The newsletter of EPCC, the supercomputing centre at the University of Edinburgh

epcc news

Issue 86 Autumn 2019

Building on our history of excellence

In this issue

From the Director

Welcome to the Autumn 2019 edition of EPCC News. EPCC has had a very busy year in our new home at the Bayes Centre in the centre of Edinburgh.

I'm very pleased that we made the move – it's opened up a wide variety of new project opportunities, many as part of the City Deal's Data Driven Innovation Programme, and often with partners in the central area of the University. For significant parts of the past year we've been running over 100 concurrent projects across all of our supercomputing and data science activities – a record for EPCC.

Bringing the DDI programme to life has been a key part of our activities in 2019. We're still in the start-up phase of the World Class Data Infrastructure but we're making great progress with both the new building at our ACF data centre and the core IT infrastructure. A key focus of this infrastructure will be our services for research using public data – often deidentified unconsented data – which we have made a particular specialism over the past five years. In what has been a long and very complex project we're also making excellent progress in hosting the national medical image archive for research, which you can read about in this issue.

We are still fully engaged in Europe. EPCC is and will remain a European organisation irrespective of political developments. The NEXTGenIO project which has recently ended very successfully is evidence of how working together with collaborators from across Europe can lead to world-leading scientific results.

I hope you enjoy this EPCC News and please do get in touch if you have suggestions for future issues.

Mark Parsons
EPCC Director
m.parsons@epcc.ed.ac.uk

EPCC at SC19

SC19 Denver, CO hpc is now.

Look out for us at Booth 1701!

We have a busy schedule of activities this year – see the EPCC website for the full programme: http://bit.ly/2N87iAf

We will also be unveiling our new exhibition stand. The photograph shows one of the arches during the test build in the Bayes Centre.

Contact us

www.epcc.ed.ac.uk info@epcc.ed.ac.uk +44 (0)131 650 5030

EPCC is a supercomputing centre based at The University of Edinburgh, which is a charitable body registered in Scotland with registration number SC005336.

Contents

New staff at EPCC

Dr Holly Judge
Applications Developer in High Performance Computing

Dr William Lucas
Applications Consultant in High Performance Computing

Dr Julien Sindt
Applications Developer in High Performance Computing

NVIDIA AI Technology Centre UK

The NVIDIA AI Technology Centre is a programme designed to support ground-breaking research in AI and data science, and foster engagement across the higher education and research community.

The first NVAITC opened in Singapore in 2015 supporting collaborative projects throughout the Asia Pacific region. The UK instance, which was launched in 2019 with EPCC, the University of Reading and STFC Hartree as inaugural members, is one of the first in Europe. The NVAITC UK provides a collaborative community for world-class talent driving AI adoption and excellence across the UK. Centre members receive access to NVIDIA expertise and resources to support their projects, and in return NVIDIA gets the opportunity to participate in cutting-edge research and contribute to enhancing the national AI capability and expertise.

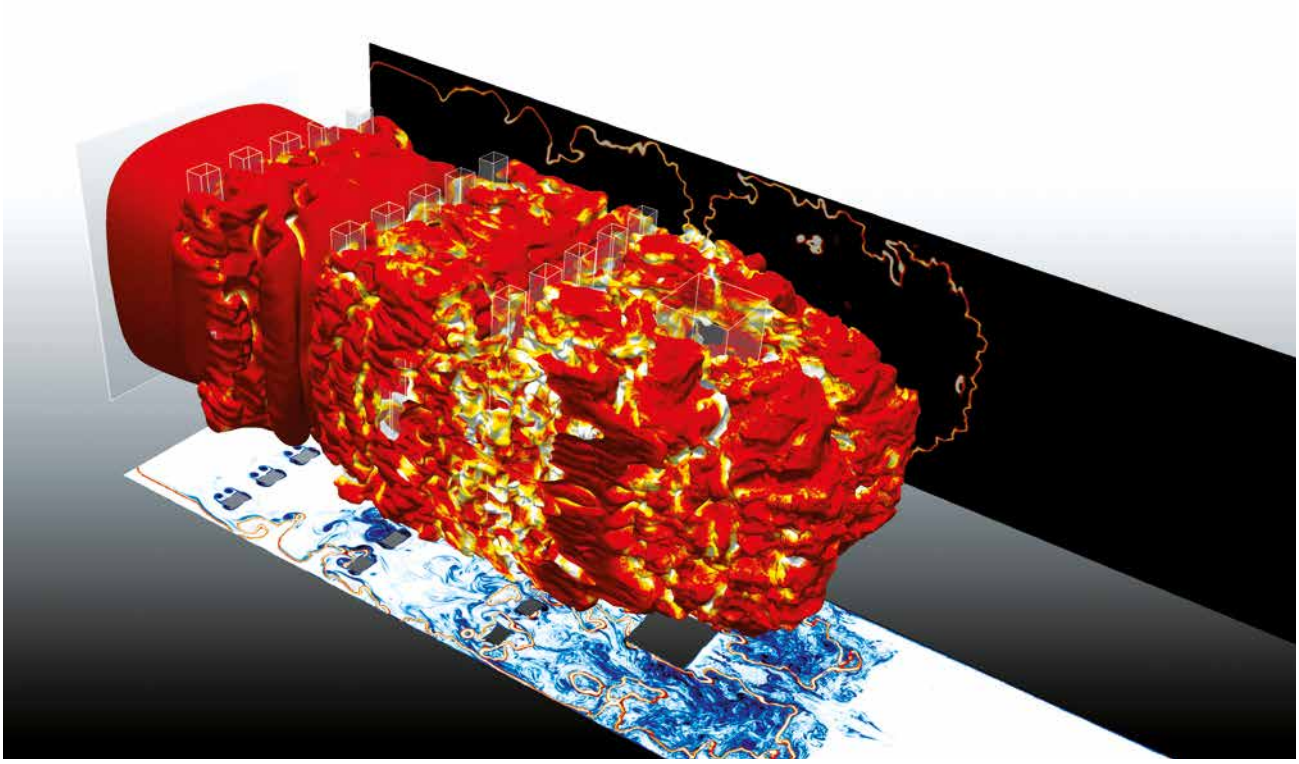
Increasingly we are seeing a convergence of HPC and AI in areas such as climate modelling, drug discovery and particle physics, where the successful application of the approach has led to game-changing improvements in the performance and capabilities of the simulations. The project with EPCC is a fantastic opportunity to explore this exciting area of research – in particular, using AI techniques and mixed precision compute to reduce time to solution for numerical simulations.

NVIDIA GPUs have hardware support for trans-precision computation, in particular via the Tensor Cores – dedicated cores on the GPU able to perform many matrix multiply and accumulate operations per clock cycle at a variety of precisions. With support from NVIDIA's Solution Architect team in both AI techniques and optimal use of the hardware, the application of EPCC's performance engineering expertise will aim to deliver an innovative new approach to pre-conditioning and solving large systems of linear equations.

