HiDALGO

CONTENTS

NEWSLETTER

Jan. 20, 2020 - Issue No. 2

HiDALGO

Welcome from the TCO
(Technical Coordinator)

Dear Colleagues,

The European society faces a number of pressing challenges in different areas ranging from sociology and economy to ecology and technology. Due to globalization processes (rapid advances in communication technologies, development of global labor market, etc), many of these challenges affect a huge number of people and, thus, are global in nature. Nowadays, none of the Global Challenges can be successfully solved based on the intuition of policy makers alone. HiDALGO aims to equip stakeholders with advanced solutions for computer-aided evidence-based policy making in the domain of Global Challenges.

The vivid examples of Global Challenges – explored by HiDALGO use cases – include modelling human migration from conflict zones, designing and developing smart cities, as well as reducing the impact of malicious information in social media. In HiDALGO, we build numerical models of Global Challenges and estimate consequences of different policies based on simulations of those models. In order to improve accuracy of simulations, we have been developing detailed social models coupled with such physical models of relevant phenomena, as well as different data sources including streaming data from sensors, mobile networks, social networks, etc. Because of the huge scale of the phenomena and associated amount of data, such an approach inevitably leads to highly-complex simulation workflows that are extremely demanding in storage, computing, and networking resources. It reveals a number of interesting scientific questions and raises technical challenges related to coupling the methods, scaling of applications with extreme computational requirements, handling streaming data on state-of-the-art HPC and HPDA systems, and integrating all components and infrastructures into a single stable user-friendly solution.

In this second newsletter, we discuss outcomes of the HiDALGO workshop on coupling methods and technologies held at ECMWF last summer, as well as give a brief description of the two keystone technical work packages of HiDALGO: WP3 addressing algorithms, methods, and tools, which will help us to resolve the challenges of running our applications on exascale HPC and HPDA systems, and WP6 dealing with development of Artificial Intelligence and Machine Learning components for our pilot applications. I hope you will enjoy this newsletter.

Sergiy Gogolenko
HLRS – High Performance Computing Center Stuttgart

HiDALGO

HiDALGO Coupling Workshop

Taking place during the second week of October 2019, the two-day HiDALGO Coupling Workshop gathered 12 participants from seven HiDALGO partners for technical discussions and planning of coupling tasks related to Work Package 3 (WP 3 - Exascale HPC and HPDA System Support) and Work Package 4 (WP4 - Pilot Applications). The main objective of the workshop was to assess the work which has been completed so far, identify problems and bottlenecks, and define a roadmap for future work.

The workshop focused on the theme of coupling within HiDALGO, which consists of coupling three pilot applications (human migration, urban air pollution, social networks analysis) with ECMWF & Copernicus EU data, telecommunications and traffic sensor data.

On the morning of the first day, Diana Suleimanova (BUL - Brunel University London), Zoltán Horváth (SZE - Szechenyi Istvan University), and Christine Gfrerer (PLUS - Paris Lodron University Salzburg) presented the current status of coupling within each pilot application. The future plans for each pilot application were also presented, followed by discussions regarding these plans.

This was followed by a session focusing on technical issues. Derek Groen (BUL) presented a status update of FabSim/FabFlee as part of the human migration simulation. Dineshkumar Rajagopal (HLRS) remotely presented an update of infrastructure work related to WP5. John Hanley (ECMWF - European Centre for Medium-Range Weather Forecasts) presented ECMWF's coupling activities, including an update on the weather and climate data (WCDA) Rest API which will enable easier access to ECMWF data through ECMWF's private cloud. Presentations were followed by discussions and the day ended with a dinner at the restaurant in Reading town centre.

