree of flexibility.

2: Preliminary Evaluation Results

A first evaluation of the platform was held concurrently in Edinburgh and Barcelona in February 2016. Figure 2 shows one of the evaluation examples: a fluidised bed simulated using the discrete element method (DEM) coupled with computational fluid dynamics (CFD). The simulation model contained around 40,000 time-steps, 12,000 particles per time-step and more than 3,000 particle-particle and particle-wall contacts per time-step. Two different visualisation clients were used for the evaluation: GiD (CIMNE, 2016) and iFX (Fraunhofer IGD, 2016). A Tier-2 cluster at Edinburgh, EDDIE, was used for the evaluation. Participants in the evaluation were able to compute the temporal and spatial discrete data in order to project the data onto an underlying continuum mesh, and display quantities of interest such as magnitudes of particle velocity onto the continuum representation. This first Hadoop implementation of the discrete-to-continuum VQuery has shown excellent results in terms of its scalability and normalised speedup. This is a promising result for the future when the platform will be tested using much larger data sets, with the final objective being the ability to efficiently analyse and visualise hundreds of millions of both particles and time-steps.

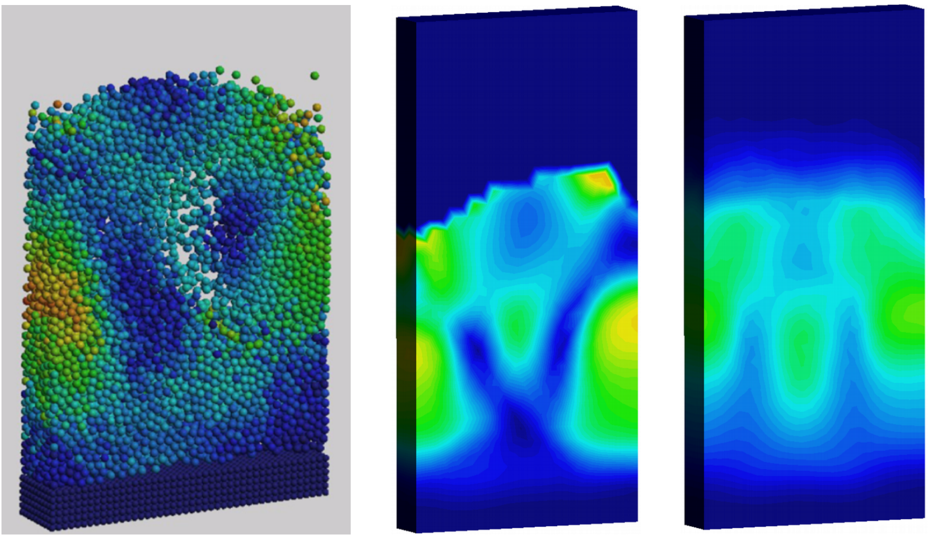

Figure 2: Example of the D2C query applied to fluidised bed data. The colours indicate modulus of velocity. The leftmost image shows discrete particle data at one time-step. The central image shows the result of spatial averaging only. The rightmost image shows spatial and temporal averaging over multiple time-steps.

ACKNOWLEDGEMENT

The development of the VELaSSCo platform is funded by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 619439.

REFERENCES

CIMNE (2016). GiD pre-post-processing system: www.gidhome.com

Fraunhofer IGD (2016). iFX post-processing framework: www.i-fx.net

Goldhirsch, I. (2010). Stress, stress asymmetry and couple stress: from discrete particles to continuum fields. Granular Matter, 12: 239–256

Labra, C., Ooi, J.Y. & Sun, J. (2013). Spatial and temporal coarse-graining for DEM analysis. AIP Conference Proc., 1542: 1258–1261, Powders and Grains 2013, Sydney, Australia

Weinhart , T., Labra, C., Luding, S. & Ooi, J.Y. (2015). Influence of coarse-graining parameters on the analysis of DEM simulations of silo flow. Powder Technology, doi:10.1016/j.powtec.2015.11.052