Paul Graham
Senior Solutions Architect, NVIDIA
pgraham@nvidia.com

EPCC has a long history of supercomputing expertise, including recent research around tackling pressing exascale challenges with novel technologies. With a dedicated team of highly experienced HPC researchers and research software engineers, it is an ideal partner for this project.

www.nvidia.com



Large Eddy Simulation in a venting chamber of an explosion using the AVBP solver (see Vermorel, O. et al (2017), Combustion and Flame, 183 (September), pp. 207-223).

Image: CERFACS.

The European Centre of Excellence for Engineering Applications

The European Centre of Excellence for Engineering Applications (EXCELLERAT) has spent its first year identifying the major challenges that must be overcome to reach Exascale performance for a set of reference applications.

There are great opportunities for engineering applications in exploiting exascale computing (that is, more than 10^{18} operations per second). This will be the case not only in academia but also industry: engineering is seen as having the biggest potential for exascale in the industrial field. EXCELLERAT brings together expertise in HPC and engineering software to support the development of applications that can exploit computing capability at this level.

In addition to developing and optimising the reference applications, EXCELLERAT will be looking at generic HPC tools and technologies. Many of the planned developments are in typical HPC technologies (for example in the areas of new solver and numerical methods, and new mesh generation and adaptation techniques). In addition the Centre is looking at data management tools and how to adapt applications to work effectively on emerging computing architectures.

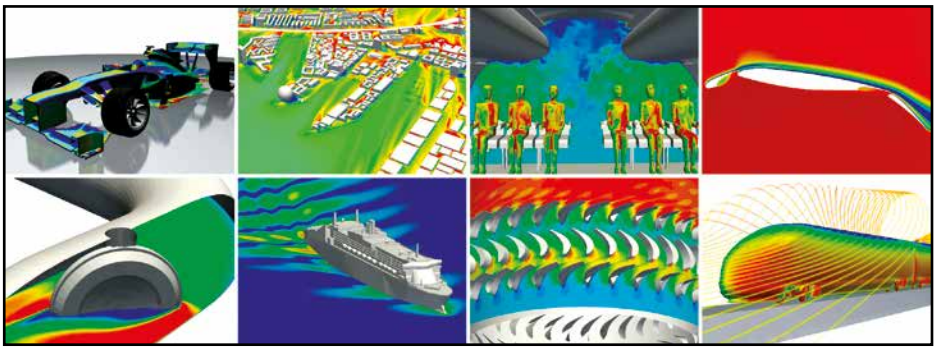
EXCELLERAT will shortly launch its first set of services to the engineering software community that will support developers and users on the road to exascale and beyond.

Mark Sawyer, EPCC
m.sawyer@epcc.ed.ac.uk

Community building

Outreach is an important activity of the Centre. Training activities are planned to enable users and developers to work more productively with HPC. We are also setting up an industry forum to allow a two-way exchange with end-users so that we can meet the objective of improving European competitiveness.

www.excellerat.eu



Solve CFD problems from your desktop

ENGYS is a global computer software company specialising in the development and distribution of computational fluid dynamics (CFD) and design optimisation solutions for enterprises based on open-source technologies. The company's software products include HELYX, a general-purpose CFD and design optimisation package for engineering applications, and ELEMENTS, a dedicated computer simulation software for automotive design.

As part of ENGYS' expansion plans to take advantage of advances in cloud and high-performance computing, in 2017 the company introduced a client-server framework across all its software products. The client-server enables users to connect interactively from their desktop to access any remote machine running the HELYX or ELEMENTS server. The seamless connection allows for setting up and solving CFD problems from scratch using powerful HPC resources on demand to achieve complex flow solutions beyond the installed hardware capacity of the user's local workstation.

With the introduction of the client-server, ENGYS approached EPCC to deploy its CFD server on Cirrus and test the cluster as a viable platform for remote computing.

With the exceptional support received from EPCC's technical team, a new SaaS offering was introduced by ENGYS to provide on-demand access to HELYX and ELEMENTS via Cirrus.

Besides the simplicity of running the software interactively on Cirrus from a desktop computer, existing customers also gained the ability to access over 10,000 computer cores to solve extremely large and complex CFD cases at a very reasonable rate per core-hour. Usage is fully monitored using the SAFE interface hosted by EPCC to provide all customers with complete control over their expenditure in real time. Security is also guaranteed by giving each customer their own protected space in the cluster and by establishing all client-server communications with Cirrus using fully encrypted SSH connections.

The open-source nature of the CFD products developed by ENGYS facilitated enormously the deployment of the software on Cirrus and the creation of the subsequent on-demand software service. The absence of complex licensing schemes and the ability to compile the source code directly in the cluster made it possible to achieve a reliable solution in a very reasonable timeframe.

Francisco Campos, ENGYS
f.campos@engys.com

Cirrus is an EPSRC Tier-2 National HPC Facility and is available for academic and industrial use. Housed in EPCC's Advanced Computing Facility, Cirrus is a flexible, state-of-the-art high performance computing system that provides an ideal platform for solving computational, simulation, modelling, and data science challenges. Users can run their own codes as well as accessing a range of commercial software tools on an on-demand, pay-per-use basis. These are used to solve CFD and FEA simulation and modelling problems in sectors such as automotive, aerospace, energy, oil and gas, general engineering, life sciences and financial services.

To apply for time on Cirrus, contact Thomas Blyth at EPCC:
t.blyth@epcc.ed.ac.uk
+44 (0) 131 651 3461

ENGYS offers HELYX and ELEMENTS on demand to all existing software customers with a valid licence for either software product. See: https://engys.com/

The World-Class Data Infrastructure: Phase 1

The WCDI will combine computing and data resources to create a novel facility that will be a fundamental enabler for innovation.

Rob Baxter, EPCC
r.baxter@epcc.ed.ac.uk

Launched at the end of 2018, the Data Driven Innovation Programme (DDI) is one of the six programmes funded within the Edinburgh & South-East Scotland City Region Deal.

The DDI Programme's aim is to make Edinburgh the "Data Capital of Europe", with ambitious targets to support, enhance and improve talent, research, commercial adoption and entrepreneurship across the region, through better use of data.

The Programme targets ten industry sectors, with interactions managed through five DDI Hubs (see opposite). The activities of these Hubs are underpinned by a World-Class Data Infrastructure.

You can think of the WCDI as a layer of storage and computing services presented as a private cloud and hosting a rich and growing collection of data. WCDI supports the long-term storage and curation of data assets, and their cataloguing, preparation and presentation as analytic-ready datasets for research and innovation. It offers a range of computing services, from web-based notebooks to rich desktop environments and seamless access to high-performance computing.

WCDI supports learners, researchers and innovators across the spectrum, with services from basic data download through simple learn-as-you-play-with-data notebooks to full-throated, GPU-enabled machine learning platforms for driving AI application development.

WCDI also provides safe haven services to health and government users, following best practice in independent governance and supporting the linkage of complex personal data for public benefit research and policy-making under national and regional safeguards. Safe haven services can also be created for organisations wishing to host and govern access to their data assets in a highly secure environment. Safe havens are isolated from the rest of WCDI, with user approvals, data ingress and egress and permitted software all controlled by information governance bodies independent of the infrastructure itself.

Are we nearly there yet?