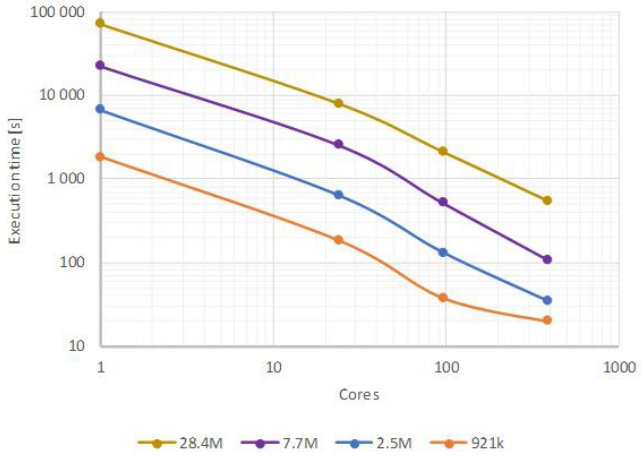
Day two started with a visit to ECMWF's super-computing and archive facilities. This was followed by a demonstration of the CKAN data management tool and a technical presentation by Marcin Lawenda (PSNC - Poznan Supercomputing and Network Center). This presentation stimulated a detailed discussion regarding data management within HiDALGO and issues related to implementation, performance and optimization.

The day ended with planning future work and drawing up a road map for the remaining two years of the project. Overall, the workshop was successful in meeting its objectives and it was agreed that technical workshops of this type focusing on particular topics of interest should be organized again in the future.

Milana Vuckovic ECMWF

HiDALGO

Exascale HPC and HPDA System Support



Execution time of the OpenFOAM Air Quality Dispersion Model with OpenFOAM v6 on Hazelhen, with a generated input mesh of various cell counts.

The work package 3 (WP 3) is designed to provide the necessary knowledge, algorithms, methods and tools for highly efficient processing and analysing data. The scope of the work covers identification of the challenges that prevents Global Challenges from achieving the highest possible performance (Exascale computation) as well as efficient and agile data manipulation. The work package is divided into six tasks covering specific areas of efficient processing. Emphasis is put on software and hardware aspects, data analytics, visualization and data management with focus on security mechanisms.

The primary goal of WP3 is to support other work packages in providing the most efficient software solutions possible. This is achieved in two different ways: by benchmarking existing developments at different perspectives (scalability, performance, data throughput, bottlenecks) and delivering revised solutions to overcome identified congestion and achieve better yield. The focal point is on simulation application originating from three different pilots: Migration, Urban Air Pollution (UAP) and Social Networks. So far the performance of most of them was improved either by adaptation of the parallelization scheme or changing the CFD simulation tool to be more efficient.

The massive data generated by Global Challenges applications require appropriate and highly scalable analytics software routines, to process data in time to be handled or provisioned to other applications. For this purpose frameworks and applications facilitating Big Data exploitation already available on the market are explored and analysed. In order to deal with large data volume, velocity, variety, but also their veracity and value standard analytics processing needs to be enhanced to guarantee efficient pre- and post-processing, coupled with secure capabilities. Efficient data manipulation is the vital aspect of the HiDALGO system and requires mastering, managing, curating and stewarding of data in the way streamlined to use case needs. That's why one of the biggest challenges here is to develop data management system which, on the one hand, addresses the demands related to processing large volumes of data and, on the ohter hand, use case specific requirements oriented towards efficient exploration and presentation of data.

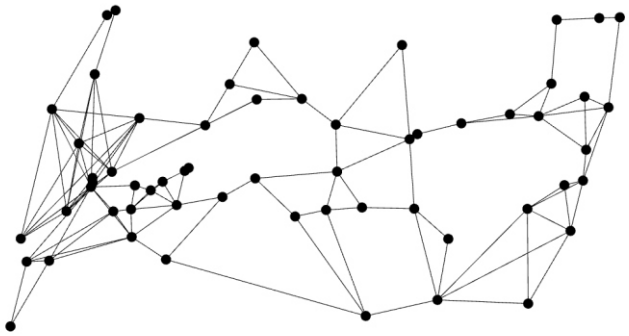
Another challenge developers are facing is the implementation of an efficient method for visualising data while taking into account their extremely large size. The environments envisioned need to be usable by experts and analysts to help in the detailed visual analysis of complex data sets, but must also be accessible to decision makers and politicians who spend more time working face-to-face than in front of a desktop computer, and to the everyday public in order to better involve them in the decision making process.

Coupling all the abovementioned modules to work together in efficient ways is a challenge in itself. Different technologies are used to combine existing applications to make them work together for an overarching purpose. Coupling technologies are a staple in the multiscale and hybrid simulation approaches, and a range of generic technologies have emerged in recent years, each with their unique added values.