For this year and next, WCDI Phase I is all about development and co-design. Working with a number of key stakeholders and early adopters we are putting together the core elements of WCDI – including a new home in EPCC's Advanced Computing Facility (ACF) Computer Room 4, due for completion in June 2020 (see p16). This is a high-resilience, state-of-

WCDI will grow and mature with the DDI Programme, expanding in capacity and capability, responding to the needs of the innovation Hubs and, through them, to learners, researchers, innovators and entrepreneurs from across the region and beyond.

your_photo/Getty Images

the-art facility with enhanced power and network connectivity to provide the long-term durability needed for a data archive.

Meanwhile, the first hardware platform is in place in temporary accommodation elsewhere in the ACF. We already have 10 petabytes (PB) of new disk capacity, split 50/50 between the safe haven and non-safe haven sides of WCDI, and planned to increase to 12 PB by July 2020. We also have a small slice of the compute and service cloud in place, around 30 virtual machines plus four NVIDIA V100 GPUs, plus access to our Cirrus HPC service.

This initial development system is helping us shape building blocks for the future WCDI. We are laying down an architecture that will enable us to scale out storage or compute as needed, and putting together the supporting software layers for a variety of use-cases from our early adopters.

Our partners in Phase 1 include the iCAIRD digital pathology archive (see p10), the National Collection of Aerial Photography, Health Data Research UK, the Administrative Data Research Partnership, the Paracrawl Internet Archive project, Albyn Housing Society, NHS Scotland, the Scottish Government, the Edinburgh Festival Fringe, SAS, the DataLoch (see p9) and Local Authorities from across the region.

But what does it do?

By Spring next year we expect the first WCDI services to come online. They will be headed by a data catalogue, a point-of-first-contact for WCDI, incorporating an open metadata repository, easy access to open data and an approvals system supporting work with, and access to, restricted data. We also expect to launch browser-based "notebook" services using Jupyter and RStudio to support data analysis using Python and R.

Around the middle of 2020 we'll launch our virtual machine services offering fully kitted-out desktops to support different kinds of data-orientated task: data analysis machines for statistical analysis with R, Python etc; a data science flavour for machine learning and data modelling (with or without GPU capability); and a data engineering variety for data flow software builders, with Spark, Scala, Kafka etc. We'll also be rolling out browser-based access to the Bayes Centre's state-of-the-art SAS Viya platform.

It's an ambitious programme. The next twelve months will be busy – and yet this is only the first step along a very exciting road for EPCC, the University and the whole region.

DDI Data-Driven Innovation

The DDI Hubs for Innovation

- Bayes Centre: Digital technology and space & satellite applications
- Usher Institute: Health and social care
- National Robotarium: Robotics and autonomous systems
- Easter Bush: Agricultural technology
- Edinburgh Futures Institute: A wide remit covering financial services, financial technology, creative industries, festivals and tourism, and the public sector.

Read more about data-driven innovation at EPCC www.epcc.ed.ac.uk/research/data

georgeclerk/Getty Images

Festival Mobility: decisions with data

Those of us who live in Edinburgh during Festival time know that, when dealing with so many people, things don't always work as well as we'd like. Our commutes may be longer. We might have to stand on the train or walk in the bus lane to bypass the crowds. Edinburgh's festivals are the envy of the world but can make life more difficult for many of it citizens.

How can we make it work better?

The Edinburgh Futures Institute, The List, EPCC, and Transport for Edinburgh have established a pilot to try to bring together event and transport datasets to better understand festival-related traffic and congestion during August. The pilot will look at how we might design a data 'product' that transport providers, city authorities and festivals could collectively explore to help them plan for the next year's festivals. For example should they increase the number of bus services, and if so how could they do it without contributing to congestion? Which roads should be closed and when?

Data-driven optimisation is difficult due to the complexity and variety of mobility data. Trajectories of individuals in motion, a multimodal transportation network, and the various types, sizes and unknown popularity of event venues all contribute to mobility problems. Data doesn't hold all of the answers, but we think that shared data, shared insights and shared decisions across the varied organisations involved with making the festivals work might help us make some marginal gains.

This is just the start of our pilot project. We can't be sure that we will come up with useful insights but it's certainly worth a try. After all, the Data-Driven Innovation (DDI) Programme was set up to use data for the benefit of our citizens.

The DDI Programme aims to help the Edinburgh and South East Scotland Region claim the title of 'Data Capital of Europe'. It will go nicely with our other famous title: the world's Festival City.

The pilot is funded by Transport Scotland, Scottish Enterprise, and the University of Edinburgh's Data-Driven Innovation Programme.

Joshua Ryan-Saha
Data-Driven Innovation Lead – Tourism & Festivals
University of Edinburgh
Joshua.Ryan-Saha@ed.ac.uk

If you've lived in Edinburgh during August, you've witnessed a city transformed. Together, Edinburgh's August festivals represent an annual event close to the size of the Olympic Games. It takes a gargantuan effort by everyone involved with running the city and they do a great job.

Data Driven Innovation Programme
https://ddi.ac.uk

A data-driven approach to public health

Wavebreakmedia/Getty Images

EPCC and other partners at the University of Edinburgh have commenced work on a new programme to develop DataLoch, a data repository for all local, regional and national health and social care data for residents of the Edinburgh & South East Scotland region. DataLoch and the associated Data Driven Innovation team will drive research and innovation, improve patient care, and reduce health inequalities across the region.

Scotland is unique in having a single healthcare provider and a health system consisting of world-leading linked national healthcare data assets from birth to death. In the East of Scotland there are excellent examples of these data assets being harnessed to benefit patients, the health service, and the general population. But these are built on specific research projects, and the current approach to data access and linkage is fragmented, inefficient and often limited to specific conditions.

The DataLoch programme aims to streamline this process by creating a comprehensive and integrated health and social care data ecosystem in partnership with the regional NHS Boards of Lothian, Borders, and Fife, allied to national and local government, and partners across the social care spectrum. Data will be securely hosted by the World Class Data Infrastructure (see page 6), and will be defined, linked and accessible via an online portal for accredited researchers. Users of the DataLoch will have access to accurate and current data on all aspects of health and social care for the population, for any condition, clinical pathway, or service across the City Region.

The programme will initially be driven by a number of exemplar projects. One of these exemplars seeks to improve the care of patients with multimorbidity – two or more chronic conditions at the same time. A data driven approach will help clinicians recognise the impact of multimorbidity on the safety and effectiveness of treatments and interventions, and will support a more personalised approach to patient management. The aim will be to improve the quality and delivery of care through innovative use of healthcare data.

Professor Stewart Mercer, Chair of Primary Care and Multimorbidity, Usher Institute, said: “Multi-morbidity is now the norm not the exception in people with chronic conditions, and health services need to respond to this by putting the patient and their choices at the centre of care.”

As the bedrock of the health and social care Data Driven Innovation programme, DataLoch enables collaboration and innovation across the public, private and third sectors. DataLoch promotes a data-driven approach to prevention, treatment and service delivery, and will enable the development of a world-leading learning health system in our region.

Ally Hume, EPCC
a.hume@epcc.ed.ac.uk

“Support from the City Region Deal and DDI programme will enable us to create a unique platform and service that aligns the needs of both our research community and our health and social care providers. DataLoch will encourage new partnerships to help us achieve a common goal of improving the health and care of the citizens of our region and beyond.”

Prof. Nicholas L Mills, Senior Responsible Officer for the Health and Social Care DDI Hub

The Industrial Centre for Artificial Intelligence Research in Digital Diagnostics

The iCAIRD project is working to establish a world-class centre of excellence in the application of artificial intelligence to digital diagnostics. The intention is that iCAIRD will allow clinicians, health planners and industry to work together, enabling research-active clinicians to collaborate with innovative SMEs to better inform clinical questions, and ultimately to solve healthcare challenges more quickly and efficiently.

Andrew Brooks, EPCC
a.brooks@epcc.ed.ac.uk

iCAIRD was announced last November as one of five successful bids to the UK Government's Industrial Strategy Challenge Fund (ISCF). It brings together a pan-Scotland collaboration of 15 partners from across industry, the NHS, and academia. Industry leadership will be provided by Canon Medical Research (radiology) and Royal Philips (digital pathology). iCAIRD has been awarded £10m by Innovate UK, while partner companies Canon Medical Research Europe Ltd and Royal Philips will provide £5m of additional supportive funding, along with seven actively-engaged SMEs.

The key iCAIRD priorities are to:

- Create infrastructure to develop and apply AI in digital diagnostics, pathology and radiology.
- Fast-track digitisation of Scottish NHS pathology data to create the largest fully digital pathology laboratory network in Europe.
- Work with partner Canon to develop a network of Safe-Haven Artificial Intelligence Platforms (SHAIP), thus allowing crucially important research while ensuring protection of patient data.
- Work with partner Philips to establish a Health Data Research UK national pathology image archive of anonymised disease cases within the National Data Safe Haven that can be used to train computers to augment and improve on current practice.

Some of the first applications will be to validate AI in stroke medicine, chest X-ray triage and mammogram interpretation, and apply AI to colon cancer data and gynaecological pathology.

One of EPCC's roles within iCAIRD is hosting the National Data Safe Haven and the national archive of digital pathology images. The archive is expected to grow to a substantial amount, in the order of many hundreds of terabytes.

We will work with Glencoe Software to deploy its OMERO Plus database system, which is ideally suited to managing such a vast collection. Users will be able to import, catalogue and view pathology images, add metadata, draw figures, collaborate with other users and groups, and analyse in a variety of ways. The other role for EPCC is to provide the technology for the application of AI to digital diagnostics, pathology and radiology. This will bring together our expertise in image archives, high performance computing, and data science.

iCAIRD is centred at the Imaging Centre of Excellence, part of the University of Glasgow's Clinical Innovation Zone at Queen Elizabeth University Hospital. The academic partners are the universities of Edinburgh, St Andrews, Aberdeen and Glasgow. It will integrate with Health Data Research UK (HDRUK) and the national Picture Archiving Communication System (PACS), working closely with partners across Scotland and beyond.

www.icaird.com

Predicting disease with artificial intelligence

ChooChin/Getty Images

EPCC is part of a £4.4 million project to turn a database of millions of clinical images into a powerful research tool to help tackle health conditions including lung cancer and dementia.

The PICTURES project will make use of the approximately 30 million NHS images collected since 2006, employing artificial intelligence (AI) to search for 'warning signs' in the images which predict the development of diseases. This will allow doctors in the future to make use of this information in routine care, greatly enhancing the clinical utility of routine scans.

The project has been funded by the Medical Research Council (MRC), Engineering and Physical Sciences Research Council (EPSRC) and industry partners to develop the technology that will unlock the huge additional potential of these images.

Dr Emily Jefferson, Director of the Health Informatics Centre at Dundee, said, "Clinical images are now core diagnostic technologies. These images can support many important areas of research to improve any or all diagnosis, monitoring of disease progression and response to treatment.

"Access to the vast bank of 'real world' images can offer a huge boost to research into major diseases and conditions, and that is what we are looking to develop through the PICTURES study, initially using lung cancer and dementia as exemplar projects."

One of these exemplar projects will develop a method to detect warning signs of coronary artery disease and lung cancer using AI to check patients' CT chest scans. This project will work in partnership with international experts from an industry partner, Aidence, to convert the research into a clinical tool which can be used to support doctors on the front line in the NHS.

Professor Edwin Van Beek of the University of Edinburgh said, "It is very exciting to be able to develop AI tools to enhance the diagnostic potential of CT scans in the chest and MRI scans of the brain, which currently don't routinely assess cardiac or dementia risk in these patients. By having these tools provide information to clinicians, earlier treatment and management changes will improve health outcomes in the future."

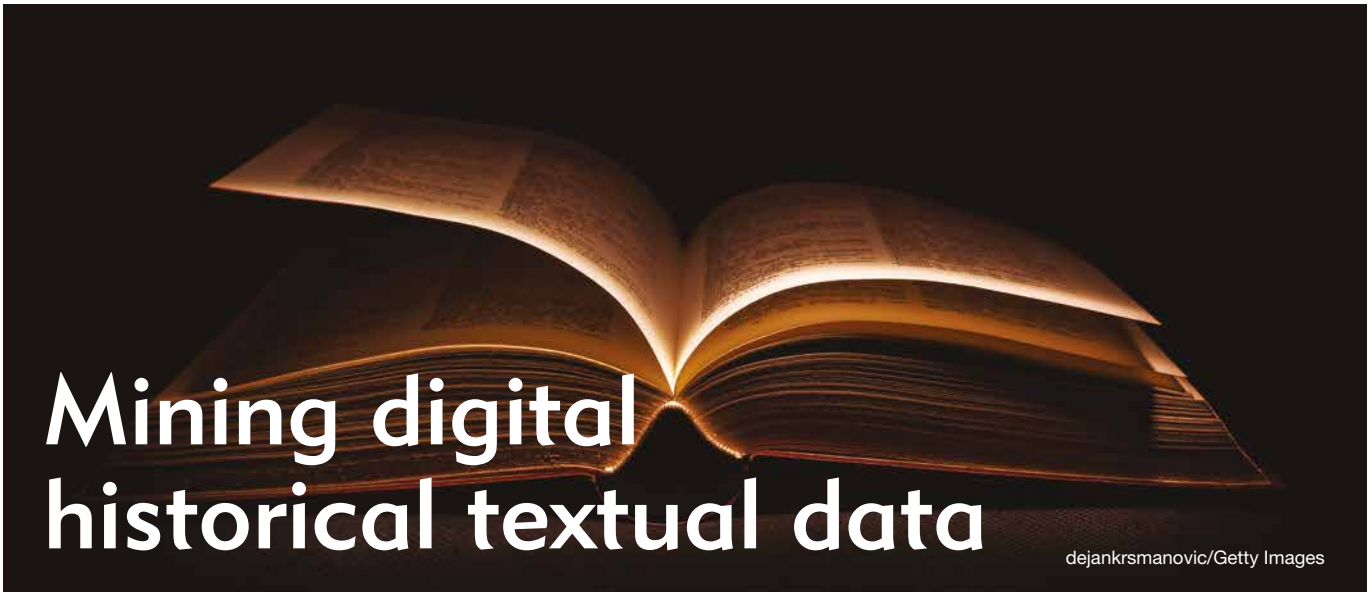
The project follows on from previous work to provide access to these images through the Scottish National Safe Haven that is expected to go live later this year. The PICTURES project enables us to carry out the research that will continue to enhance this service over the next five years.