Marcin Lawenda PSNC

HiDALGO

Artificial Intelligence in HiDALGO



Triangle inequality pruned real world road network in central Mali

In our world, Artificial Intelligence (AI) is indispensable and has a significant impact on our daily life. For the HiDALGO endeavour, it is used to approach solutions for the three major global challenges Migration, Urban Air Pollution, and the spread of information in Social Networks. One key goal is to use the power of AI and Machine Learning to assist these three use cases in various tasks, like feature extraction, feature engineering and model reduction.

Within WP6, we are currently working on two use cases in the HiDALGO project. The Social Networks Use Case has the goal to model, simulate, and analyse the spread of messages in social networks like Twitter and to understand their influence on our economic and social behaviour. In the Migration Use Case, the goal is to provide a simulation, and a data analysis tool to forecast refugee movement in times of conflict and, to support decision makers on where to provide humanitarian resources, food, and infrastructure.

Artificial Intelligence in the Social Networks Use Case is centered around the estimation of retweet probabilities for a tweet posted by a user. The idea is to create a model that classifies a tweet whether it is likely or unlikely to be retweeted. In a further step, the goal is to estimate the individual retweet probabilities of a user's followers for a tweet. Three small datasets with the topics „Migration“, „Fluechtlinge“ (refugees) and „vegan“ were provided by the use case leaders for an initial analysis. These datasets are used to extract features for the modelling of retweet probabilities. Hashtags in a tweet, the number of days a user account exists and the number of followers a user has are among the most influential features for a tweet to be retweeted. More than 20 features were chosen from the analysis and used to model the retweet probabilities. Two methods were considered to classify a tweet into one of the two target classes, namely a logistic regression model and a tree-based method. For the logistic regression model, a recursive feature elimination technique is utilised to eliminate features that only have a minor influence on retweetability, such as the time a tweet is posted. The model is then trained with the remaining features. For a new tweet, logistic regression returns one of the two categories, likely to be retweeted and unlikely to be retweeted. It is also possible to access the number (between 0 and 1) that the model returns and leads to the decision. With this number, it is planned to estimate the individual retweet probability for each follower of a user, which eventually can be used to simulate the message flow in the network. A decision tree also classifies a new tweet into one of the two target classes. In the training phase it returns threshold values for the features, which are then used to classify a new tweet.

For the Migration Use Case, a graph structure with the physical locations of the region of interest is needed. Artificial Intelligence is used to perform automated extraction of this location graph. The method is based on two steps, (1) extraction of a list of relevant locations either within a country or within a pre-defined rectangular region specified by geo-coordinates and (2) the computation of walking and driving distances between any of these locations. Together with the use case leader, we are testing and benchmarking several heuristics to find the optimal trade-off between computational complexity and accuracy of the result.

Christoph Schweimer Know-Center GmbH

HiDALGO

Partner Institutions

Events HiDALGO participated in

29/11/2019 in Brussels, Belgium **Conference on modelling for policy support: experiences, challenges and the way ahead** Participant: SZE	14/10/2019 in Helsinki, Finland **EBDVF** Participant: ATOS
9/10/2019 in Stuttgart, Germany **30th Workshop on Sustained Simulation Performance** Participant: USTUTT	08/10/2019 in Brussels, Belgium **European HPC Training Stakeholder Workshop** Participant: USTUTT
29/9/2019 in Bucharest, Romania **FOSS4G** Participant: ECMWF	23/9/2019 in Waukesha, WI, USA **GEHC First Global Data Science Symposium** Participant: SZE
9/9/2019 in Copenhagen, Denmark **European Conference for Applied Meteorology and Climatology 2019** Participant: ECMWF	28/8/2019 in Graz, Austria **4th Workshop on Model Reduction of Complex Dynamical Systems - MODRED 2019** Participant: KNOW
16/7/2019 in Adama, Ethiopia **First Flee training workshop** Participant: UBRU	12/6/2019 in Bruyères-le-Châtel, France **TERATEC 2019 Forum** Participant: ATOS, HLRS

HiDALGO workshop organized at

HiPEAC 2020
20-22 January 2020
Bologna, Italy

Project #824115 funded by the Horizon 2020 Framework Programme of the European Union