PICTURES is a collaboration between the University of Edinburgh, Dundee University, Abertay University and NHS Scotland.

Ally Hume, EPCC
a.hume@epcc.ed.ac.uk

Each year millions of clinical images such as X-rays, CT, MRI, ultrasound, nuclear medicine, and retinal images are generated by the NHS in Scotland and stored in the national imaging database. In addition to containing important clinical information, these images also potentially contain a great deal of information about the health of the individual which is not currently made use of in health care.

PICTURES



Mining digital historical textual data

dejankrsmanovic/Getty Images

Over the last three decades the collections of libraries, archives and museums have been transformed by large-scale digitisation. The volume and quality of available digitised text makes searching and linking these data now feasible, for example the extensive digital collection of the National Library of Scotland (NLS) can be accessed online and also downloaded for further digital humanities research.

At EPCC we have been working across several projects to mitigate some of the obstacles to mining historical texts at scale by providing a new Apache Spark-based text mining toolbox. Called defoe, it enables parallel text analyses across large collections, has a rich set of text mining queries, includes NLP pre-processing techniques to mitigate OCR errors, and can be easily installed.

For the NLS digital collection, we have collaborated with Prof. Melissa Terras (College of Arts, Humanities, and Social Sciences, University of Edinburgh) to run text analyses across the first eight editions of Encyclopaedia Britannica, which comprise a total of 143 volumes in ALTO and METS formats. As defoe already supported both schemas, only a few changes were needed to be able to ingest the data.

After downloading the full available dataset, we started by exploring the popularity of different topics over time, such as sports, Scottish cities or Scottish philosophers.

We used defoe's keysearch_by_year query, which counts the number of occurrences for single words or sentences, and groups the results by year. For normalising the frequencies of previous results, we also need to run the normalise query, which iterates through the Encyclopaedia's editions and counts the number of documents, pages and words per year. The normalised frequencies of topics are displayed later using n-grams.

We are currently working on a new text-mining platform to allow non-computational experts to select and run large-scale text mining queries across different digital collections in a transparent way. With a simple user interface, researchers will be able to customise the query to run by selecting which corpus to use and some configuration parameters. defoe will be at the backend of this platform, receiving the query and its configuration parameters, running the query across the specified dataset, and returning the results to the platform. Different visualisations will be also offered for displaying results. In this way, we hope to make the platform more accessible to humanities researchers, so allowing them to access new text-mining facilities.

Rosa Filgueira, EPCC
r.filgueira@epcc.ed.ac.uk

" All this work provides the means to search across large-scale datasets and to return results for further analysis and interpretation by historians."
Melissa Terras, Professor of Digital Cultural Heritage, College of Arts, Humanities, and Social Sciences at the University of Edinburgh

For a more detailed description of this work, see:
http://bit.ly/33S61nN

National Library of Scotland Digital Resources
www.nls.uk/digital-resources

Reducing uncertainty in geological modelling

Cognitive Geology Ltd is an independent technology company based in Edinburgh, UK. It delivers innovative geological modelling software to the oil and gas industry, with the goal of improving efficiency and reducing uncertainty in geological modelling workflows and the business decisions which are based on them. EPCC has worked with the company to investigate ways to reduce large ensembles of geological models while maintaining the range of plausible scenarios described by the set.

Cognitive Geology's software, Hutton, allows the exploration of the thousands of possible ways sub-surface properties may be distributed through an oilfield, with all the generated scenarios being based on geological first principles. For example, deciding the production quality of a reservoir under a specific scenario is computationally challenging and, given the high number of possible scenarios, investigating all possibilities is simply not feasible. The collaboration between EPCC and Cognitive Geology looked at machine learning approaches to select a few scenarios that would truthfully represent the original ensemble.

During the first part of the collaboration we tested whether Cognitive Geology's novel reduced basis representation of a complex 3D geological model of the earth still contains enough information to derive properties of the sub-surface that are of interest to the petroleum industry. Using supervised learning techniques we confirmed that the relevant patterns can be learned from a rather small sample of 2000 geological models, which corresponds to less than 15% of the possible models.

The second part of the work looked at grouping the geological models based on their associated outcome to identify a representative subset. Such a grouping cannot rely on statistically significant similarities, as such an approach disregards rare and anomalous scenarios which in turn might lead to wrong decisions regarding the reservoir development. Clustering techniques generally allow identification of such outliers, but need a user to define a metric that describes similarity of different models. We have shown how the machine learning techniques developed in the previous step can be applied to this task. The choice of such a metric was further validated by experiments using various clustering algorithms and different metrics which demonstrated the importance of the metric over the specific algorithm.

The collaboration led to a reduction of 13824 models to 64 representative models, and this reduction was done based on a detailed evaluation of less than 15% of the original model ensemble. This work also provided new insights into the properties of the reduced basis representation used in this work as well as in Cognitive Geology's software.

Lucy MacGregor
Cognitive Geology
lucy.macgregor@cognitivegeology.com

Anna Roubíčková, EPCC
a.roubikova@epcc.ed.ac.uk

From EPCC's perspective, the project was a great opportunity to deepen our experience with oil and gas exploration thanks to excellent domain specific support from Cognitive Geology. The findings of this work provide a proof of concept for the techniques developed by Cognitive Geology and highlighted several new lines of enquiry to follow up in the future.

www.cognitivegeology.com

Using FPGAs to model the atmosphere

xxmmxx/Getty Images

The Met Office relies on some of the world's most computationally intensive codes and works to very tight time constraints. It is important to explore and understand any technology that can potentially accelerate its codes, ultimately modelling the atmosphere and forecasting the weather more rapidly.

To explore the potential role of Field Programmable Gate Arrays (FPGAs) in accelerating HPC codes, we focused on the advection kernel of an atmospheric model called MONC, a code that we developed with the Met Office a few years ago. Advection, which is the movement of quantities through the air due to wind, is one of the most computationally intensive parts of the model. We decided to port the advection kernel to FPGAs using High Level Synthesis (HLS), which is an approach where the programmer writes C or C++ code and the tooling synthesises it to the Register Transfer Level (RTL), which can be thought as the assembly code level of FPGAs. These are then exported as IP blocks, which can be dropped into designs called shells and hooked up to other components such as memory.

Initially I made good progress, developing the HLS code and associated shell in just a few days. But I soon realised that when writing FPGA code it is necessary to entirely rethink the algorithm and rewrite it from a data-flow perspective for even adequate performance. This can be a real challenge from a code perspective because, while HLS accepts something syntactically similar to C or C++, to get good performance the programmer must embrace the fact that the semantics are entirely different.

From redesigning how we load data into the kernel to how the computation is structured, the optimised FPGA code bears no resemblance to its CPU equivalent. However, the effort was rewarded: the optimised FPGA version goes over 800 times faster than our first attempt based directly on the CPU code! More generally, for all but the largest systems we out-perform an 18-core Broadwell CPU at significantly lower energy.

One challenge we face is, because the FPGA is mounted on a PCIe card, input data has to be transferred from the host to the on-board DRAM and the results copied back. For our largest problem this involved transferring over 12 GB of data, and even though we optimised it using an approach similar to CUDA streams, data transfer still accounts for around 40% of the runtime.

The investment in programming tools for FPGAs has been really positive and made them a realistic future HPC technology. However code cannot simply be copied from the CPU to the FPGA: problems must be completely rethought and algorithms recast in a data-flow style. But the potential performance and energy benefits are significant.

Nick Brown, EPCC
n.brown@epcc.ed.ac.uk

FPGAs provide a large number of configurable logic blocks sitting within a sea of configurable interconnect. It has recently become easier for developers to convert their algorithms to configure these fundamental components and so execute their HPC codes in hardware rather than software. This has significant potential benefits for both performance and energy usage, but as FPGAs are so different from CPUs or GPUs, a key challenge is how we design our algorithms to leverage them.

I will present a paper at SC19 on the data movement optimisations we have done as part of this work at the H2RC workshop on the Sunday morning. See: http://bit.ly/2N87iAf

EPCC's ARM system: comparing the performance of MPI implementations

MVAPICH is a high performance implementation of MPI. It is specialised for InfiniBand, Omni-Path, Ethernet/iWARP, and RoCE communication technologies, but people generally use the default module loaded on their system. This is important because, as HPC programmers, we often optimise our codes but overlook the potential performance gains of better choice of MPI implementation.

EPCC installed a new ARM-based system called Fulhame earlier this year. With 64 nodes, each containing two 32-core Marvell ThunderX2 CPUs (4096 cores in total), this system is funded under the Catalyst UK programme to further develop the ARM software ecosystem for HPC. MVAPICH's performance has been benchmarked and explored at length on traditional x86 HPC systems, but I was interested how it would perform on Fulhame.

Some very interesting patterns were highlighted and generally MVAPICH performed very competitively against the other implementations. For instance on Cirrus, our x86 system, it out-performed MPT consistently and in some cases very significantly. On Fulhame the performance patterns were more nuanced and complex. MVAPICH demonstrates significant performance benefits with 2D pencil decomposed FFT codes, which is because the AlltoAll collective significantly out performs OpenMPI or MPT. In other situations however OpenMPI or MPT performed slightly better, but OpenMPI must first be configured to select the correct communication protocol before the results come close to those of MVAPICH, which provides good performance out of the box.

The MVAPICH team at Ohio State University, USA, has been given access to Fulhame to explore the performance properties of MVAPICH on this system and ARM in general. As the default performance was really very good, I am excited to see what can be achieved with this tuning. I believe MVAPICH is a good choice for MPI implementation, and definitely worth using. Furthermore, the use of ARM in HPC is also a very exciting prospect and throughout my work with this system I found that not only did it perform well, but it also provided a stable and mature HPC ecosystem. I expect we will see significant adoption of ARM-based systems (likely with MVAPICH installed!) across the Top500 in the next few years.

Nick Brown, EPCC
n.brown@epcc.ed.ac.uk

ARM in HPC is still very new, and the MVAPICH team has not yet finished fully tuning their library to the architecture. This is very important because MVAPICH contains many advanced algorithms which suit different situations, and it is likely that the rules about which algorithm to use for ARM are in need of more refinement. Bearing this in mind, it is impressive that MVAPICH performs so well.

Building a home for the World-Class Data Infrastructure

Data-Driven Innovation

The World-Class Data Infrastructure (WCDI) is a core component of the Data Driven Innovation (DDI) Programme that forms an integral part of the Edinburgh and South East Scotland City Deal investment proposition.

The building currently under construction at EPCC's Advanced Computing Facility (ACF) will house a new High Resiliency Computer Room (cr4) and associated plant areas to host and support new computer equipment.

When complete, cr4 will extend the existing ACF, which already has a number of computer rooms, associated plant rooms and office space, and will be connected via a unique and impressive walkway.

The work to build cr4 started in June 2019, and is due to finish in summer 2020.

From an engineering perspective, the new facility will provide innovative cooling solutions to help cr4 to be one of the most efficient computer rooms possible, adding to the ACF's current portfolio of energy and cooling efficient rooms.

The WCDI will be underpinned by cutting-edge networking advancements, making the ACF and cr4 one of the fastest available. We will provide secure data hosting and analytics infrastructure, built on years of experience of running the National Safe Haven, to hold unique datasets and allow approved, authorised researchers and users to find innovative solutions to novel questions.

The room will be capable of hosting up to 270 cabinets, giving WCDI and the DDI programme a platform to grow over the next 10 years.

Paul Clark, EPCC
p.clark@epcc.ed.ac.uk

Building on existing excellence, our new high resiliency room will provide the state-of-the-art computing and data storage infrastructure needed to develop the technologies that will underpin data-driven innovation.

Optimising traffic handling in cloud environments for cyber security

Illuminate Technologies works with EPCC and Intel® on a collaborative project funded by Scottish Enterprise under the CodeOpt Scotland programme to actively seek out innovative performance-enhancing, packet-processing approaches within the type of Cloud environment expected to be seen in 5G deployments.

Illuminate Technologies provides innovative capabilities for cyber-threat detection, legal compliance for critical communications infrastructure, and public safety, focusing on telecommunications networks. This project aims to establish a realistic, carrier-grade, Cloud environment, using standard components such as the OpenStack Cloud operating system and state-of-the-art, performance-enhancing tools like the Data Plane Development Kit (DPDK). This will be used to establish a benchmark for data-traffic processing against which further enhancements can be measured. These enhancements will use emerging programmable networking approaches, optimised for a Cloud environment.

The partners in this project have gained valuable experience in overcoming challenges in setting up the benchmarking environment. The nature of these problems is to be expected in such a rapidly evolving technology space. This pace of evolution clearly needs to continue in order to deliver low-latency, high-capacity 5G services.

With its core Research & Development centre based in Edinburgh, Illuminate Technologies values the opportunity to collaborate in industry projects with leading companies and centres of technical excellence to develop a sustainable knowledge base in Scotland for innovative cyber security products and the CodeOpt Scotland programme brings in technical expertise from EPCC and Intel®. With the project currently at its early stages, EPCC has delivered a dedicated testbed for the project, with a customised version of OpenStack that explores advanced DPDK configurations. EPCC has also developed early versions of the code that will next be optimised.

Intel® and EPCC have collaborated on accelerating network traffic on the EPCC testbed and Intel® has secured access to suitable larger-scale testbeds for further experimentation at the later stages of the project. Numerous insights have been delivered to Illuminate Technologies in what is already a very close collaboration, and Illuminate Technologies has developed new know-how, including improved methods for synthesising realistic network traffic patterns. The project runs until January 2020.

CodeOpt Scotland

The project was part-funded by the CodeOpt Scotland programme, which was designed to increase company competitiveness by optimising the performance of business-critical software. CodeOpt Scotland is a partnership of EPCC, Scottish Enterprise and Intel®.

Jeremy Fuller, R&D Project Manager, Illuminate Technologies
contact_uk@illuminate.solutions

Kostas Kavoussanakis, EPCC
k.kavoussanakis@epcc.ed.ac.uk

"Illuminate Technologies is ensuring its solutions are well matched to the industry's shift towards Cloud technology and the roll out of new communications networks. This partnership project gives us a competitive edge by providing an opportunity to push technology boundaries and evolve our Edinburgh-based expertise for creating new products."
Roy Macnaughton
VP Engineering, Illuminate Technologies

"This collaboration with Illuminate Technologies and Intel® expands EPCC's world-class systems engineering expertise into the area of cybersecurity. We are now working to optimise the performance of the code, taking it beyond the current state of the art."
Mark Parsons
Director, EPCC

Launch of Society of Research Software Engineering

The Society of Research Software Engineering launched in September 2019. You can now become a member of the Society and, in doing so, help support software in research and the people who develop it.

The Society developed from the Research Software Engineer (RSE) movement that formed after discussions at the Software Sustainability Institute's 2012 Collaborations Workshop. It is the successor to the UK RSE Association but, unlike the informal Association, it is a legal entity that can hold funds and employ staff. Regardless of its more formal incarnation, the Society is still a grass-roots organisation run by Trustees elected from its membership.

The UK RSE Association was created in 2013 to recognise the work of software developers in academia and to improve the career structures available to them. Since then the movement has gone from strength to strength: encouraging and supporting the creation of sister organisations across the world, and organising an international RSE survey that provides insight into current working conditions and practices.

EPCC played a pivotal role in the RSE story. EPCC's Rob Baxter coined the term Research Software Engineer at the initial workshop, and EPCC has always provided a home and careers for RSEs. More recently, the Edinburgh Research Software Engineering Community has been formed to provide support for RSEs in Edinburgh and a forum to share best practice. EPCC's Andy Turner has served on the committee of the UK RSE Association and as an initial trustee of the Society of Research Software Engineering.

The launch of the Society took place at the inaugural Annual General Meeting, held during the Research Software Engineering Conference at the University of Birmingham. This included pitches from nominees, including most of the current Trustees, who wish to form the first elected committee to run the Society. Voting is now open for the Trustees who will steer the Society in its first year.

The Research Software Engineering community has grown from a handful of people in 2012 to an international community of thousands in 2019. This would not have been possible without the foresight and hard work of a number of people within the community and the continued support of the Software Sustainability Institute and the EPCC.

Neil Chue Hong, EPCC
n.chuehong@epcc.ed.ac.uk

Simon Hettrick, Software Sustainability Institute
s.hettrick@software.ac.uk

The Society is a truly independent organisation with a single focus: supporting the UK's research software engineering community.

RSE
SOCIETY OF RESEARCH SOFTWARE ENGINEERING

https://society-rse.org/

"An excellent way to have international collaborations with excellent scientists"

The HPC-Europa3 programme funds collaborative visits of up to three months for researchers at any level, in any discipline, who need access to HPC resources.

HPC-Europa3 has now been welcoming visitors to EPCC – and another eight HPC centres around Europe – for nearly two years.

The successful applicants to date are from a total of 43 countries. Priority is given to researchers working in the EU and Associated States, but limited places are available for researchers working elsewhere who can make a strong case for funding.

There is a roughly equal split among the visitors between post-graduates, postdocs, and experienced researchers, and a wide variety of disciplines are represented.

We have been pleased to see that, compared to previous programmes, there has been an increase in the participation of female researchers, who have accounted for 23% of successful applications so far.

We know that it takes some time for the research started during these visits to reach conclusions and be published. However, with the visitor programme now having been running for nearly two years, we have seen a recent surge in the number of journal publications and conference presentations arising from the visits. We are aware of at least 50 publications and presentations which have resulted from the 265 visits completed to date, including recent or upcoming publications in journals such as ChemPhysChem, Advanced Materials, and the Journal of Computational Physics: X.

Feedback on the programme from both visitors and host researchers has been overwhelmingly positive, with one visitor stating that it had been "the best professional and personal experience of my life".

Hosts have commented that it is an "excellent way to have international collaborations with excellent scientists", and have said they appreciate that HPC-Europa takes care of the administrative aspects as well as the HPC support, leaving them to concentrate on the core science.

Closing dates for applications are held 4 times per year. The next closing dates are 20th February and 14th May 2020.

HPC-Europa3 is funded until April 2021 by the European Union's Horizon 2020 research and innovation programme.

Catherine Inglis, EPCC
c.inglis@epcc.ed.ac.uk

Read more about our visitors and their research on our blog: https://edin.ac/2lfizgz

Or watch short interviews: http://bit.ly/HPCEuropa3-videos

To apply for an HPC-Europa3 visit or to become a host researcher, see: www.hpc-europa.org

If you have any questions which are not answered on the website, please contact: staff@hpc-europa.org.

In demand! Masters programmes at EPCC

Class of 2019

The class of 2019 is the eighteenth cohort of EPCC MSc students and the first to be based at the Bayes Centre home. Our new surroundings have been immensely beneficial to the student experience, with the student desk space quite literally on the doorstep of staff offices and plenty of space to breakout for group study. Despite 2018/19 being our largest MSc intake in terms of entrants to that point, it was an incredibly cohesive group.

It is always sad to see the students depart EPCC, but it's also a source of great joy when they successfully complete their studies and embark on the next stage of their careers. Students from this year have gone on to graduate destinations as diverse as Atos, Deloitte, Craneware, SeeByte, and EPCC itself. At least three students have gone on to PhD study, including one who is remaining at EPCC as part of the ASiMoV Prosperity Partnership.

Class of 2020

As much as we miss last year's students, we have been delighted to welcome the next MSc cohort to EPCC.

The class of 2020 is by far the largest we have had at EPCC so far, with 69 new students from 13 countries joining the programmes, alongside two continuing part-time students. Demand for the skills provided by our MSc programme is clearly demonstrated by the growth of both programmes, along with the large number of potential employers who contact us to pass job vacancies on to our alumni and graduating students.

We hope to be able to properly introduce you to many of the students in this year's cohort in a future issue of EPCC News.

Looking ahead: milestones and new frontiers

The 2020/21 academic year still seems a long way off, but it marks a number of important milestones: EPCC's thirtieth anniversary, the MSc in High Performance Computing's twentieth cohort, and the launch of our online MSc programmes.

Ben Morse, EPCC
b.morse@epcc.ed.ac.uk

This year's cohort of students is our largest ever, reflecting the increasing demand for the skills we offer.

Above: Some of the students of the 2018/19 MSc programmes celebrate the end of their studies with EPCC staff. Photo from Julita Inca, taken by EPCC's Dr Darren White.

Left: The Bayes Centre, where EPCC staff and students are based. Image: Keith Hunter.

Below: Students' communal working areas. Image: Mark Reynolds.

Online learning

For many years now, both the MSc in High Performance Computing, and High Performance Computing with Data Science have been available on-campus as one year full-time, or part-time over two or three years. However one major obstacle to potential part-time students has always been their ability to access classes: even if they are based in Edinburgh.

For the 2020/21 academic year both programmes will also be available online as part-time degrees, which students can complete over three to six years on an intermittent basis to enable flexibility around working/family commitments (ie students do not need to complete a set amount of credits within a particular year or Semester of the programme and there is no annual fee. Instead students pay on a per-course basis as and when they take them).

The programmes will also have multiple entry routes, including Postgraduate Diploma (PGDip), Postgraduate Certificate (PGCert), and Postgraduate Professional Development (PPD). PPD is a non-graduating route which allows students to take up to 50 credits of courses from the curriculum.

Given the increasing demand for places on the MSc programmes and the skills they teach, we hope the new online part-time study options will provide further avenues for a wide variety of people, including members of the RSE community, and those looking to develop or refresh skills.

Further information

On-campus programmes

MSc/PGDip High Performance Computing:
https://edin.ac/2vu0nJn

MSc High Performance Computing with Data Science:
https://edin.ac/2MPBErf

Online programmes

MSc/PGDip/PGCert/PPD in High Performance Computing:
https://edin.ac/2J1Y9YL

MSc/PGDip/PGCert/PPD in High Performance Computing with Data Science:
https://edin.ac/2pA7TCA

NEXTGenIO: the end is just the beginning

After four years of hard work, the NEXTGenIO project has now come to an end. It has been an extremely enjoyable and successful collaboration with a dedicated group of HPC users, software and tools developers, and hardware providers from across Europe.

Michèle Weiland, EPCC
m.weiland@epcc.ed.ac.uk

In late March this year, the system we designed and developed in the project – a 34-node prototype with Intel's new DCPMM technology – was delivered to EPCC's Advanced Computing Facility. Since then we have knuckled down and worked hard on integrating the system software stack, ensuring it is user friendly and robust. We have also spent a lot of effort on analysing the performance of our key applications, and the performance we see from the NEXTGenIO prototype is extremely impressive.

Some early results for the CASTEP computation chemistry code and the European Centre for Medium-Range Weather Forecasts (ECMWF) weather forecasting simulation IFS are published in our SC19 Paper "An Early Evaluation of Intel's Optane DC Persistent Memory Module and its Impact on High-Performance Scientific Applications". For OpenFOAM, a popular CFD package, we see a huge improvement in the performance when writing data directly to the DCPMM, rather than writing to the parallel file system.

In September we ran the first NEXTGenIO hackathon. Held at ECMWF, attendees were given hands-on experience with the NEXTGenIO prototype and learned how to program the DCPMM (see photograph).

EPCC is committed to running the NEXTGenIO prototype for the next three years – so even though the project itself has ended, there is still a lot of life (and research) left in NEXTGenIO!

NEXTGenIO at SC19

Paper: Optane DC Persistent Memory Module and Its Impact on High-Performance Scientific Applications

The new byte-addressable persistent non-volatile memory technology from Intel, DCPMM, promises to be an exciting opportunity, with unprecedented levels of capacity at near-DRAM speeds.

This paper will be presented at SC19. See: http://bit.ly/2OEToG

The first NEXTGenIO hackathon.

Accessing the NEXTGenIO prototype

If you are interested in using the NEXTGenIO prototype for your research, please get in touch to discuss your requirements: nextgenio-epcc@mlist.is.ed.ac.uk

Project website

www.nextgenio.eu

EPCC at New Scientist Live

Above: our parallel beanbag sorting game. Below: some of the EPCC Outreach team. Thanks to Spyro Nita for the photographs.

New Scientist Live is a four-day annual science festival held in the city of London. With more than 40,000 visitors over four days, it is a great opportunity to promote STEM careers to young people and to explain what we do.

New Scientist Live has five main themes, reflected in the exhibition floor layout: Humans, Technology, Earth, Engineering, and Space. Our HPC and data science domains touch on all five themes, however this year we were situated in the heart of the Technology Zone.

Visitors to the EPCC stand could learn about supercomputing hardware, as well as applications. Many people were interested to know how performant ARCHER is and what its current standing is in the global rankings, and we used this as an opportunity to tell people about the new ARCHER2 system that will be hosted by EPCC from next year. For exhibits, we brought our tried and tested high-throughput crowd favourites:

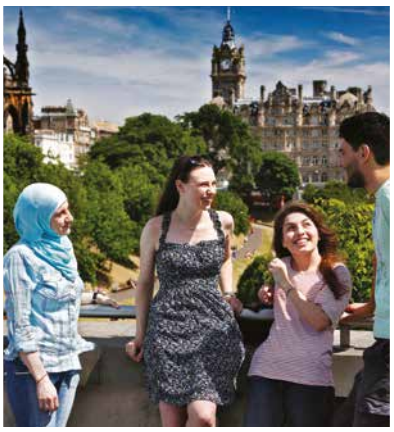
- Wee Archie plus two demos based on ARCHER use cases. Firstly a wind-tunnel simulation, where users design a wing aerofoil and see whether this generates enough upward force for the aeroplane to take off. In the second demo, the user places coastal defences on a map screen and Wee Archie simulates how they will affect the height of tidal waves. It also compares the value of the savings made by protecting a town against a baseline where no defences are in place.
- An XC30 blade like those found in ARCHER, plus lifesize banners of a cabinet to show how these are part of a far larger system.
- A parallel beanbag sorting game, which demonstrates how sorting efficiency changes as the number of sorters increases. Firstly it increases linearly, before contention for access to the unsorted beanbag supply will cause it to drop off.

New Scientist Live is such a busy event that we were all ready for a rest at the end of it, however the great conversations we had with visitors made it one of our most enjoyable outreach trips. We are already planning demos for next year!

Jane Kennedy, EPCC
j.kennedy@epcc.ed.ac.uk

EPCC's Weronika Filinger will host a Birds of a Feather session at SC19 entitled "Community Building for Sustainable and Scalable HPC Outreach" on Tuesday 19th November. If you're attending the conference and are interested in participating in the HPC Outreach community, please come along!

Outreach at EPCC
www.epcc.ed.ac.uk/discover-and-learn




Prepare for the future with a postgraduate degree from EPCC

Studying with us can take your career to the next level

EPCC at the University of Edinburgh offers two postgraduate Master of Science (MSc) programmes:

- High Performance Computing (HPC)
- HPC with Data Science

Available both online and on-campus

On-campus: 1 year full-time or part-time MSc over either 2 or 3 years
Online: part-time intermittent study for between 3-6 years for full MSc
Online study: Offers the opportunity to take specific courses from the programmes on a standalone basis as Postgraduate Professional Development (PPD) to develop and improve specific skills without committing to a full programme of study.

Study with us and you'll benefit from world-class research, connect with top industry professionals, and graduate with a globally recognised degree from a university ranked among the top 50 in the world.*
*Times Higher World University Ranking.

THE UNIVERSITY of EDINBURGH

www.epcc.ed.ac.uk/msc

epcc